GAMs to Analyze Plankton Comunity Using Fish, not River Herring

Curtis C. Bohlen, Casco Bay Estuary Partnership

6/16/2022

Contents

Introduction

This notebook reprises relevant analyses presented in the "Final-Gams.pdf" notebook, just substituting "Fish" as a predictor where previously we had looked at "RH, for River Herring. Most of the code should just run"out of the box" based on a glopbal seach and replace. . . ..

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`
```

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
#>   <dttm>              <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>, discharge_week_cftpersec <dbl>,
#> #   discharg_day <dbl>, discharge_week_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>, ...
```

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')),
         Yearf = factor(Year)) %>%
  rename(Season = season,
         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, DOY, riv_km,
         Temp, Sal, Turb, AvgTurb, DOsat, Chl,
         Fish, RH,
         combined_density,H, SEI,
         Acartia, Balanus, Eurytemora, Polychaete, Pseudocal, Temora) %>%
  arrange(Date, Station)
head(base_data)
#> # A tibble: 6 x 25
#>   Date                Station  Year Yearf Month Season   DOY riv_km  Temp
#>   <dttm>              <fct>   <dbl> <fct> <fct> <fct>  <dbl>  <dbl> <dbl>
#> 1 2013-05-28 00:00:00 1        2013 2013  May   Spring   148  22.6  11.7
#> 2 2013-05-28 00:00:00 2        2013 2013  May   Spring   148  13.9   9.40
#> 3 2013-05-28 00:00:00 3        2013 2013  May   Spring   148   8.12  6.97
#> 4 2013-05-28 00:00:00 4        2013 2013  May   Spring   148   2.78  9.51
```

```
#> 5 2013-07-25 00:00:00 1         2013 2013  Jul   Summer   206   22.6   18.5
#> 6 2013-07-25 00:00:00 2         2013 2013  Jul   Summer   206   13.9   13.6
#> # ... with 16 more variables: Sal <dbl>, Turb <dbl>, AvgTurb <dbl>,
#> #   DOsat <dbl>, Chl <dbl>, Fish <dbl>, RH <dbl>, combined_density <dbl>,
#> #   H <dbl>, SEI <dbl>, Acartia <dbl>, Balanus <dbl>, Eurytemora <dbl>,
#> #   Polychaete <dbl>, Pseudocal <dbl>, Temora <dbl>
```

```
rm(station_data)
```

Add Transformed Predictors

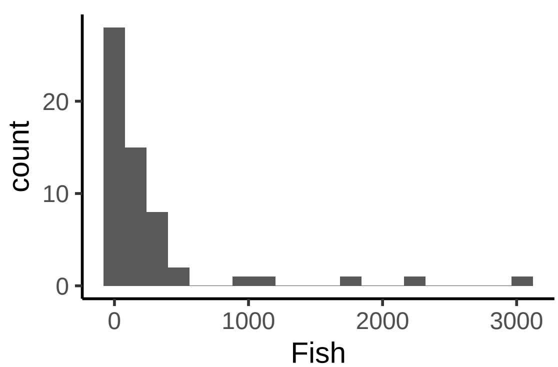
We can treat the sampling history as “spring”, “summer” and “fall” observations each year from 2013 through 2017. This breaks the temporal pattern down into integer valued time, generating a “quasi regular” time series, and allowing us to simplify the analysis of temporal autocorrelation. The “real world” time difference across the winter is longer that between seasons, but I could not find a ready way to address that.

We need both the numerical sequence and a factor later, for different purposes.

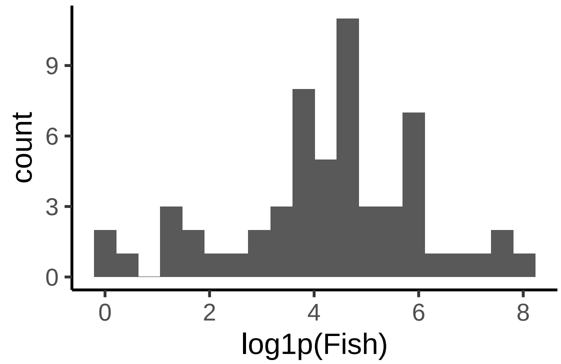
```
base_data <- base_data %>%
  mutate(sample_seq = as.numeric(Season) + (Year-2013)*3,
         sample_event = factor(sample_seq))
```

Check Distributuion of the “Fish” Abundance

```
base_data %>%
  ggplot(aes(x = Fish)) +
  geom_histogram(bins = 20)
#> Warning: Removed 2 rows containing non-finite values (stat_bin).
```



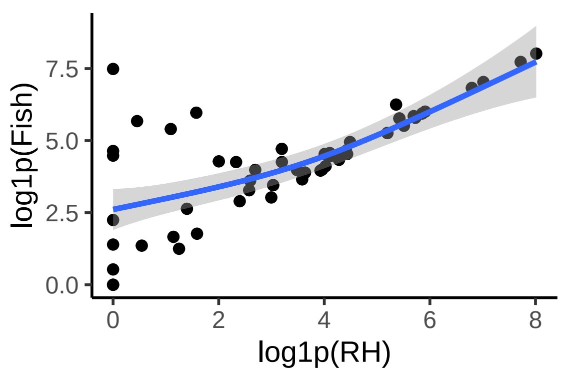

```
base_data %>%
  ggplot(aes(x = log1p(Fish))) +
  geom_histogram(bins = 20)
#> Warning: Removed 2 rows containing non-finite values (stat_bin).
```



A log transform should work O.K. as it did for River Herring, so all code should run just changing "RH" to "Fish"...

Compare "Fish" and "RH"

```
base_data %>%
  ggplot(aes(x = log1p(RH), y = log1p(Fish))) +
  geom_point() +
  geom_smooth(method = 'gam', formula = y~s(x))
#> Warning: Removed 2 rows containing non-finite values (stat_smooth).
#> Warning: Removed 2 rows containing missing values (geom_point).
```



So, the two transformed measures are correlated, as expected. They diverge where river herring abundance is low.

Model of Fish Abundance

```
fish_gam <- gamm(log1p(Fish) ~ Station +
                    Season +
                    s(Temp, bs="ts") +
                    s(Sal, bs="ts") +
```

```
                     s(log(Turb), bs="ts") +
                     s(log(Chl), bs="ts") +
                     s(log1p(combined_density),bs="ts"),
                   random = list(Yearf = ~ 1, sample_event = ~ 1),
                   data = base_data, family = 'gaussian')
summary(fish_gam$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Fish) ~ Station + Season + s(Temp, bs = "ts") + s(Sal,
#>     bs = "ts") + s(log(Turb), bs = "ts") + s(log(Chl), bs = "ts") +
#>     s(log1p(combined_density), bs = "ts")
#>
#> Parametric coefficients:
#>               Estimate Std. Error t value Pr(>|t|)
#> (Intercept)     6.5740     0.8577   7.665 5.12e-10 ***
#> Station2       -1.5718     0.8103  -1.940   0.0580 .
#> Station3       -1.7365     0.7515  -2.311   0.0250 *
#> Station4       -1.8053     0.7952  -2.270   0.0275 *
#> SeasonSummer   -1.0156     0.7012  -1.448   0.1537
#> SeasonFall     -1.9043     0.7316  -2.603   0.0121 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                                  edf Ref.df     F p-value
#> s(Temp)                    3.771e-08      9 0.000  0.2524
#> s(Sal)                     7.680e-01      9 0.385  0.0557 .
#> s(log(Turb))               3.180e-09      9 0.000  0.8968
#> s(log(Chl))                5.184e-09      9 0.000  0.7949
#> s(log1p(combined_density)) 6.414e-01      9 0.197  0.1351
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.123
#>   Scale est. = 2.2403    n = 58
```

```
anova(fish_gam$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Fish) ~ Station + Season + s(Temp, bs = "ts") + s(Sal,
#>     bs = "ts") + s(log(Turb), bs = "ts") + s(log(Chl), bs = "ts") +
#>     s(log1p(combined_density), bs = "ts")
#>
#> Parametric Terms:
#>         df     F p-value
#> Station  3 2.051  0.1184
#> Season   2 3.411  0.0408
```

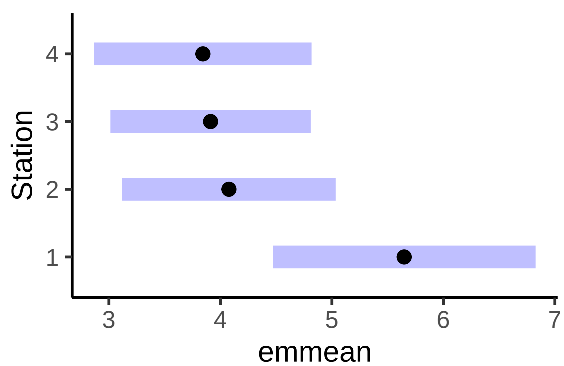
```
#>
#> Approximate significance of smooth terms:
#>                                 edf    Ref.df     F p-value
#> s(Temp)                   3.771e-08 9.000e+00 0.000  0.2524
#> s(Sal)                    7.680e-01 9.000e+00 0.385  0.0557
#> s(log(Turb))              3.180e-09 9.000e+00 0.000  0.8968
#> s(log(Chl))               5.184e-09 9.000e+00 0.000  0.7949
#> s(log1p(combined_density)) 6.414e-01 9.000e+00 0.197  0.1351
```

In comparison to the River Herring model, Station is no longer statistically significant by ANOVA, but Season is. Salinity is now only marginally significant.

My instinct here would be to simplify this model and see if that clarifies relationships any. We're carrying a lot of predictors that have low effect in these models, which may nevertheless alter interpretation of the other terms.

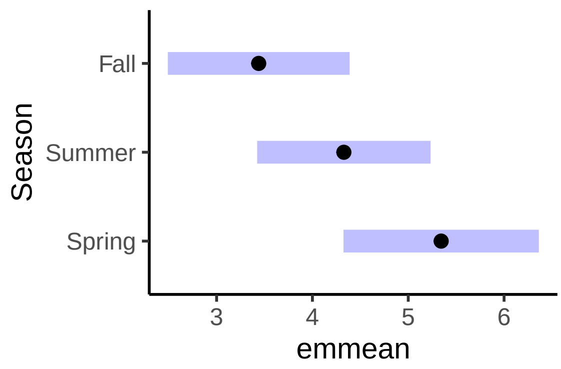
Station and Season (Station not significant)

```
Sta_emms <- emmeans(fish_gam, ~Station, type = 'response',
                    data = base_data)
plot(Sta_emms)
```




```
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast            estimate    SE   df t.ratio p.value
#>  Station1 - Station2   1.5718 0.810 50.6   1.940  0.3480
#>  Station1 - Station3   1.7365 0.751 50.6   2.311  0.1497
#>  Station1 - Station4   1.8053 0.795 50.6   2.270  0.1648
#>  Station2 - Station3   0.1647 0.585 50.6   0.282  1.0000
#>  Station2 - Station4   0.2335 0.592 50.6   0.394  1.0000
#>  Station3 - Station4   0.0688 0.594 50.6   0.116  1.0000
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

