Tumor Response GAM models with comments

Ariel Mundo

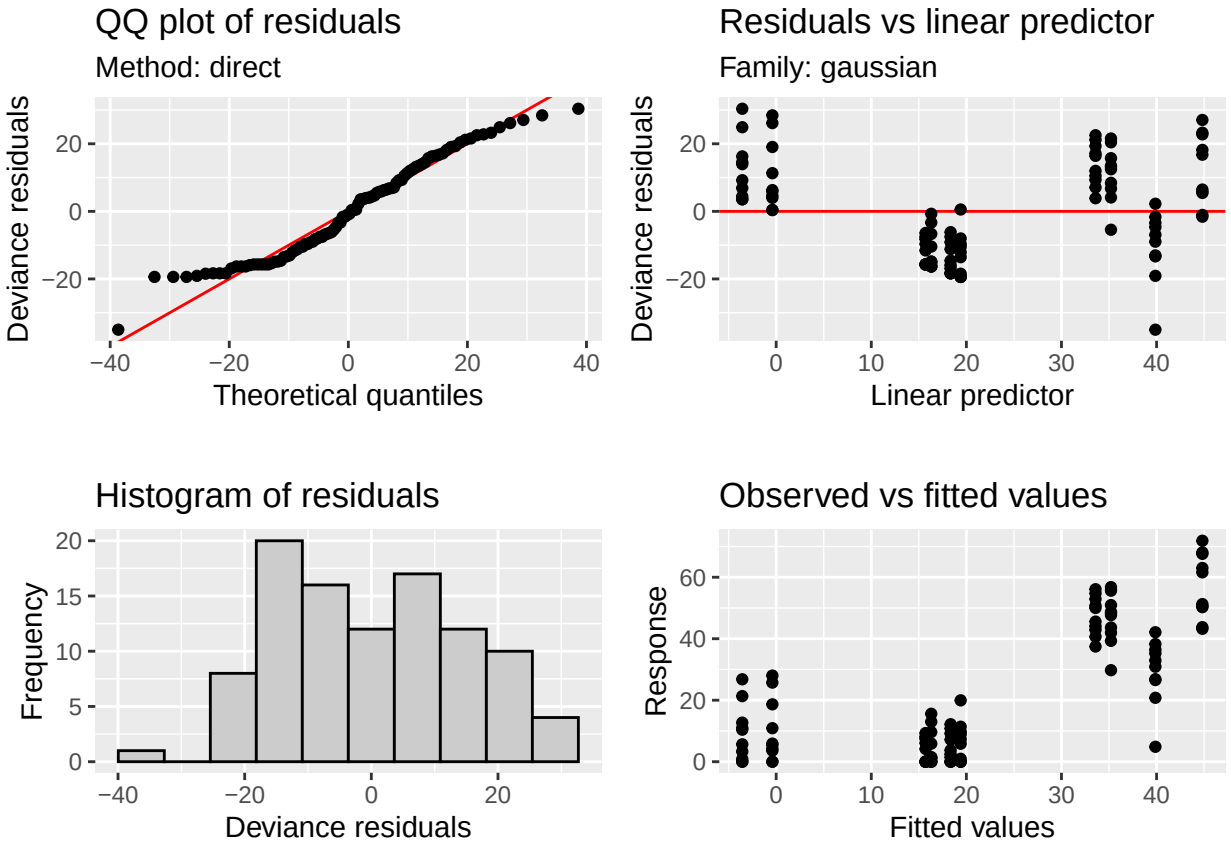
5/2/2021

number of observations – currently using 5 use measurement error of sd as 10 (rough estimate)

Data are simulated from Fig. 3 (C) of Vishwanath et. al: "Using optical spectroscopy to longitudinally monitor physiological changes within solid tumors." Neoplasia 11.9 (2009): 889-900. https://www.sciencedirect.com/science/article/pii/S1476558609800420

From the figure the maximum standard deviation for the measurements is approximately 10%.