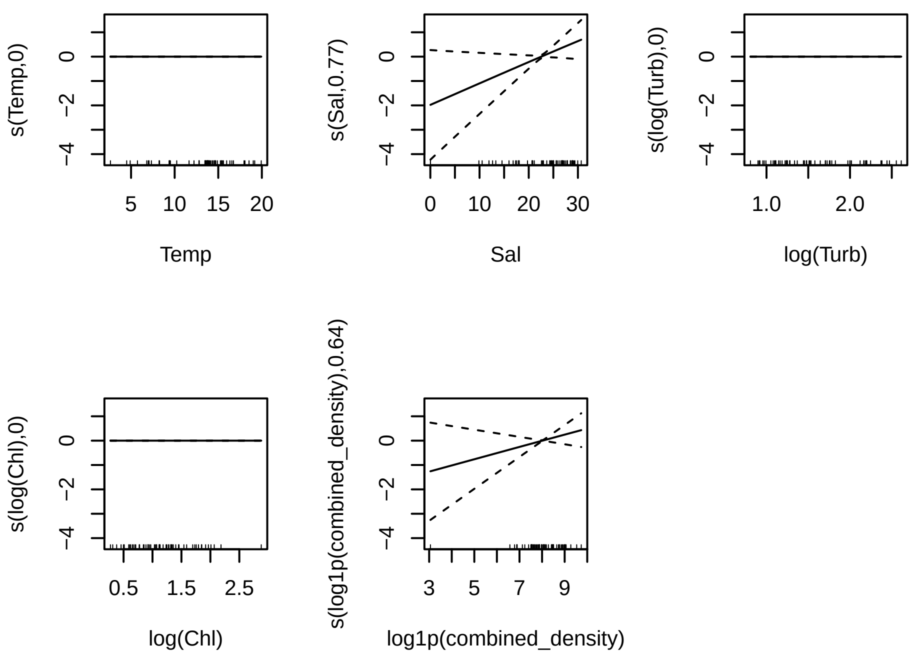
```
Seas_emms <- emmeans(fish_gam, ~Season, type = 'response',
                     data = base_data)
plot(Seas_emms)
```




```
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate    SE   df t.ratio p.value
#>  Spring - Summer    1.016 0.701 50.6   1.448  0.4610
#>  Spring - Fall      1.904 0.732 50.6   2.603  0.0363
#>  Summer - Fall      0.889 0.624 50.6   1.424  0.4815
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

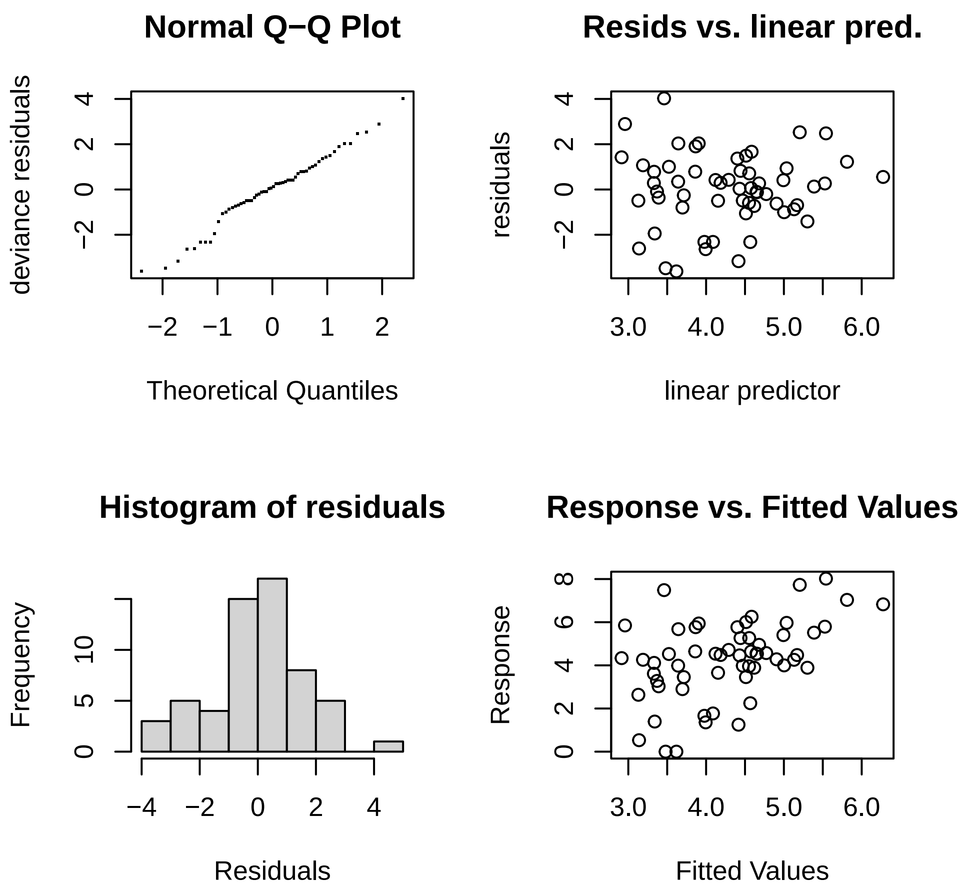
Plot GAM results

```
oldpar <- par(mfrow = c(2,3))
plot(fish_gam$gam)
par(oldpar)
```



Model Diagnostics

```r
oldpar <- par(mfrow = c(2,2))
gam.check(fish_gam$gam)
```




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)                      9.00e+00 3.77e-08    1.04    0.54
#> s(Sal)                       9.00e+00 7.68e-01    0.93    0.26
#> s(log(Turb))                 9.00e+00 3.18e-09    0.91    0.22
#> s(log(Chl))                  9.00e+00 5.18e-09    1.05    0.57
#> s(log1p(combined_density)) 9.00e+00 6.41e-01    0.93    0.23
par(oldpar)
```

The model is pretty good, with only slightly skewed residuals.

Total Zooplankton Density

Summary and Anova

```
density_gam_fish <- gamm(log(combined_density) ~
                          Station +
                          Season +
                          s(Temp, bs="ts") +
                          s(Sal, bs="ts") +
                          s(log(Turb), bs="ts") +
                          s(log(Chl), bs="ts") +
                          s(log1p(Fish),bs="ts"),
                        random = list(Yearf = ~ 1, sample_event = ~ 1),
                        data = base_data, family = 'gaussian')
summary(density_gam_fish$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log(combined_density) ~ Station + Season + s(Temp, bs = "ts") +
#>     s(Sal, bs = "ts") + s(log(Turb), bs = "ts") + s(log(Chl),
#>     bs = "ts") + s(log1p(Fish), bs = "ts")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    9.3298     0.4471  20.868  < 2e-16 ***
#> Station2      -1.0127     0.2760  -3.669 0.000624 ***
#> Station3      -0.7620     0.2627  -2.900 0.005673 **
#> Station4      -1.1834     0.2943  -4.020 0.000211 ***
#> SeasonSummer  -0.8744     0.3377  -2.589 0.012798 *
#> SeasonFall    -0.7889     0.3203  -2.463 0.017533 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                      edf Ref.df      F  p-value
#> s(Temp)        9.539e-05      9  0.000 0.112921
#> s(Sal)         3.437e+00      9 12.044  < 2e-16 ***
#> s(log(Turb))   8.029e-01      9  0.420 0.049330 *
#> s(log(Chl))    1.186e+00      9  2.021 0.000268 ***
#> s(log1p(Fish)) 3.036e-05      9  0.000 0.833827
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.258
#>   Scale est. = 0.17578   n = 58
```

Comparison to River Herring Model

```r
density_gam_rh<- gamm(log(combined_density) ~
                          Station +
                          Season +
                          s(Temp, bs="ts") +
                          s(Sal, bs="ts") +
                          s(log(Turb), bs="ts") +
                          s(log(Chl), bs="ts") +
                          s(log1p(RH),bs="ts"),
                        random = list(Yearf = ~ 1, sample_event = ~ 1),
                        data = base_data, family = 'gaussian')

anova(density_gam_fish$lme, density_gam_rh$lme)
#>                        Model df      AIC      BIC    logLik
#> density_gam_fish$lme       1 14 139.2853 168.1315 -55.64263
#> density_gam_rh$lme         2 14 139.2851 168.1313 -55.64257
```

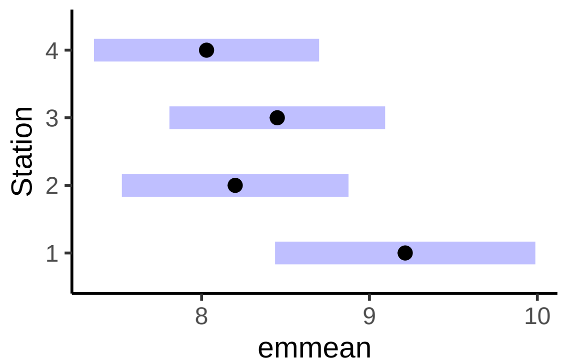
So, in this particular test, the different fish predictors make essentially no difference. I reviewed parameter values (not shown), and the don't change. I suppose that is not too surprising, given that River Herring was not an important predictor. . . .

```r
anova(density_gam_fish$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log(combined_density) ~ Station + Season + s(Temp, bs = "ts") +
#>     s(Sal, bs = "ts") + s(log(Turb), bs = "ts") + s(log(Chl),
#>     bs = "ts") + s(log1p(Fish), bs = "ts")
#>
#> Parametric Terms:
#>         df     F p-value
#> Station  3 6.020 0.00149
#> Season   2 3.802 0.02955
#>
#> Approximate significance of smooth terms:
#>                      edf    Ref.df      F  p-value
#> s(Temp)        9.539e-05 9.000e+00  0.000 0.112921
#> s(Sal)         3.437e+00 9.000e+00 12.044  < 2e-16
#> s(log(Turb))   8.029e-01 9.000e+00  0.420 0.049330
#> s(log(Chl))    1.186e+00 9.000e+00  2.021 0.000268
#> s(log1p(Fish)) 3.036e-05 9.000e+00  0.000 0.833827
```

Station, Season, Salinity, Turbidity, and CHlorophyll are all significnat predictors in this model.

Station and Season

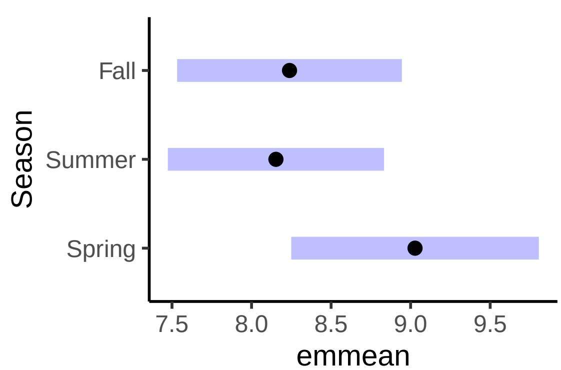
```r
Sta_emms <- emmeans(density_gam_fish, ~Station, type = 'response',
                    data = base_data)
plot(Sta_emms)
```




```r
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast            estimate    SE   df t.ratio p.value
#>  Station1 - Station2    1.013 0.276 46.6   3.669  0.0037
#>  Station1 - Station3    0.762 0.263 46.6   2.900  0.0340
#>  Station1 - Station4    1.183 0.294 46.6   4.020  0.0013
#>  Station2 - Station3   -0.251 0.188 46.6  -1.331  1.0000
#>  Station2 - Station4    0.171 0.201 46.6   0.851  1.0000
#>  Station3 - Station4    0.421 0.179 46.6   2.357  0.1362
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

Station 1 has significantly higher zooplankton than other stations.

```r
Seas_emms <- emmeans(density_gam_fish, ~Season, type = 'response',
                     data = base_data)
plot(Seas_emms)
```



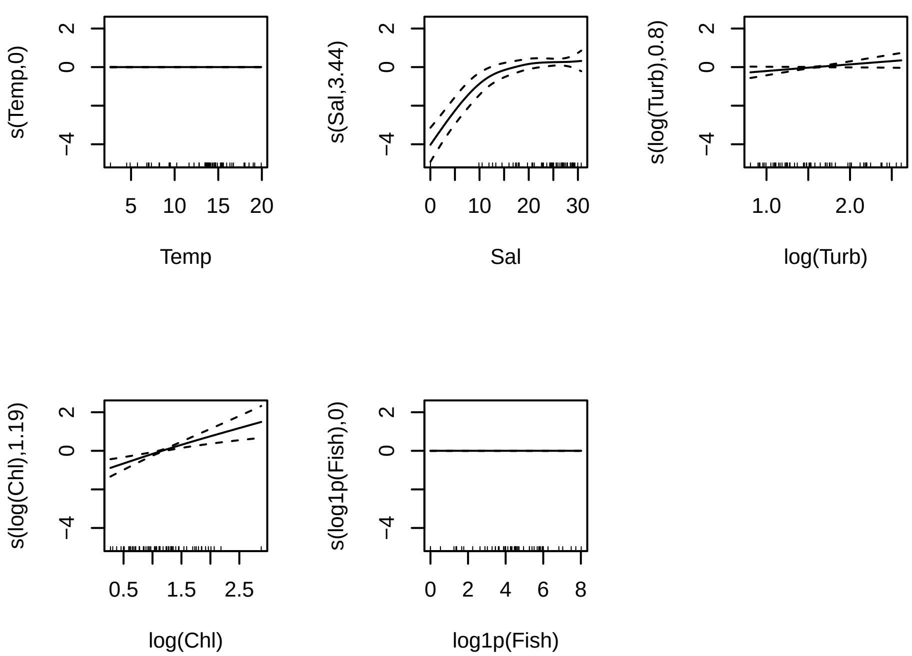

```r
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate    SE   df t.ratio p.value
#>  Spring - Summer   0.8744 0.338 46.6   2.589  0.0384
```

```
#>  Spring - Fall      0.7889 0.320 46.6   2.463  0.0526
#>  Summer - Fall     -0.0854 0.262 46.6  -0.326  1.0000
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

And Spring has significantly higher Zooplanton density than the fall.

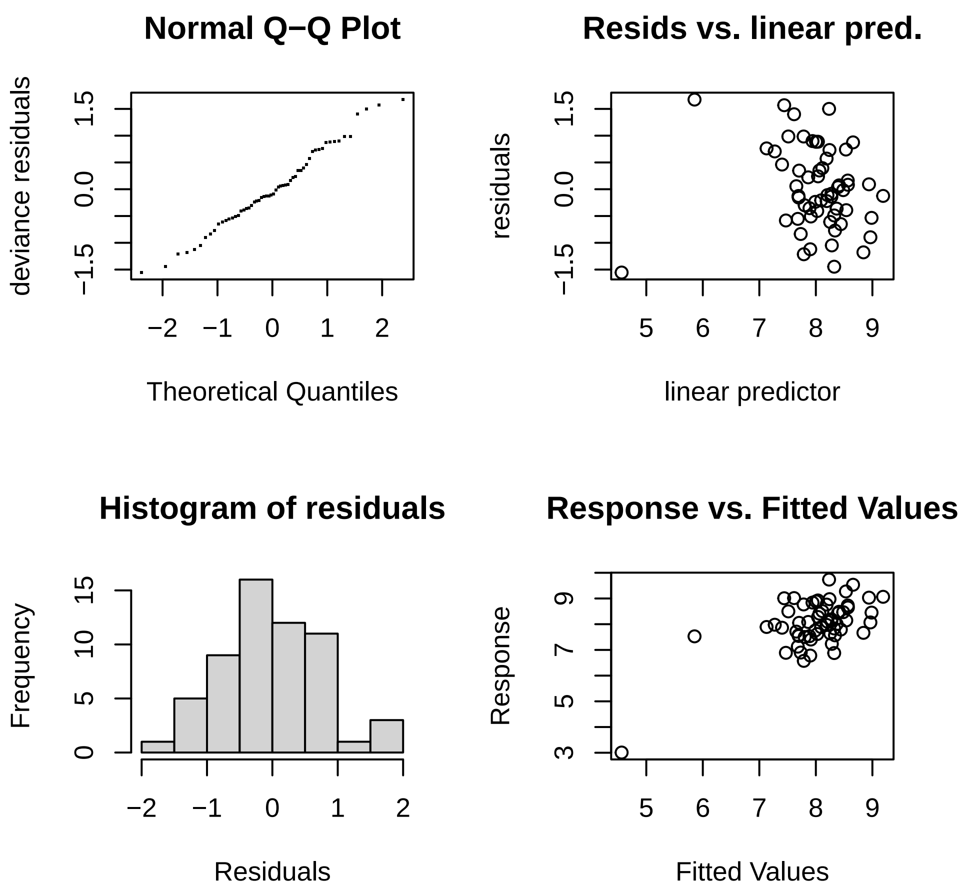
Plot the GAM

```
oldpar <- par(mfrow = c(2,3))
plot(density_gam_fish$gam)
par(oldpar)
```



Model Diagnostics

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




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)        9.00e+00 9.54e-05    1.07    0.64
#> s(Sal)         9.00e+00 3.44e+00    1.03    0.56
#> s(log(Turb))   9.00e+00 8.03e-01    0.89    0.16
#> s(log(Chl))    9.00e+00 1.19e+00    0.94    0.28
#> s(log1p(Fish)) 9.00e+00 3.04e-05    1.00    0.45
par(oldpar)
```

So, nothing much has changed. One big outlier – presumably one of those spring “washout” samples.

Shannon Diversity

Summary and Anova

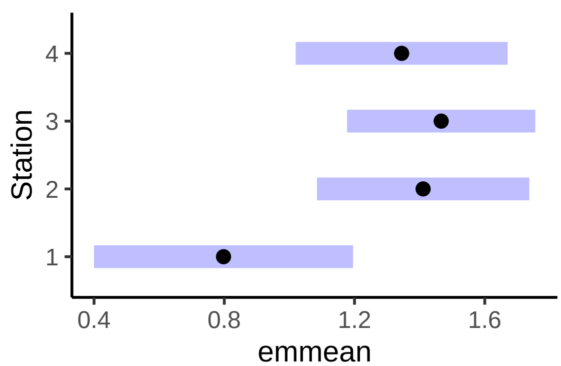
```
shannon_gam <- gamm(H ~ Station +
                     Season +
                     s(Temp, bs="ts") +
                     s(Sal, bs="ts") +
                     s(log(Turb), bs="ts") +
                     s(log(Chl), bs="ts") +
                     s(log1p(Fish),bs="ts"),
                   random = list(Yearf = ~ 1, sample_event = ~ 1),
                   data = base_data, family = 'gaussian')
#> Warning in lme.formula(y ~ X - 1, random = rand, data = strip.offset(mf), : nlminb problem, converge
#>   message = iteration limit reached without convergence (10)
summary(shannon_gam$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> H ~ Station + Season + s(Temp, bs = "ts") + s(Sal, bs = "ts") +
#>     s(log(Turb), bs = "ts") + s(log(Chl), bs = "ts") + s(log1p(Fish),
#>     bs = "ts")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   0.86216    0.27100   3.181  0.00254 **
#> Station2      0.61303    0.24771   2.475  0.01683 *
#> Station3      0.66851    0.22908   2.918  0.00530 **
#> Station4      0.54704    0.24141   2.266  0.02788 *
#> SeasonSummer  0.09502    0.20150   0.472  0.63933
#> SeasonFall   -0.01715    0.21941  -0.078  0.93803
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                     edf Ref.df     F p-value
#> s(Temp)        7.388e-06      9 0.000 0.46622
#> s(Sal)         2.398e+00      9 1.716 0.00338 **
#> s(log(Turb))   1.412e-07      9 0.000 0.93052
#> s(log(Chl))    1.450e-06      9 0.000 0.87186
#> s(log1p(Fish)) 4.429e-01      9 0.098 0.20885
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.258
#>   Scale est. = 0.16688   n = 58
```

```
anova(shannon_gam$gam)
#>
#> Family: gaussian
```

```
#> Link function: identity
#>
#> Formula:
#> H ~ Station + Season + s(Temp, bs = "ts") + s(Sal, bs = "ts") +
#>     s(log(Turb), bs = "ts") + s(log(Chl), bs = "ts") + s(log1p(Fish),
#>     bs = "ts")
#>
#> Parametric Terms:
#>         df     F p-value
#> Station  3 2.888  0.0448
#> Season   2 0.245  0.7836
#>
#> Approximate significance of smooth terms:
#>                      edf    Ref.df     F p-value
#> s(Temp)        7.388e-06 9.000e+00 0.000 0.46622
#> s(Sal)         2.398e+00 9.000e+00 1.716 0.00338
#> s(log(Turb))   1.412e-07 9.000e+00 0.000 0.93052
#> s(log(Chl))    1.450e-06 9.000e+00 0.000 0.87186
#> s(log1p(Fish)) 4.429e-01 9.000e+00 0.098 0.20885
```

Station and Season (Season is not significant)

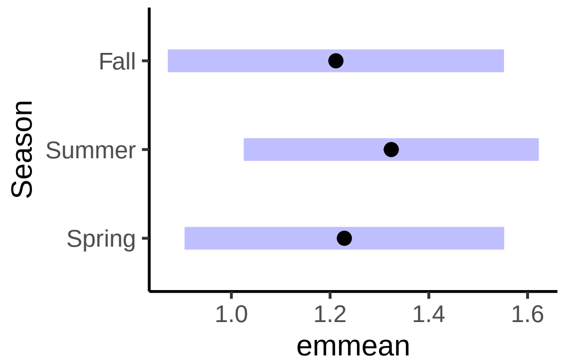
```
Sta_emms <- emmeans(shannon_gam, ~Station, type = 'response',
                    data = base_data)
plot(Sta_emms)
```




```
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast            estimate    SE   df t.ratio p.value
#>  Station1 - Station2  -0.6130 0.248 49.2  -2.475  0.1010
#>  Station1 - Station3  -0.6685 0.229 49.2  -2.918  0.0318
#>  Station1 - Station4  -0.5470 0.241 49.2  -2.266  0.1673
#>  Station2 - Station3  -0.0555 0.164 49.2  -0.339  1.0000
#>  Station2 - Station4   0.0660 0.162 49.2   0.407  1.0000
#>  Station3 - Station4   0.1215 0.164 49.2   0.740  1.0000
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

Station 1 has significantly lower zooplankton diversity.