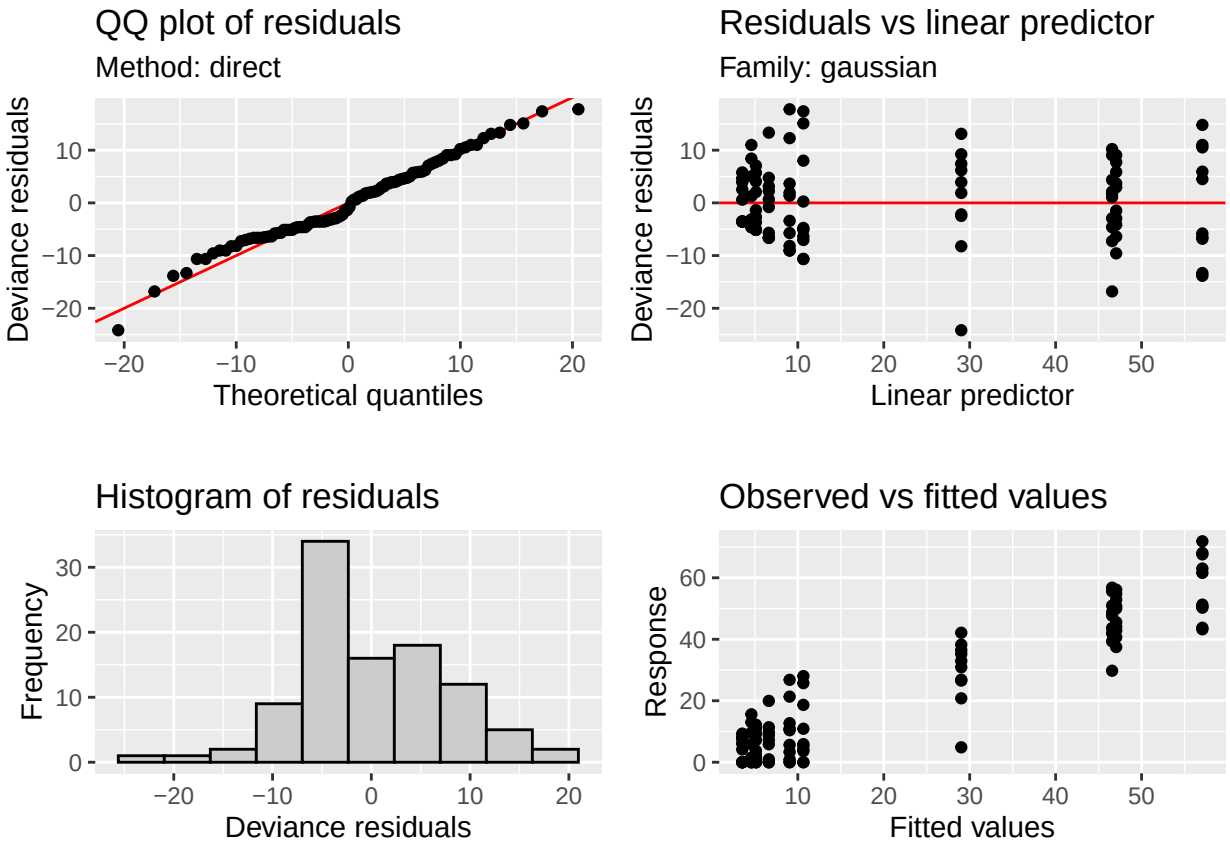
```
mod_simple  <- gam(StO2_sim ~ s(Day,by=Group,k=5),data  = dat_sim)
appraise(mod_simple)
```



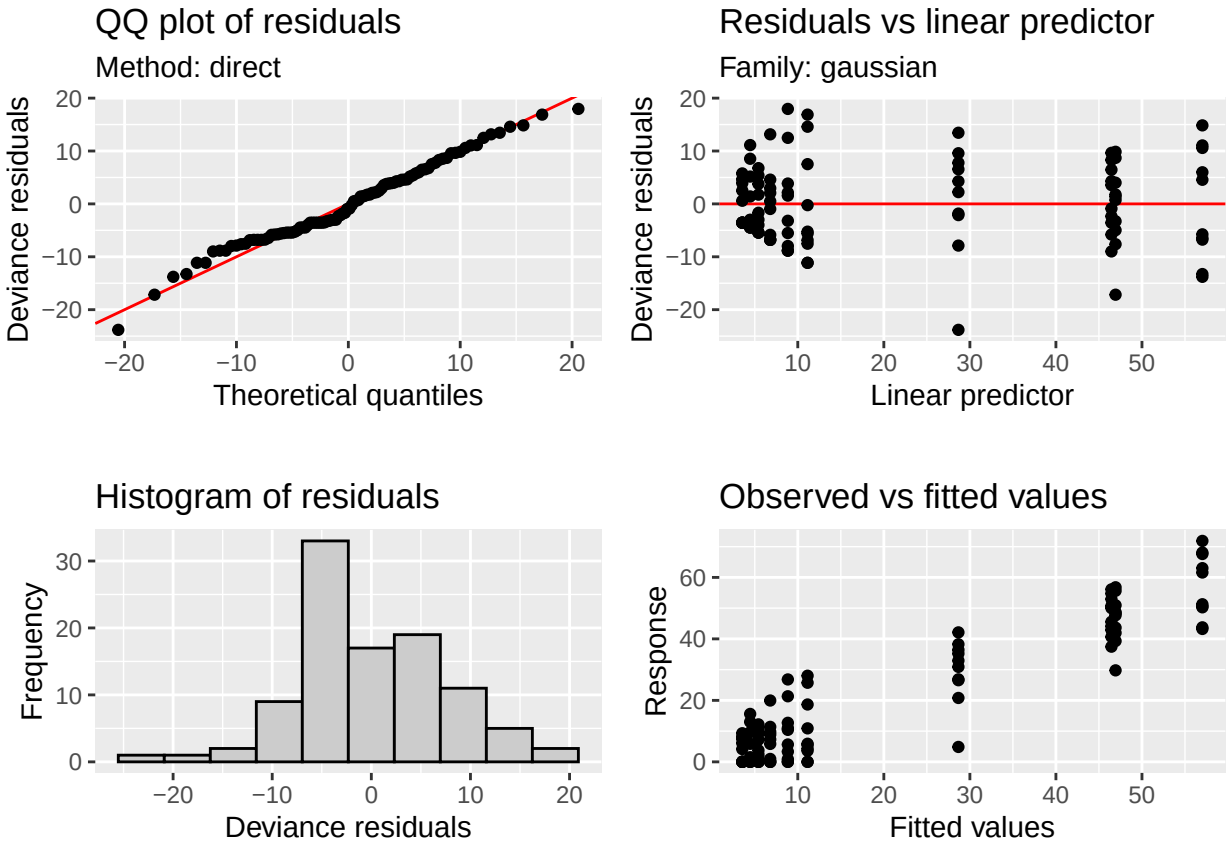

```
#SIMPLEST MODEL: A SMOOTH FOR TIME AS THE COVARIATE, WHICH CAN VARY BY GROUP BUT... APPRAISING THE MODE

#Model with interaction using tensor: there is some different specification that does not allow to plot
```

```
mod_int1 <- gam(StO2_sim ~ Group+te(Day,by=Group, k = 5),
                data  = dat_sim)
appraise(mod_int1)
```




```
#Model with interaction using s
mod_int2 <- gam(StO2_sim ~ Group+s(Day, by = Group, k = 5),method='REML', data  = dat_sim)
appraise(mod_int2)
```



The simplest model attempts to explain how StO_2 changes over time, and gives each treatment group a separate smooth. However, from the diagnostics, it is clear that there are some issues because the QQ plot, the residuals vs linear predictor and the histogram show that the assumptions of normality are not met.

Next, we build a model that accounts for the *interaction* of time and treatment. In this sense, mod_int1 and mod_int2 are doing essentially the same thing: We have a smooth to explain how StO_2 changes per group, but we also account separately for the "main effect" of treatment. In other words, we are trying to explain the changes in StO_2 by considering how it changes over time, but also disentangling the treatment effect from the response.