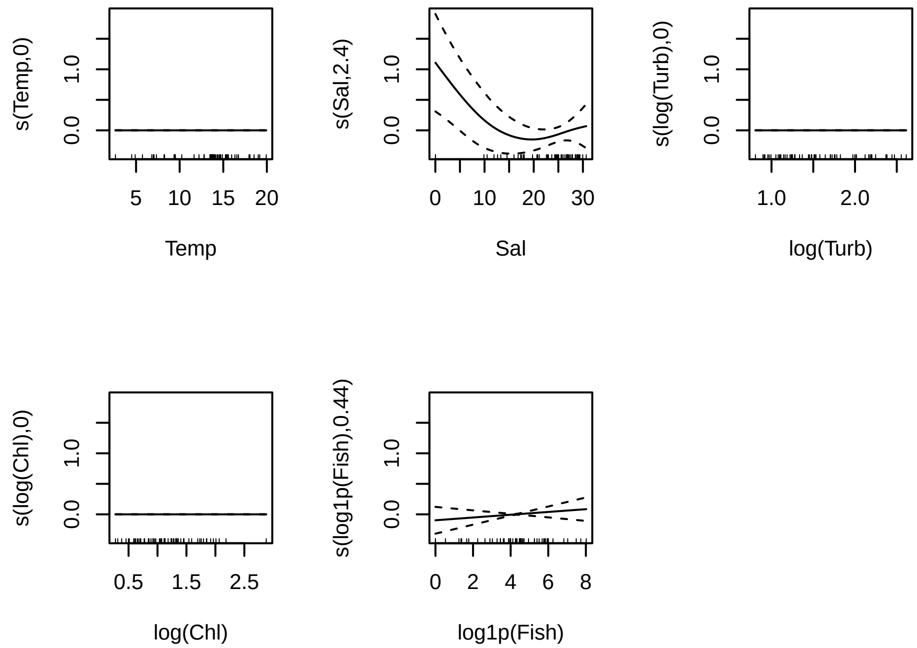
```
Seas_emms <- emmeans(shannon_gam, ~Season, type = 'response',
                     data = base_data)
plot(Seas_emms)
```




```
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate    SE   df t.ratio p.value
#>  Spring - Summer  -0.0950 0.202 49.2  -0.472  1.0000
#>  Spring - Fall     0.0171 0.219 49.2   0.078  1.0000
#>  Summer - Fall     0.1122 0.174 49.2   0.643  1.0000
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

Plot the GAM

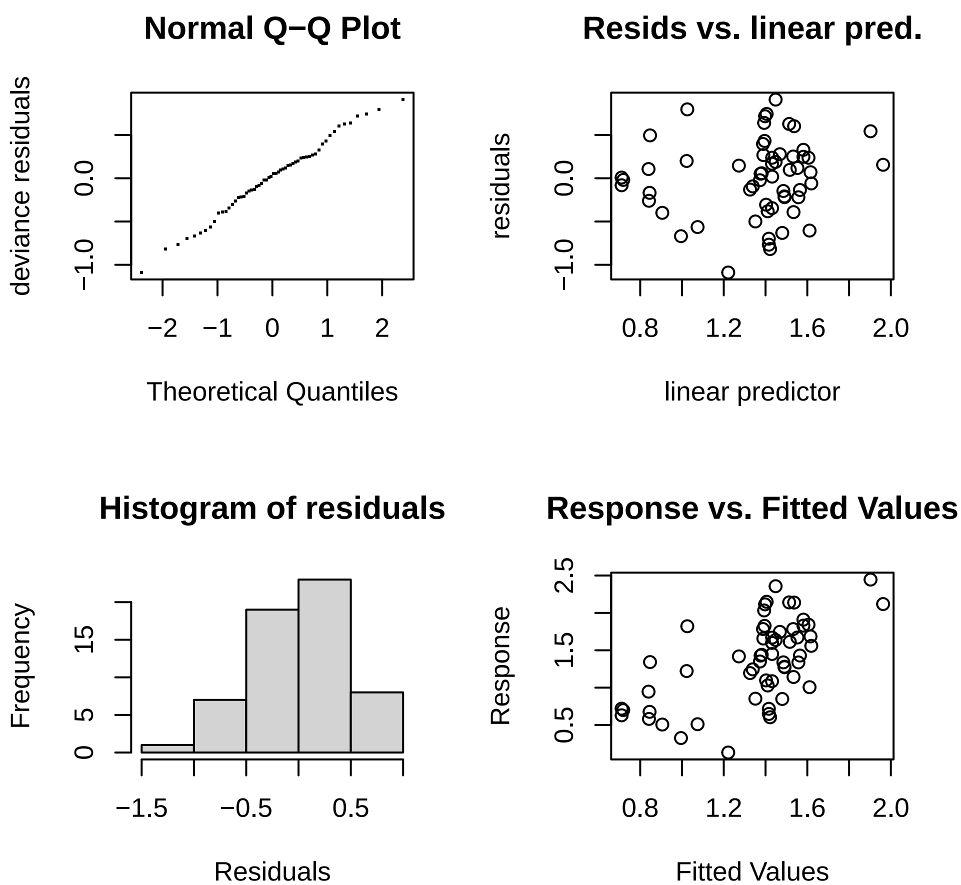
```
oldpar <- par(mfrow = c(2,3))
plot(shannon_gam$gam)
par(oldpar)
```



Again, not much changes, although here the relationship with Fish is a bit stronger than the relationship was with River Herring. Still not statistically significant in this model.

Diagnostic Plots

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




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)        9.00e+00 7.39e-06    1.00    0.44
#> s(Sal)         9.00e+00 2.40e+00    1.01    0.52
#> s(log(Turb))   9.00e+00 1.41e-07    1.28    0.98
#> s(log(Chl))    9.00e+00 1.45e-06    1.07    0.69
#> s(log1p(Fish)) 9.00e+00 4.43e-01    0.98    0.40
par(oldpar)
```

Single Species Models

Model Choice

Our model alternatives are similar to the choices we had for the Total Density model. The problem is, we can't use any of the continuous data distributions in GAMS with zero values (at least relying on the canonical link functions) because (log(0) = -Inf; 1/0 = Inf, 1 / 0*0 = Inf). The easiest solution is to add some finite small quantity to the density data, and predict that. Here we predict log(Density + 1) using Gaussian 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.

```
spp_data <- base_data %>%
  select(Yearf, Month, Season, sample_event, 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 mush 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`. Lowe numbers constrain the GAM to fit smoother lines.

```
my_gamm <- function(.dat) {

  gamm(log1p(Density) ~ Station +
        Season +
        s(Temp, bs="ts", k = 4) +
        s(Sal, bs="ts", k = 4) +
        s(log(Turb), bs="ts", k = 4) +
        s(log(Chl), bs="ts", k = 4) +
        s(log1p(Fish),bs="ts", k = 4),
      random = list(Yearf = ~ 1, sample_event = ~ 1),
      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_gamm))
```

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

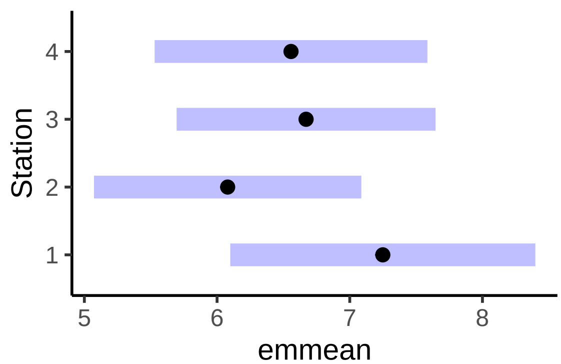
Summary and ANOVA

```
spp = 'Acartia'
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    5.6273     0.7413   7.591 7.78e-10 ***
#> Station2      -1.1691     0.5359  -2.182 0.033934 *
#> Station3      -0.5781     0.4975  -1.162 0.250799
#> Station4      -0.6917     0.5471  -1.264 0.212049
#> SeasonSummer   2.0830     0.5633   3.698 0.000546 ***
#> SeasonFall     2.1174     0.5695   3.718 0.000514 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                      edf Ref.df     F p-value
#> s(Temp)        1.255e-09      3 0.000 0.48621
#> s(Sal)         1.611e+00      3 4.780 0.00205 **
#> s(log(Turb))   8.159e-01      3 1.165 0.06015 .
#> s(log(Chl))    2.607e-01      3 0.156 0.22529
#> s(log1p(Fish)) 8.538e-08      3 0.000 0.41394
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.436
#>   Scale est. = 0.75338   n = 58
cat('\n')
anova(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric Terms:
```

```
#>          df     F  p-value
#> Station  3 1.764 0.166244
#> Season   2 8.361 0.000747
#>
#> Approximate significance of smooth terms:
#>                      edf    Ref.df     F p-value
#> s(Temp)        1.255e-09 3.000e+00 0.000 0.48621
#> s(Sal)         1.611e+00 3.000e+00 4.780 0.00205
#> s(log(Turb))   8.159e-01 3.000e+00 1.165 0.06015
#> s(log(Chl))    2.607e-01 3.000e+00 0.156 0.22529
#> s(log1p(Fish)) 8.538e-08 3.000e+00 0.000 0.41394
```

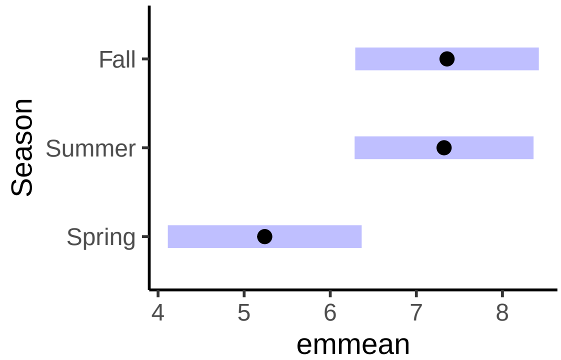
Comparison of Station and Season (Station not significant)

```
Sta_emms <- emmeans(mod, ~Station, type = 'response',
                    data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Sta_emms)
```




```
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast            estimate    SE   df t.ratio p.value
#>  Station1 - Station2    1.169 0.536 49.3   2.182  0.2036
#>  Station1 - Station3    0.578 0.498 49.3   1.162  1.0000
#>  Station1 - Station4    0.692 0.547 49.3   1.264  1.0000
#>  Station2 - Station3   -0.591 0.370 49.3  -1.599  0.6976
#>  Station2 - Station4   -0.477 0.390 49.3  -1.224  1.0000
#>  Station3 - Station4    0.114 0.353 49.3   0.322  1.0000
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

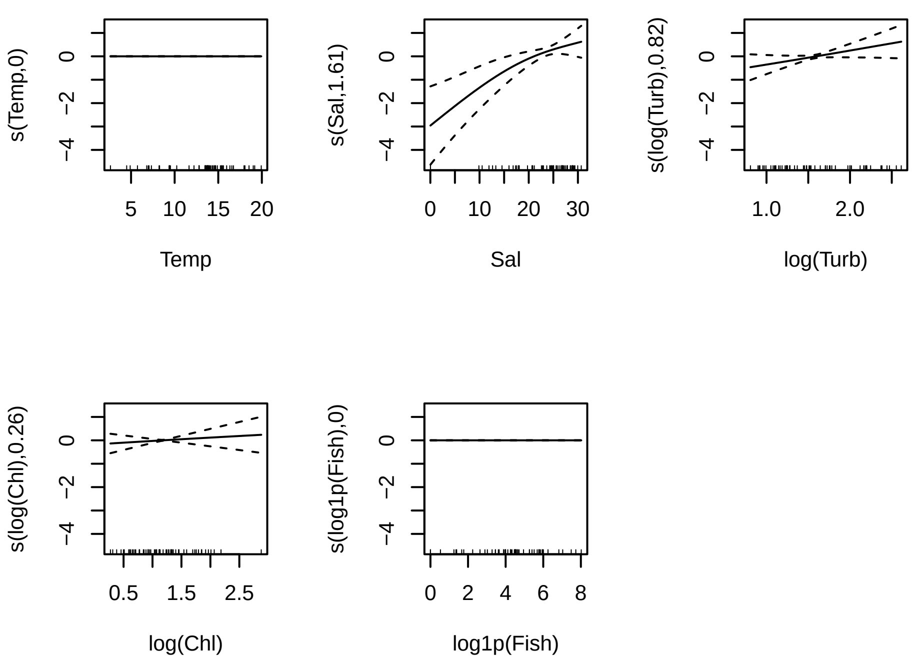
```
Seas_emms <- emmeans(mod, ~Season, type = 'response',
                     data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Seas_emms)
```




```
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate    SE   df t.ratio p.value
#>  Spring - Summer  -2.0830 0.563 49.3  -3.698  0.0016
#>  Spring - Fall    -2.1174 0.570 49.3  -3.718  0.0015
#>  Summer - Fall    -0.0344 0.478 49.3  -0.072  1.0000
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

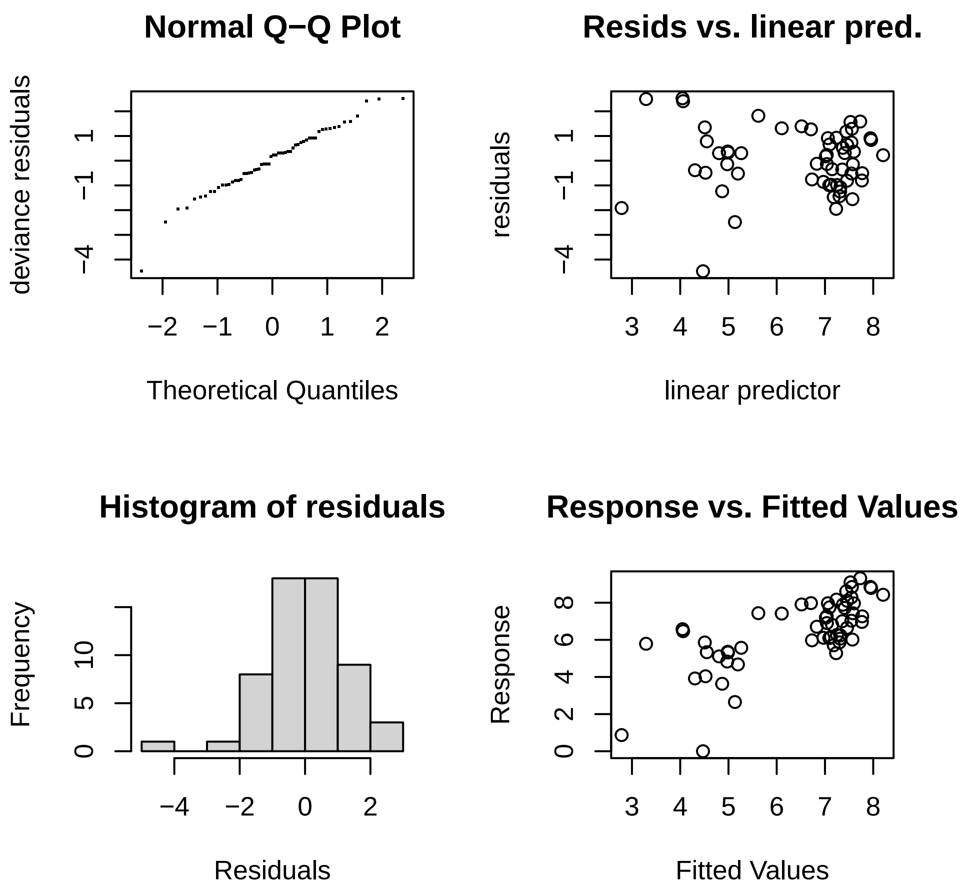
Plot GAM

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



Model Diagnostics

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




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)         3.00e+00 1.26e-09    1.03   0.555
#> s(Sal)          3.00e+00 1.61e+00    1.21   0.945
#> s(log(Turb))    3.00e+00 8.16e-01    1.08   0.670
#> s(log(Chl))     3.00e+00 2.61e-01    0.83   0.075 .
#> s(log1p(Fish)) 3.00e+00 8.54e-08    0.86   0.150
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(oldpar)
```