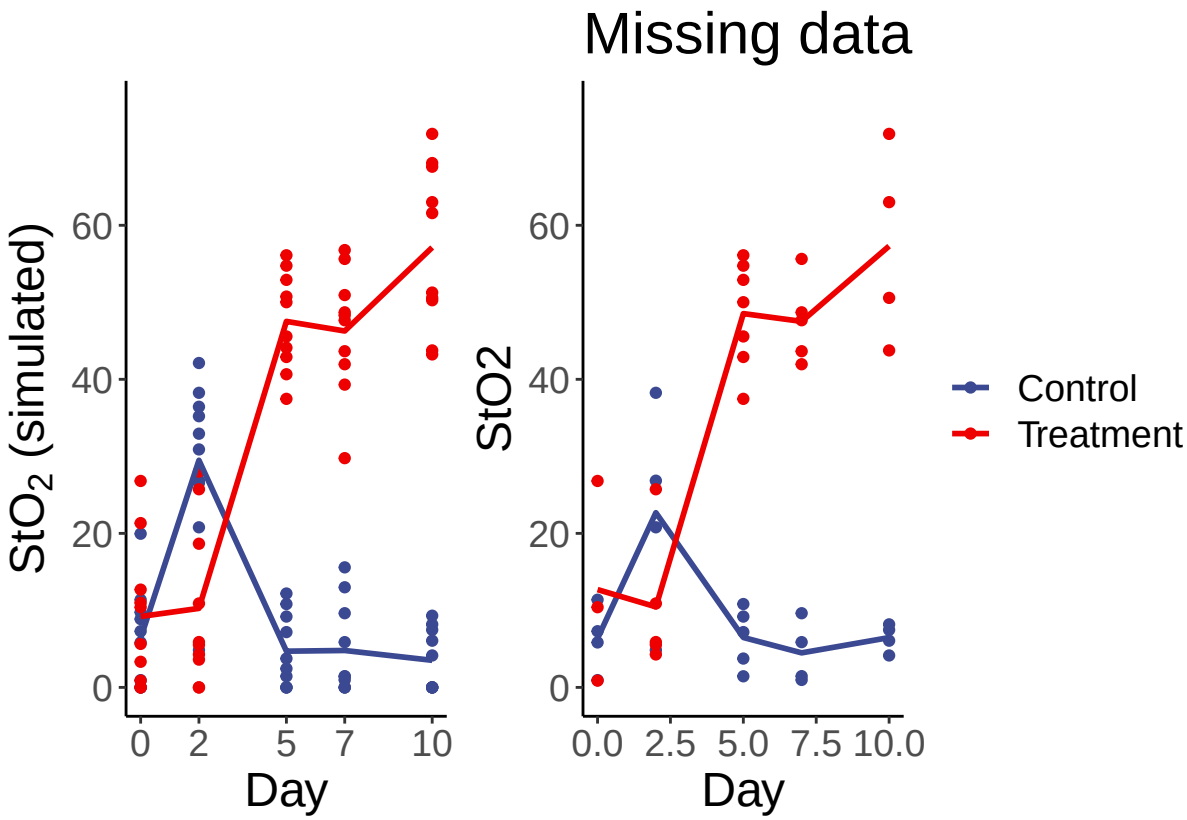
"To avoid confounding problems the smooths are all subject to sum to zero constraints, which usually means that the main effect of g should also be included in the model specification. For example, g + te(z,x,by=g)(pg.327)" (From Simon Wood's book). This gives a good explanation on why mod_int1 and mod_int2 work.

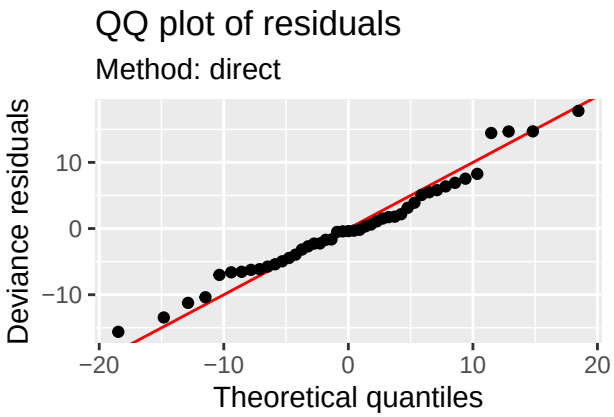
The idea for this came from https://stats.stackexchange.com/questions/486118/model-building-in-generalized-additive-mixed-models-gamms

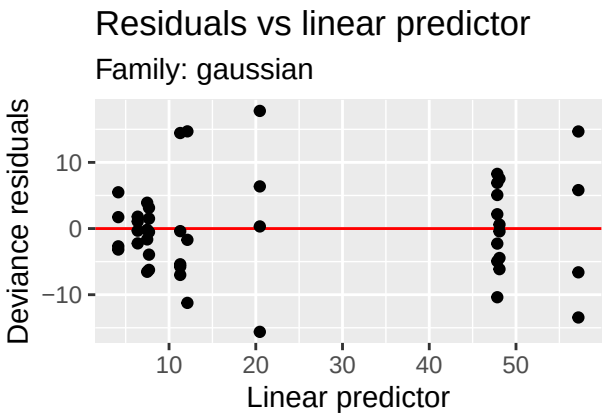
GAMs do not require equally-spaced observations and can accommodate missing observations as well. In a case where 40% of the observations are missing:

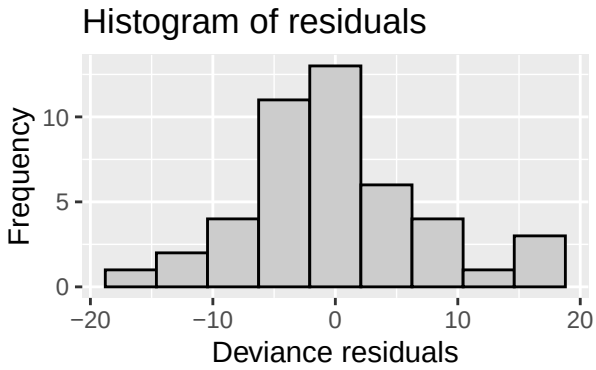
```
## # A tibble: 10 x 3
## # Groups:   Day, Group [10]
##       Day Group         n
##     <dbl> <fct>     <int>
##  1     0 Control       4
##  2     0 Treatment     3
##  3     2 Control       4
```

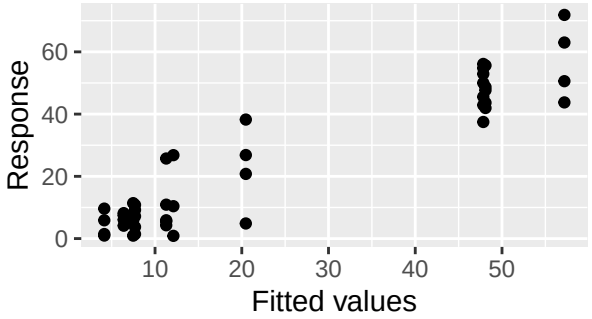
```
##  4     2 Treatment     5
##  5     5 Control       5
##  6     5 Treatment     7
##  7     7 Control       4
##  8     7 Treatment     5
##  9    10 Control       4
## 10    10 Treatment     4
```