Balanus

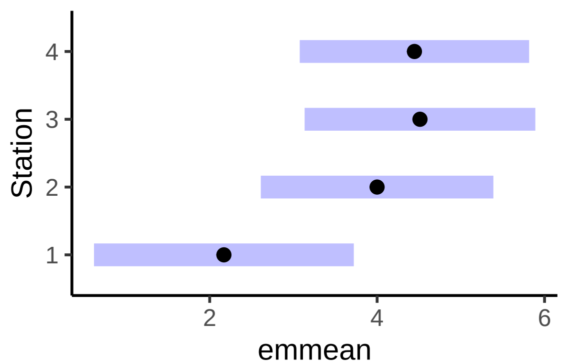
Summary and ANOVA

```
spp = 'Balanus'
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    3.5909     1.0034   3.579 0.000778 ***
#> Station2       1.8299     0.6385   2.866 0.006066 **
#> Station3       2.3423     0.5805   4.035 0.000187 ***
#> Station4       2.2761     0.6247   3.644 0.000639 ***
#> SeasonSummer  -2.0651     1.3806  -1.496 0.141003
#> SeasonFall    -2.8044     1.2947  -2.166 0.035097 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                     edf Ref.df     F p-value
#> s(Temp)        9.544e-01      3 1.428 0.04595 *
#> s(Sal)         3.537e-08      3 0.000 0.27145
#> s(log(Turb))   4.340e-09      3 0.000 0.69121
#> s(log(Chl))    1.033e+00      3 3.043 0.00628 **
#> s(log1p(Fish)) 7.143e-08      3 0.000 0.36177
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.205
#>   Scale est. = 1.1471    n = 58
cat('\n')
anova(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric Terms:
```

```
#>         df     F p-value
#> Station  3 6.032 0.00137
#> Season   2 2.389 0.10209
#>
#> Approximate significance of smooth terms:
#>                     edf    Ref.df     F p-value
#> s(Temp)        9.544e-01 3.000e+00 1.428 0.04595
#> s(Sal)         3.537e-08 3.000e+00 0.000 0.27145
#> s(log(Turb))   4.340e-09 3.000e+00 0.000 0.69121
#> s(log(Chl))    1.033e+00 3.000e+00 3.043 0.00628
#> s(log1p(Fish)) 7.143e-08 3.000e+00 0.000 0.36177
```

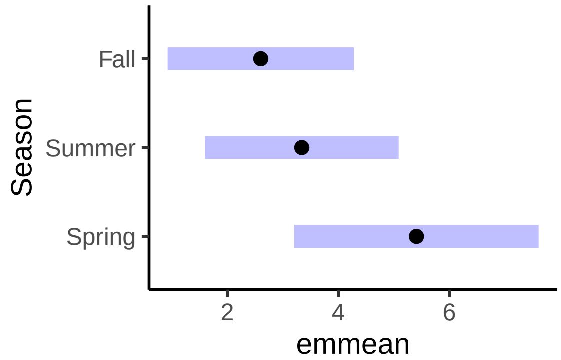
Comparison of Station and Season (Season not significant)

```
Sta_emms <- emmeans(mod, ~Station, type = 'response',
                    data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Sta_emms)
```




```
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast            estimate    SE df t.ratio p.value
#>  Station1 - Station2  -1.8299 0.638 50  -2.866  0.0364
#>  Station1 - Station3  -2.3423 0.580 50  -4.035  0.0011
#>  Station1 - Station4  -2.2761 0.625 50  -3.644  0.0038
#>  Station2 - Station3  -0.5124 0.423 50  -1.213  1.0000
#>  Station2 - Station4  -0.4461 0.432 50  -1.033  1.0000
#>  Station3 - Station4   0.0663 0.440 50   0.151  1.0000
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

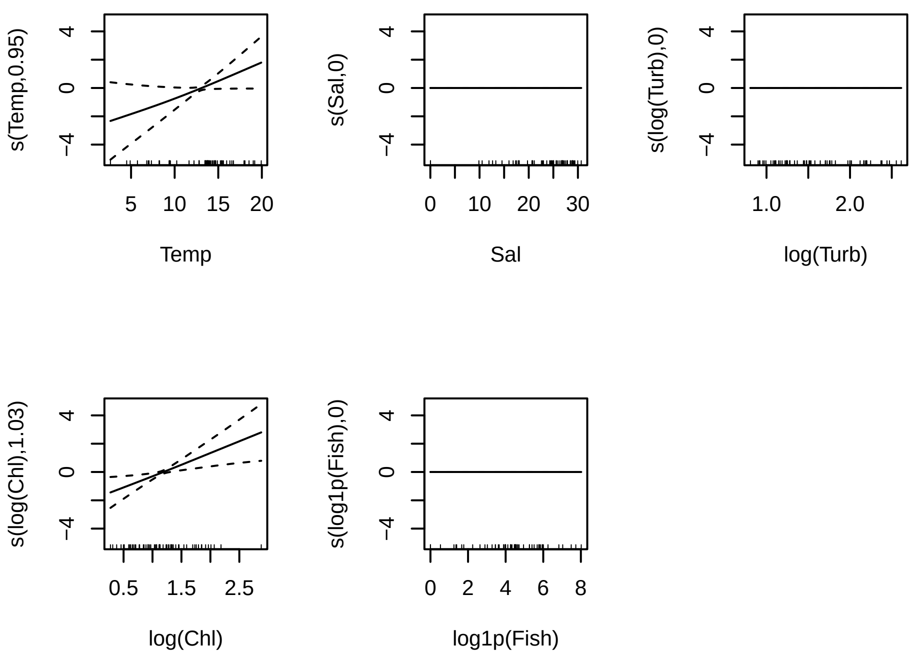
```
Seas_emms <- emmeans(mod, ~Season, type = 'response',
                     data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Seas_emms)
```




```
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate    SE df t.ratio p.value
#>  Spring - Summer    2.065 1.381 50   1.496  0.4230
#>  Spring - Fall      2.804 1.295 50   2.166  0.1053
#>  Summer - Fall      0.739 0.899 50   0.823  1.0000
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

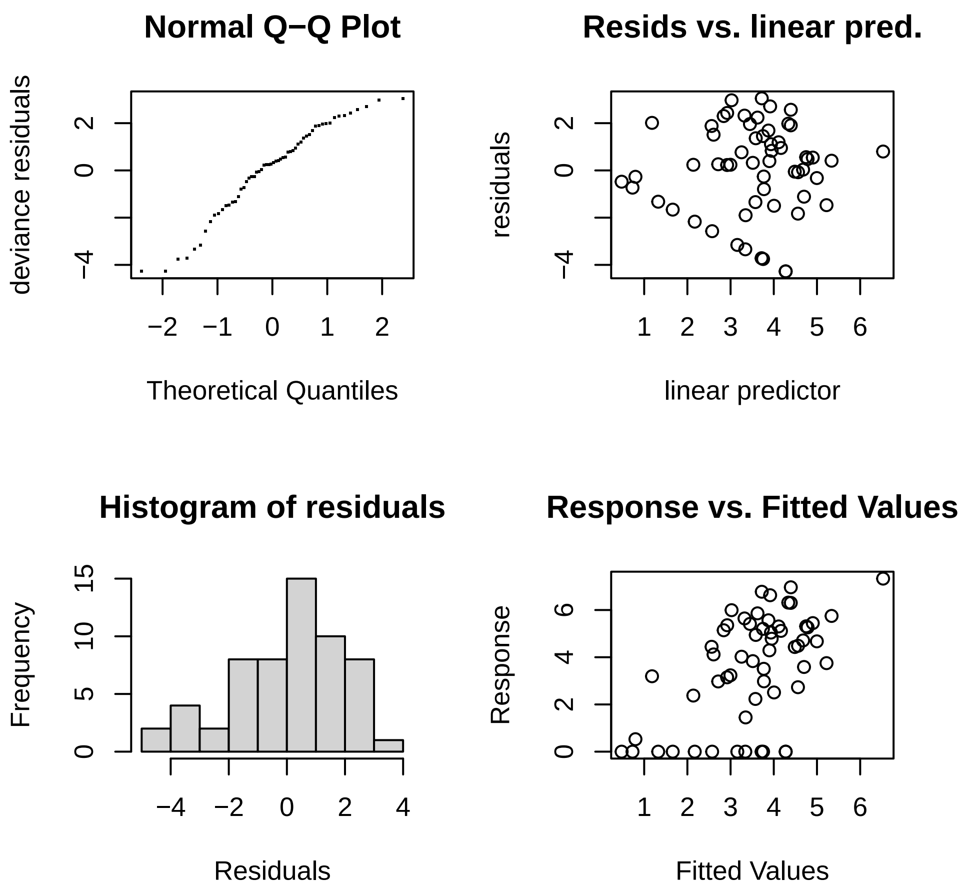
Plot GAM

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



Model Diagnostics

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




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)         3.00e+00 9.54e-01    0.87    0.15
#> s(Sal)          3.00e+00 3.54e-08    0.89    0.16
#> s(log(Turb))    3.00e+00 4.34e-09    0.99    0.38
#> s(log(Chl))     3.00e+00 1.03e+00    1.10    0.78
#> s(log1p(Fish)) 3.00e+00 7.14e-08    1.15    0.83
par(oldpar)
```

Eurytemora

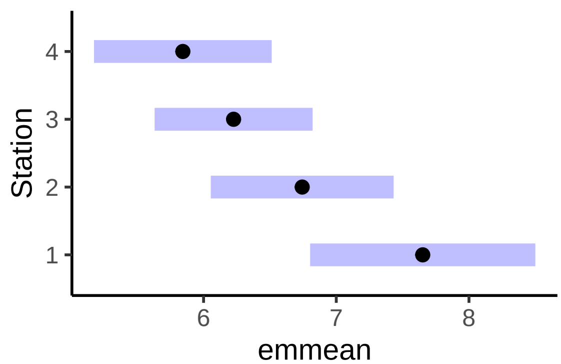
Summary and ANOVA

```
spp =  "Eurytemora"
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    8.0329     0.5290  15.186  < 2e-16 ***
#> Station2      -0.9082     0.4638  -1.958  0.05599 .
#> Station3      -1.4252     0.4343  -3.281  0.00192 **
#> Station4      -1.8077     0.4651  -3.886  0.00031 ***
#> SeasonSummer  -0.8402     0.3983  -2.110  0.04009 *
#> SeasonFall    -0.8912     0.4046  -2.203  0.03240 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                     edf Ref.df      F p-value
#> s(Temp)        8.520e-09      3  0.000  0.2810
#> s(Sal)         2.550e+00      3 17.702  <2e-16 ***
#> s(log(Turb))   8.089e-01      3  1.190  0.0571 .
#> s(log(Chl))    3.536e-01      3  0.239  0.2012
#> s(log1p(Fish)) 7.578e-10      3  0.000  0.8324
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =   0.45
#>   Scale est. = 0.59061   n = 58
cat('\n')
anova(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric Terms:
```

```
#>         df     F p-value
#> Station  3 6.034 0.00142
#> Season   2 2.757 0.07351
#>
#> Approximate significance of smooth terms:
#>                     edf    Ref.df      F p-value
#> s(Temp)        8.520e-09 3.000e+00  0.000  0.2810
#> s(Sal)         2.550e+00 3.000e+00 17.702  <2e-16
#> s(log(Turb))   8.089e-01 3.000e+00  1.190  0.0571
#> s(log(Chl))    3.536e-01 3.000e+00  0.239  0.2012
#> s(log1p(Fish)) 7.578e-10 3.000e+00  0.000  0.8324
```

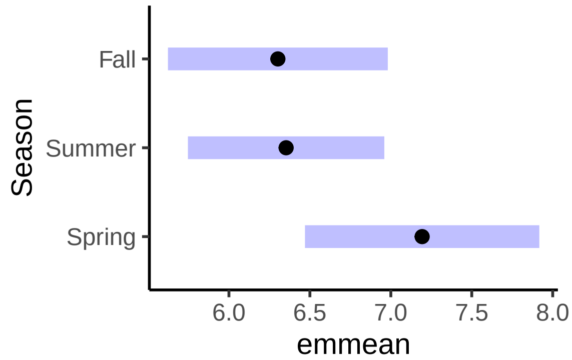
Comparison of Station and Season (Season marginally significant)

```
Sta_emms <- emmeans(mod, ~Station, type = 'response',
                    data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Sta_emms)
```




```
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast            estimate    SE   df t.ratio p.value
#>  Station1 - Station2    0.908 0.464 48.3   1.958  0.3360
#>  Station1 - Station3    1.425 0.434 48.3   3.281  0.0115
#>  Station1 - Station4    1.808 0.465 48.3   3.886  0.0019
#>  Station2 - Station3    0.517 0.331 48.3   1.563  0.7481
#>  Station2 - Station4    0.899 0.338 48.3   2.662  0.0632
#>  Station3 - Station4    0.382 0.313 48.3   1.221  1.0000
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

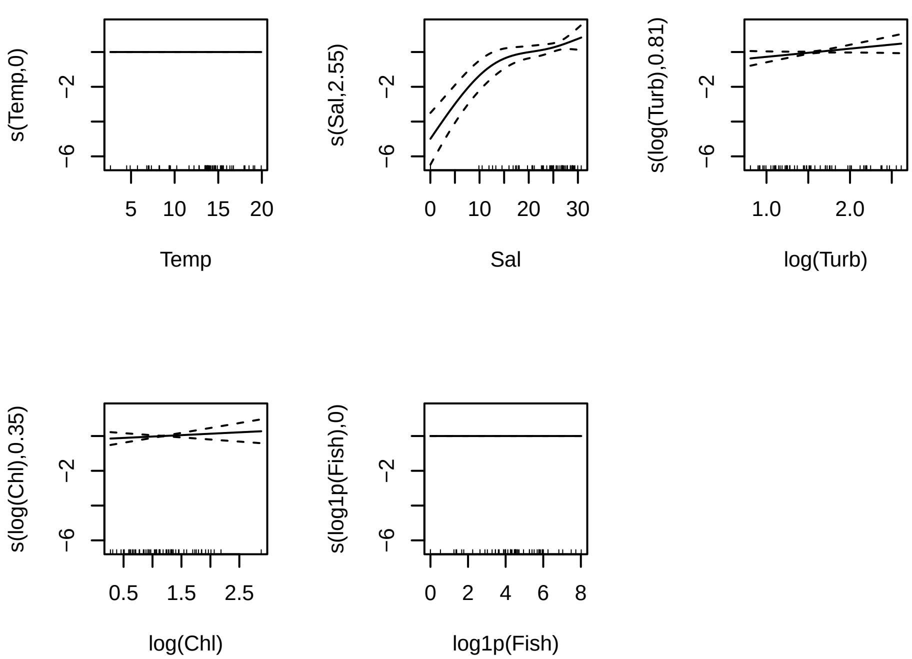
```
Seas_emms <- emmeans(mod, ~Season, type = 'response',
                     data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Seas_emms)
```




```
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate    SE   df t.ratio p.value
#>  Spring - Summer    0.840 0.398 48.3   2.110  0.1203
#>  Spring - Fall      0.891 0.405 48.3   2.203  0.0972
#>  Summer - Fall      0.051 0.316 48.3   0.162  1.0000
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

Plot GAM

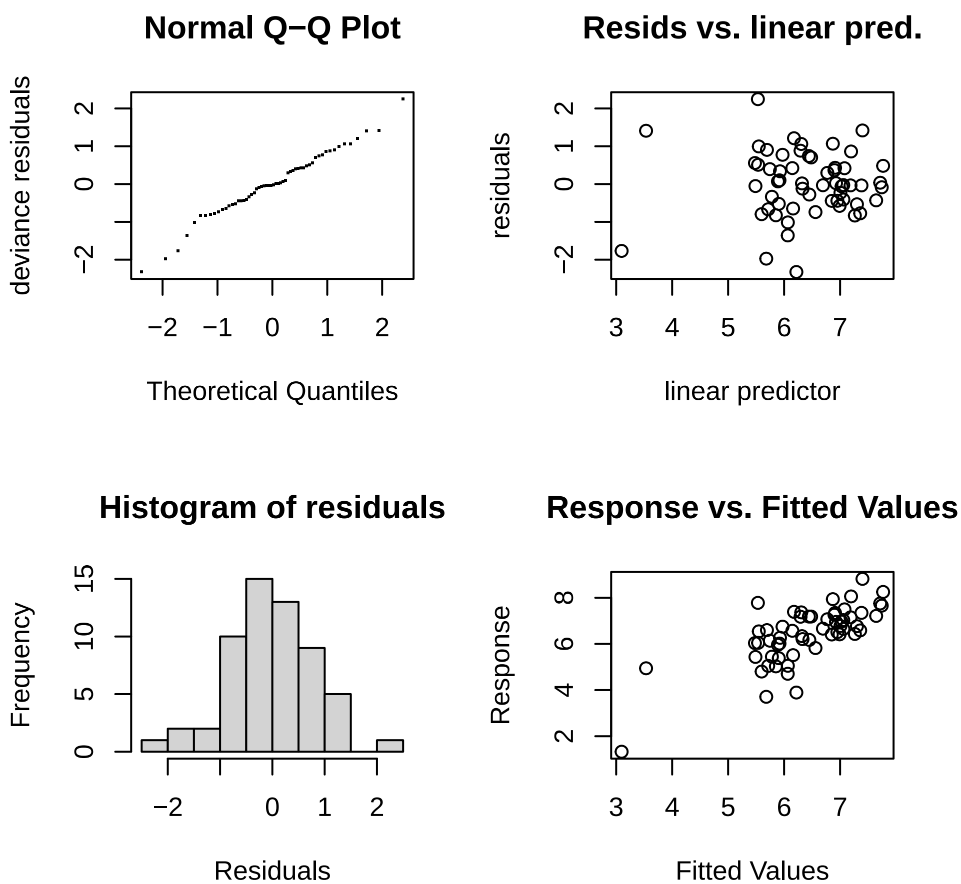
```
oldpar <- par(mfrow = c(2,3))
plot(mod$gam)
par(oldpar)
```



Salinity is the only smooter term that shows as significant.

Model Diagnostics

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




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)        3.00e+00 8.52e-09    1.06    0.59
#> s(Sal)         3.00e+00 2.55e+00    1.19    0.90
#> s(log(Turb))   3.00e+00 8.09e-01    0.90    0.21
#> s(log(Chl))    3.00e+00 3.54e-01    1.04    0.57
#> s(log1p(Fish)) 3.00e+00 7.58e-10    1.03    0.56
par(oldpar)
```

Some outliers – again, probably the "washout" samples.

Polychaete

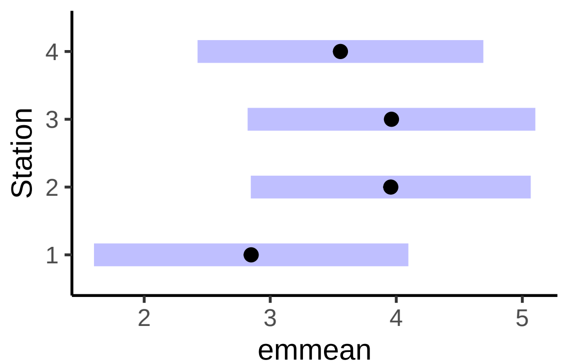
Summary and ANOVA

```
spp =  "Polychaete"
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    5.1189     0.9732   5.260 2.95e-06 ***
#> Station2       1.1075     0.5278   2.099 0.040885 *
#> Station3       1.1131     0.5075   2.193 0.032935 *
#> Station4       0.7082     0.5709   1.241 0.220524
#> SeasonSummer  -4.6422     1.2227  -3.797 0.000395 ***
#> SeasonFall    -3.3737     1.1753  -2.871 0.005973 **
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                     edf Ref.df     F p-value
#> s(Temp)        9.579e-10      3 0.000  0.5767
#> s(Sal)         1.114e-09      3 0.000  0.6269
#> s(log(Turb))   3.041e-07      3 0.000  0.2477
#> s(log(Chl))    1.085e+00      3 2.733  0.0106 *
#> s(log1p(Fish)) 4.649e-01      3 0.290  0.2237
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.354
#>   Scale est. = 1.4856    n = 58
cat('\n')
anova(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric Terms:
```

```
#>         df     F p-value
#> Station  3 2.030 0.12146
#> Season   2 7.582 0.00132
#>
#> Approximate significance of smooth terms:
#>                    edf    Ref.df     F p-value
#> s(Temp)        9.579e-10 3.000e+00 0.000  0.5767
#> s(Sal)         1.114e-09 3.000e+00 0.000  0.6269
#> s(log(Turb))   3.041e-07 3.000e+00 0.000  0.2477
#> s(log(Chl))    1.085e+00 3.000e+00 2.733  0.0106
#> s(log1p(Fish)) 4.649e-01 3.000e+00 0.290  0.2237
```