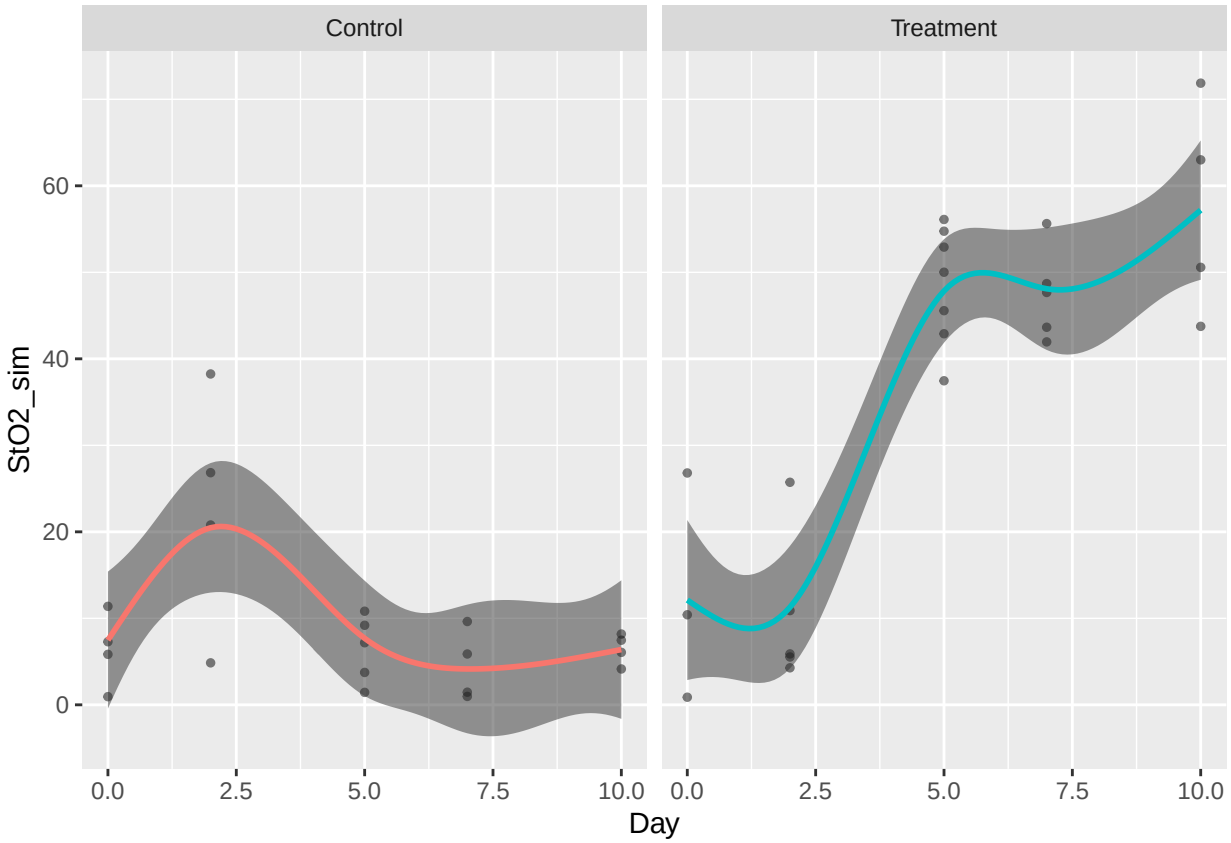


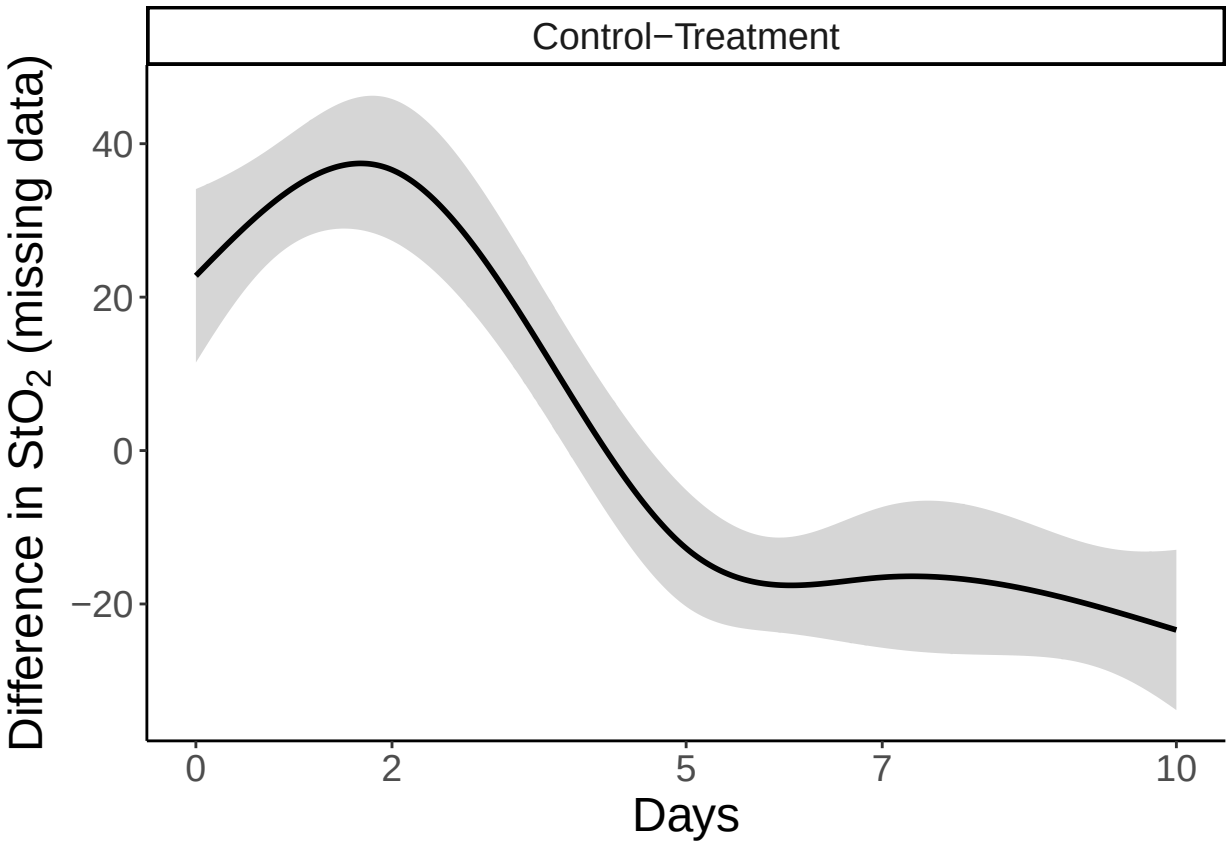

QQ plot of residuals
Method: direct
Deviance residuals
10
0
−10
−20
−10
0
10
20
Theoretical quantiles