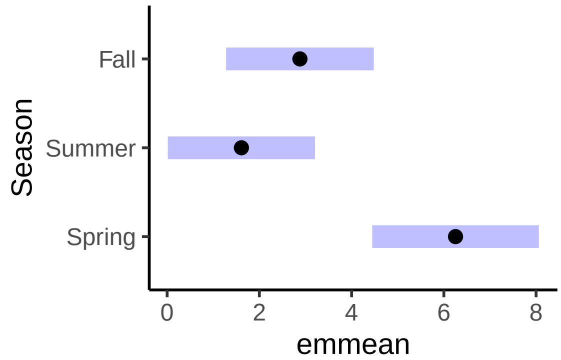
Comparison of Station and Season (Station not significnat)

```
Sta_emms <- emmeans(mod, ~Station, type = 'response',
                    data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Sta_emms)
```




```
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast             estimate    SE   df t.ratio p.value
#>  Station1 - Station2 -1.10754 0.528 50.5  -2.099  0.2453
#>  Station1 - Station3 -1.11311 0.508 50.5  -2.193  0.1976
#>  Station1 - Station4 -0.70815 0.571 50.5  -1.241  1.0000
#>  Station2 - Station3 -0.00557 0.476 50.5  -0.012  1.0000
#>  Station2 - Station4  0.39938 0.487 50.5   0.821  1.0000
#>  Station3 - Station4  0.40496 0.500 50.5   0.809  1.0000
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

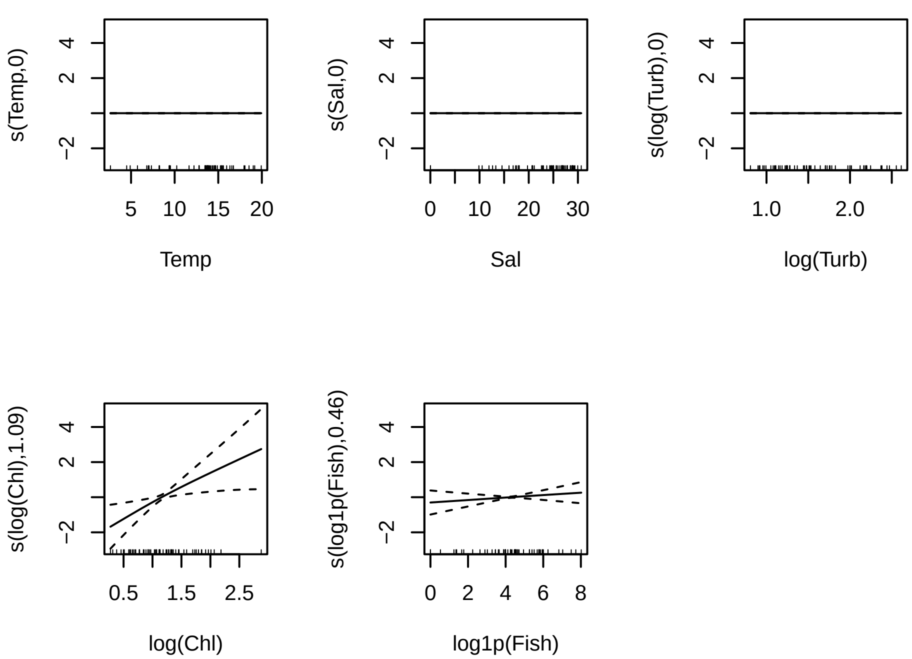
```
Seas_emms <- emmeans(mod, ~Season, type = 'response',
                     data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Seas_emms)
```




```
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate   SE   df t.ratio p.value
#>  Spring - Summer     4.64 1.22 50.5   3.797  0.0012
#>  Spring - Fall       3.37 1.18 50.5   2.871  0.0179
#>  Summer - Fall      -1.27 1.11 50.5  -1.138  0.7816
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

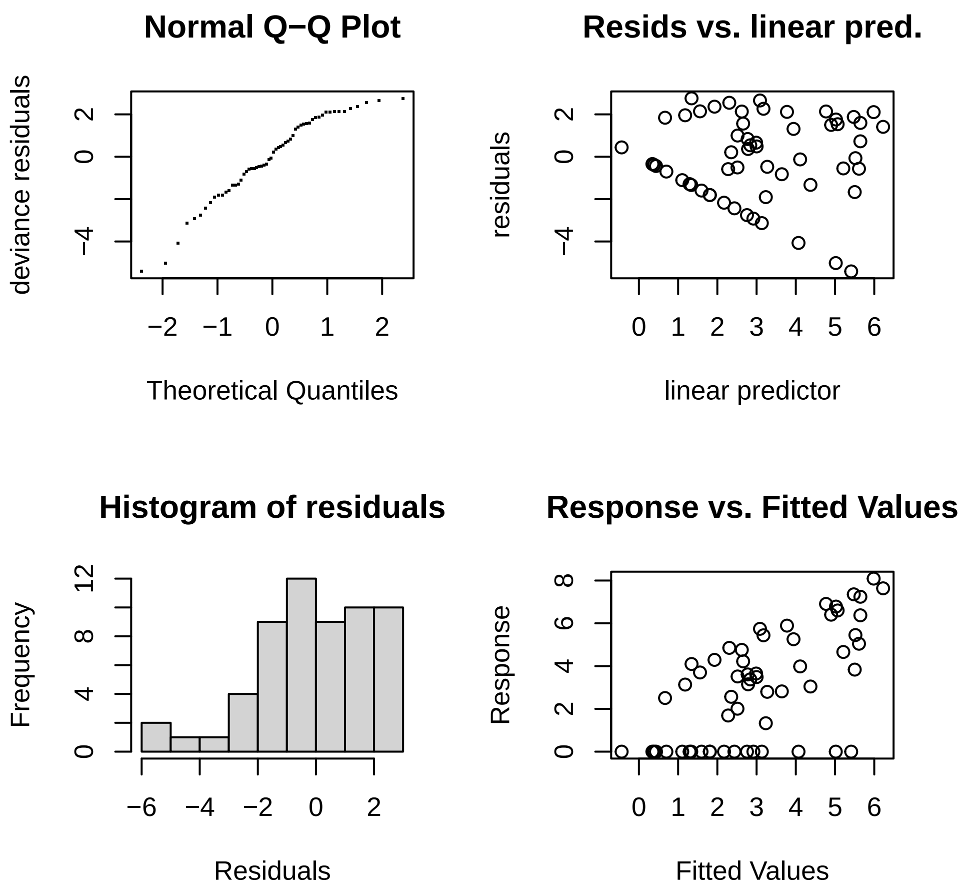
Plot GAM

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



Model Diagnostics

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




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)         3.00e+00 9.58e-10    0.93    0.24
#> s(Sal)          3.00e+00 1.11e-09    1.01    0.47
#> s(log(Turb))    3.00e+00 3.04e-07    0.77    0.04 *
#> s(log(Chl))     3.00e+00 1.09e+00    0.93    0.27
#> s(log1p(Fish)) 3.00e+00 4.65e-01    1.19    0.92
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(oldpar)
```

Pseudocal

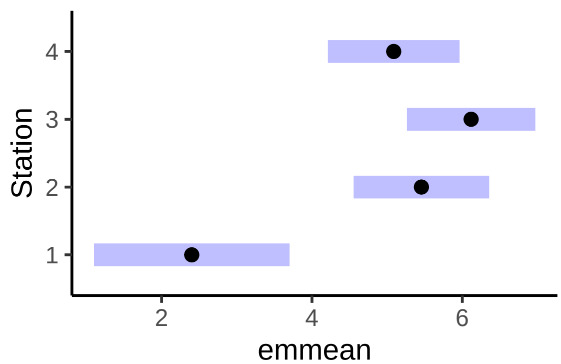
Summary and ANOVA

```
spp =  "Pseudocal"
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    4.0433     0.8249   4.901 1.08e-05 ***
#> Station2       3.0541     0.7683   3.975 0.000230 ***
#> Station3       3.7154     0.6887   5.395 1.96e-06 ***
#> Station4       2.6859     0.7366   3.646 0.000643 ***
#> SeasonSummer  -2.1727     1.0411  -2.087 0.042095 *
#> SeasonFall    -3.8028     1.0327  -3.683 0.000575 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                      edf Ref.df     F p-value
#> s(Temp)        1.947e+00      3 5.655 0.00125 **
#> s(Sal)         9.475e-01      3 2.221 0.01594 *
#> s(log(Turb))   6.567e-09      3 0.000 0.67171
#> s(log(Chl))    1.006e-08      3 0.000 0.88410
#> s(log1p(Fish)) 1.128e-08      3 0.000 0.94832
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.652
#>   Scale est. = 1.2913    n = 58
cat('\n')
anova(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric Terms:
```

```
#>          df     F  p-value
#> Station  3 10.13 2.64e-05
#> Season   2 10.74 0.000135
#>
#> Approximate significance of smooth terms:
#>                      edf    Ref.df     F p-value
#> s(Temp)        1.947e+00 3.000e+00 5.655 0.00125
#> s(Sal)         9.475e-01 3.000e+00 2.221 0.01594
#> s(log(Turb))   6.567e-09 3.000e+00 0.000 0.67171
#> s(log(Chl))    1.006e-08 3.000e+00 0.000 0.88410
#> s(log1p(Fish)) 1.128e-08 3.000e+00 0.000 0.94832
```

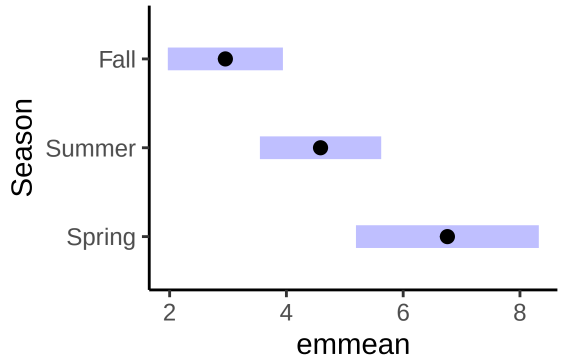
Comparison of Station and Season

```
Sta_emms <- emmeans(mod, ~Station, type = 'response',
                    data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Sta_emms)
```




```
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast            estimate    SE   df t.ratio p.value
#>  Station1 - Station2   -3.054 0.768 49.1  -3.975  0.0014
#>  Station1 - Station3   -3.715 0.689 49.1  -5.395  <.0001
#>  Station1 - Station4   -2.686 0.737 49.1  -3.646  0.0039
#>  Station2 - Station3   -0.661 0.452 49.1  -1.463  0.8997
#>  Station2 - Station4    0.368 0.450 49.1   0.818  1.0000
#>  Station3 - Station4    1.029 0.453 49.1   2.275  0.1640
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

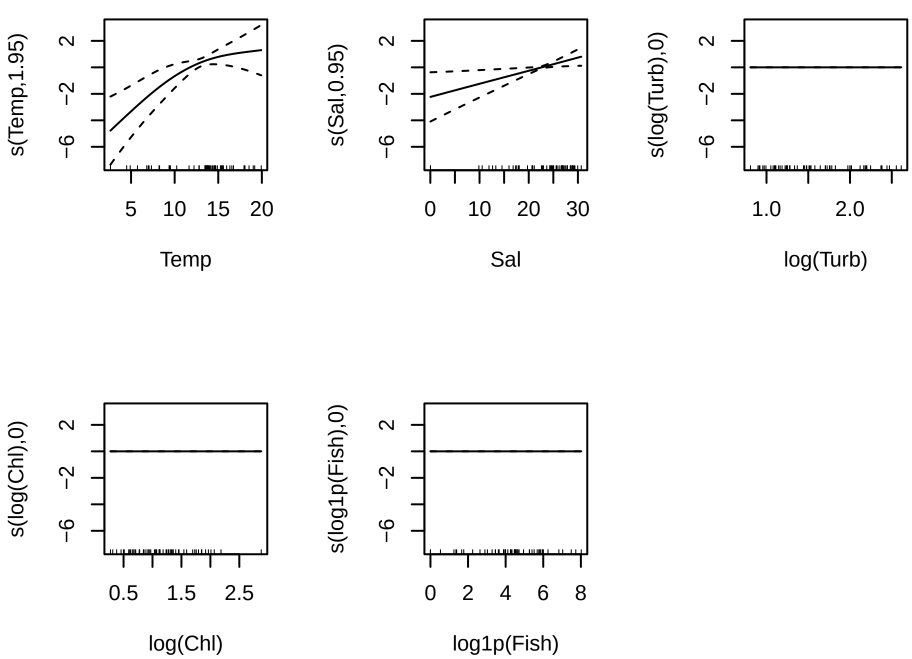
```
Seas_emms <- emmeans(mod, ~Season, type = 'response',
                     data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Seas_emms)
```




```
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate    SE   df t.ratio p.value
#>  Spring - Summer     2.17 1.041 49.1   2.087  0.1263
#>  Spring - Fall       3.80 1.033 49.1   3.683  0.0017
#>  Summer - Fall       1.63 0.464 49.1   3.515  0.0029
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

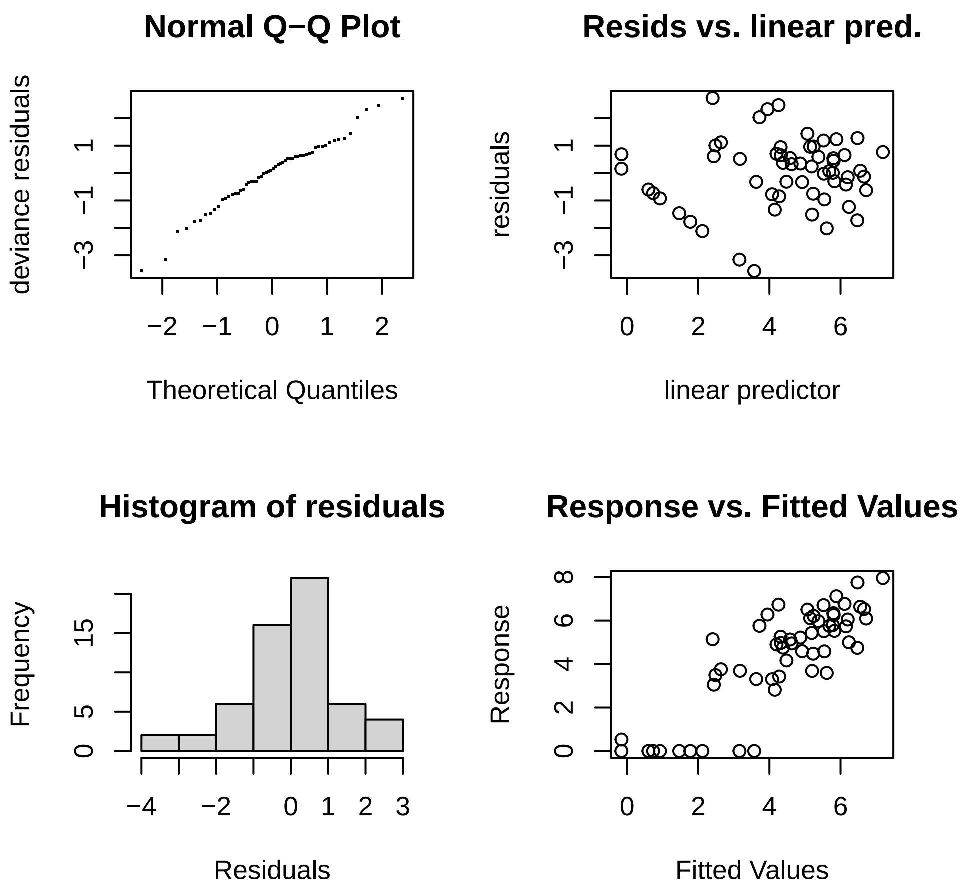
Plot GAM

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



Model Diagnostics

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




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)         3.00e+00 1.95e+00    0.79   0.055 .
#> s(Sal)          3.00e+00 9.48e-01    0.87   0.135
#> s(log(Turb))    3.00e+00 6.57e-09    0.98   0.300
#> s(log(Chl))     3.00e+00 1.01e-08    1.32   0.995
#> s(log1p(Fish)) 3.00e+00 1.13e-08    1.26   0.975
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(oldpar)
```