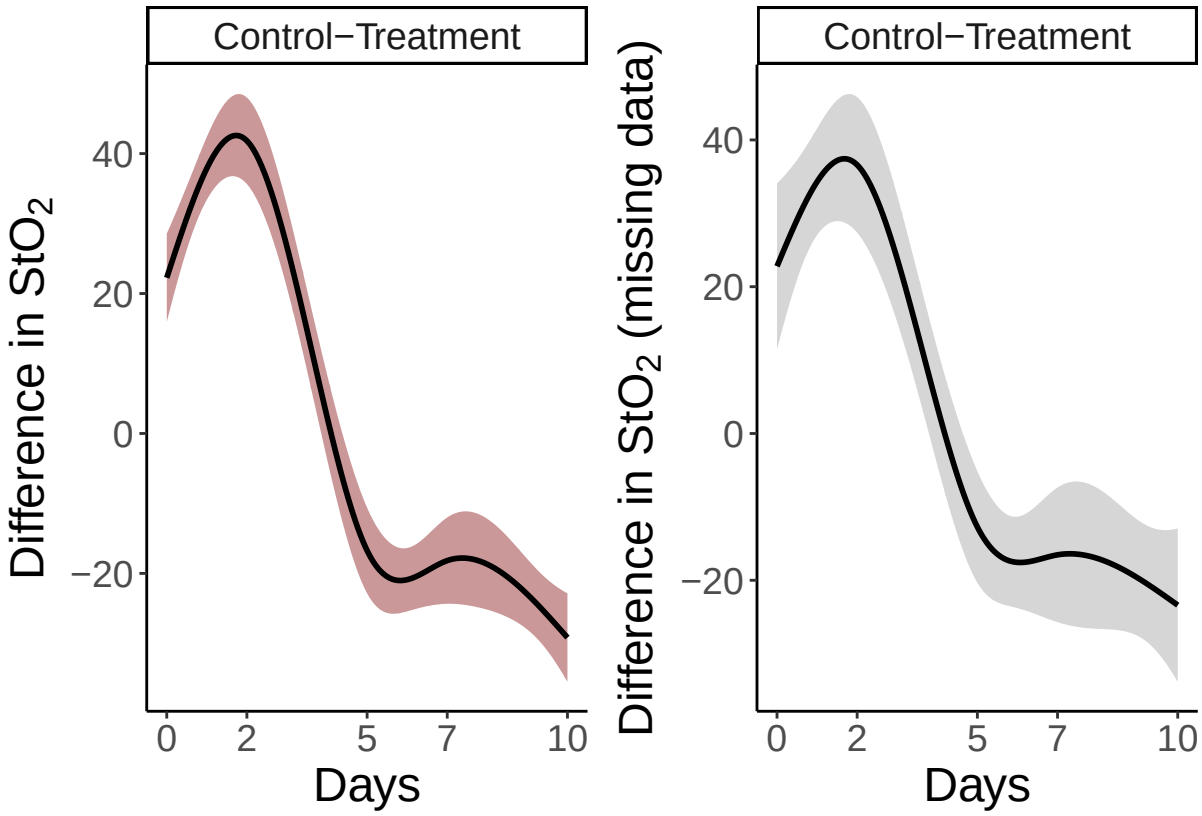

Residuals vs linear predictor
Family: gaussian
Deviance residuals
10
0
−10
10
20
30
40
50
Linear predictor


Histogram of residuals
Frequency
10
5
0
−20
−10
0
10
20
Deviance residuals


Observed vs fitted values
Response
60
40
20
0
10
20
30
40
50
Fitted values


Control
Treatment
StO2_sim
60
40
20
0
0.0
2.5
5.0
7.5
10.0
Day


Control−Treatment
Difference in StO2 (missing data)
40
20
0
−20
0
2
5
7
10
Days



It is reassuring to see that we get the same diagnostics by either having a dataframe with missing observations as NA or a dataframe where the missing observations are not present at all.

So two questions remains after this: -How do we want to present the pairwise smooth confidence interval comparison? The GAMs are referred as "empirical Bayesian", but I could not find a straightforward reference that had information in detail of what the pairwise comparisons can tell at certain timepoints. Sure, they are different, but what else?

-Do we want to address prediction? I recently saw a webinar from Gavin Simpson regarding this, wondered if it would be feasible to include in a concise manner.