Temora

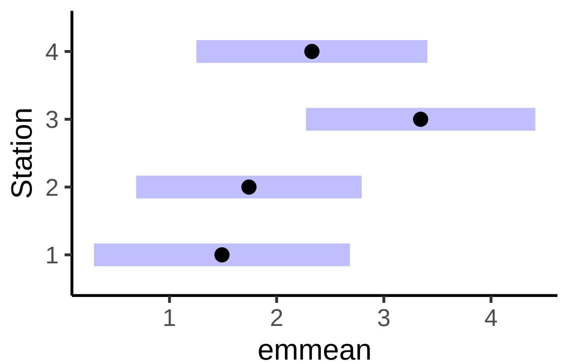
Summary and ANOVA

```
spp =  "Temora"
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    1.5639     0.7733   2.022  0.04838 *
#> Station2       0.2519     0.6882   0.366  0.71588
#> Station3       1.8536     0.6736   2.752  0.00819 **
#> Station4       0.8392     0.7227   1.161  0.25102
#> SeasonSummer  -0.7134     0.7446  -0.958  0.34255
#> SeasonFall    -0.2892     0.6722  -0.430  0.66883
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>                     edf Ref.df     F p-value
#> s(Temp)        1.562e-09      3 0.000 0.67355
#> s(Sal)         1.255e-08      3 0.000 0.40191
#> s(log(Turb))   1.878e-08      3 0.000 0.38103
#> s(log(Chl))    9.754e-01      3 2.997 0.00651 **
#> s(log1p(Fish)) 2.906e-09      3 0.000 0.84705
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.211
#>   Scale est. = 2.8181    n = 58
cat('\n')
anova(mod$gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Station + Season + s(Temp, bs = "ts", k = 4) +
#>     s(Sal, bs = "ts", k = 4) + s(log(Turb), bs = "ts", k = 4) +
#>     s(log(Chl), bs = "ts", k = 4) + s(log1p(Fish), bs = "ts",
#>     k = 4)
#>
#> Parametric Terms:
```

```
#>         df     F p-value
#> Station  3 3.136  0.0333
#> Season   2 0.501  0.6089
#>
#> Approximate significance of smooth terms:
#>                    edf    Ref.df     F p-value
#> s(Temp)        1.562e-09 3.000e+00 0.000 0.67355
#> s(Sal)         1.255e-08 3.000e+00 0.000 0.40191
#> s(log(Turb))   1.878e-08 3.000e+00 0.000 0.38103
#> s(log(Chl))    9.754e-01 3.000e+00 2.997 0.00651
#> s(log1p(Fish)) 2.906e-09 3.000e+00 0.000 0.84705
```

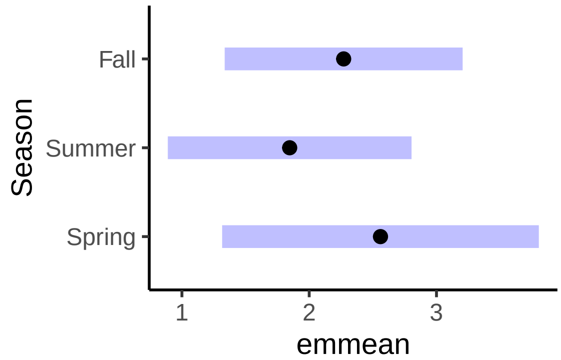
Comparison of Station and Season 9Season not significant)

```
Sta_emms <- emmeans(mod, ~Station, type = 'response',
                    data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Sta_emms)
```




```
pairs(Sta_emms, adjust ='bonferroni')
#>  contrast            estimate    SE df t.ratio p.value
#>  Station1 - Station2   -0.252 0.688 51  -0.366  1.0000
#>  Station1 - Station3   -1.854 0.674 51  -2.752  0.0491
#>  Station1 - Station4   -0.839 0.723 51  -1.161  1.0000
#>  Station2 - Station3   -1.602 0.651 51  -2.462  0.1034
#>  Station2 - Station4   -0.587 0.664 51  -0.885  1.0000
#>  Station3 - Station4    1.014 0.674 51   1.506  0.8295
#>
#> Results are averaged over the levels of: Season
#> P value adjustment: bonferroni method for 6 tests
```

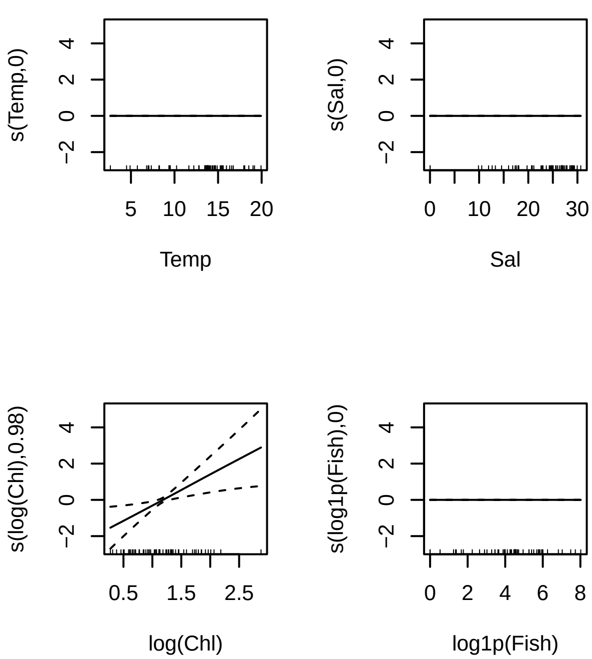
```
Seas_emms <- emmeans(mod, ~Season, type = 'response',
                    data = spp_analysis$data[spp_data$Species == spp][[1]])
plot(Seas_emms)
```




```
pairs(Seas_emms, adjust ='bonferroni')
#>  contrast        estimate    SE df t.ratio p.value
#>  Spring - Summer    0.713 0.745 51   0.958  1.0000
#>  Spring - Fall      0.289 0.672 51   0.430  1.0000
#>  Summer - Fall     -0.424 0.577 51  -0.735  1.0000
#>
#> Results are averaged over the levels of: Station
#> P value adjustment: bonferroni method for 3 tests
```

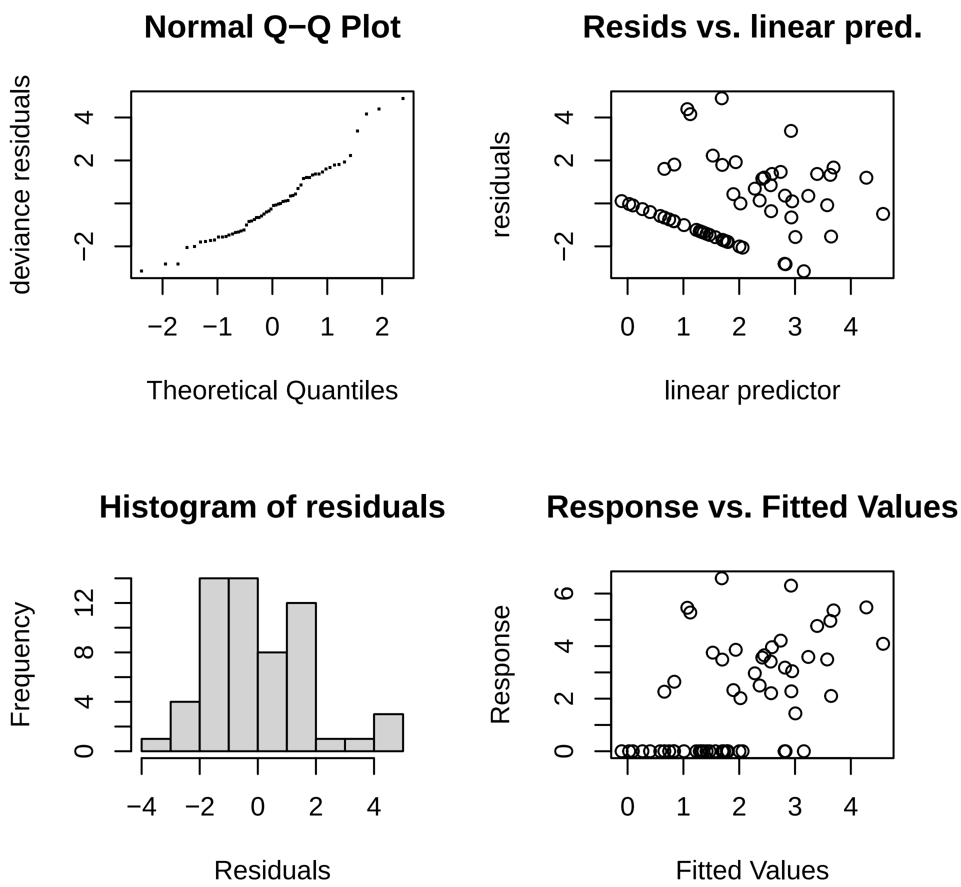
Plot GAM

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



Model Diagnostics

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




```
#>
#> 'gamm' based fit - care required with interpretation.
#> Checks based on working residuals may be misleading.
#> 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(Temp)         3.00e+00 1.56e-09    1.08    0.68
#> s(Sal)          3.00e+00 1.25e-08    1.13    0.82
#> s(log(Turb))    3.00e+00 1.88e-08    1.10    0.71
#> s(log(Chl))     3.00e+00 9.75e-01    0.97    0.34
#> s(log1p(Fish)) 3.00e+00 2.91e-09    1.12    0.74
par(oldpar)
```