GAM fits for hirsutism data

Group 1: Pariente Antonio, Bosch Guillem, Ebner Lena

13/12/2023

Hirsutism dataset

`Hirsutism` is the excessive hairiness on women in those parts of the body where terminal hair does not normally occur or is minimal -for example, a beard or chest hair. It refers to a male pattern of body hair (androgenic hair) and it is therefore primarily of cosmetic and psychological concern. Hirsutism is a symptom rather than a disease and may be a sign of a more serious medical condition, especially if it develops well after puberty.

The amount and location of the hair is measured by a Ferriman-Gallwey score. The original method used 11 body areas to assess hair growth, but was decreased to 9 body areas in the modified method: Upper lip, Chin, Chest, Upper back, Lower back, Upper abdomen, Lower abdomen, Upper arms, Thighs, Forearms (deleted in the modified method) and Legs (deleted in the modified method). In the modified method, hair growth is rated from 0 (no growth of terminal hair) to 4 (extensive hair growth) in each of the nine locations. A patient's score may therefore range from a minimum score of 0 to a maximum score of 36.

A clinical trial was conducted to evaluate the effectiveness of an antiandrogen combined with an oral contraceptive in reducing hirsutism for 12 consecutive months. It is known that contraceptives have positive effects on reduction of hirsutism. The degree of hirsutism is measured by the modified Ferriman-Gallwey scale. Patients were randomized into 4 treatment levels: levels 0 (only contraceptive), 1, 2, and 3 of the antiandrogen in the study (always in combination with the contraceptive). The clinical trial was double-blind.

The data set `hirsutism.dat` contains artificial values of measures corresponding to some patients in this study. The variables are the following:

- `Treatment`, with values 0, 1, 2 or 3.
- `FGm0`, it indicates the baseline hirsutism level at the randomization moment (the beginning of the clinical trial). Only women with baseline FG values grater than 15 where recruited.
- `FGm3`, FG value at 3 months.
- `FGm6`, FG value at 6 months
- `FGm12`, FG value at 12 months, the end of the trial.
- `SysPres`, baseline systolic blood pressure.
- `DiaPres`, baseline diastolic blood pressure.
- `weight`, baseline weight.
- `height`, baseline height.

(Note: The term "baseline" means that these variables were measured at the beginning of the clinical trial).

GAMs for hirsutism data

Fit several GAM models (including semiparametric models) explaining `FGm12` as a function of the variables that were measured at the beginning of the clinical trial (including `FGm0`, but NOT `FGm3` or `FGm6`) and `Treatment` (treated as factor, which can by used as value of parameter `by` in function `s()`). Use functions `summary`, `plot`, `vis.gam` and `gam.check` to get an insight into the fitted models. Then use function `anova` to select among them the model (or models) that you think is (are) the most appropriate.

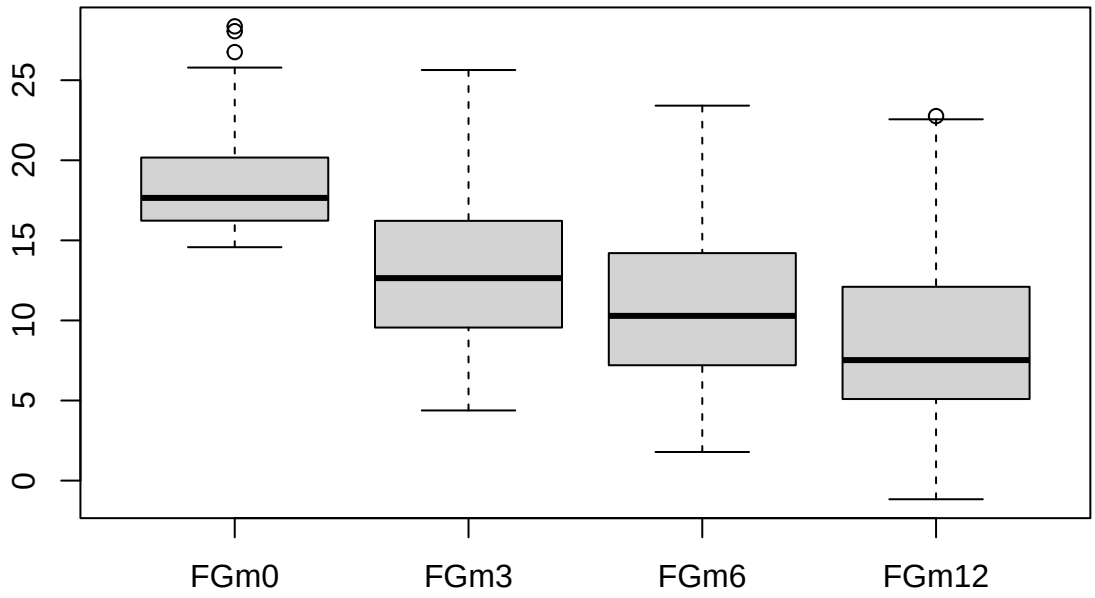
Load Data

```
# load data
hirs <- read.table("hirsutism.dat",header=T, sep="\t",fill=TRUE)
hirs$Treatment = as.factor(hirs$Treatment)
summary(hirs)
```

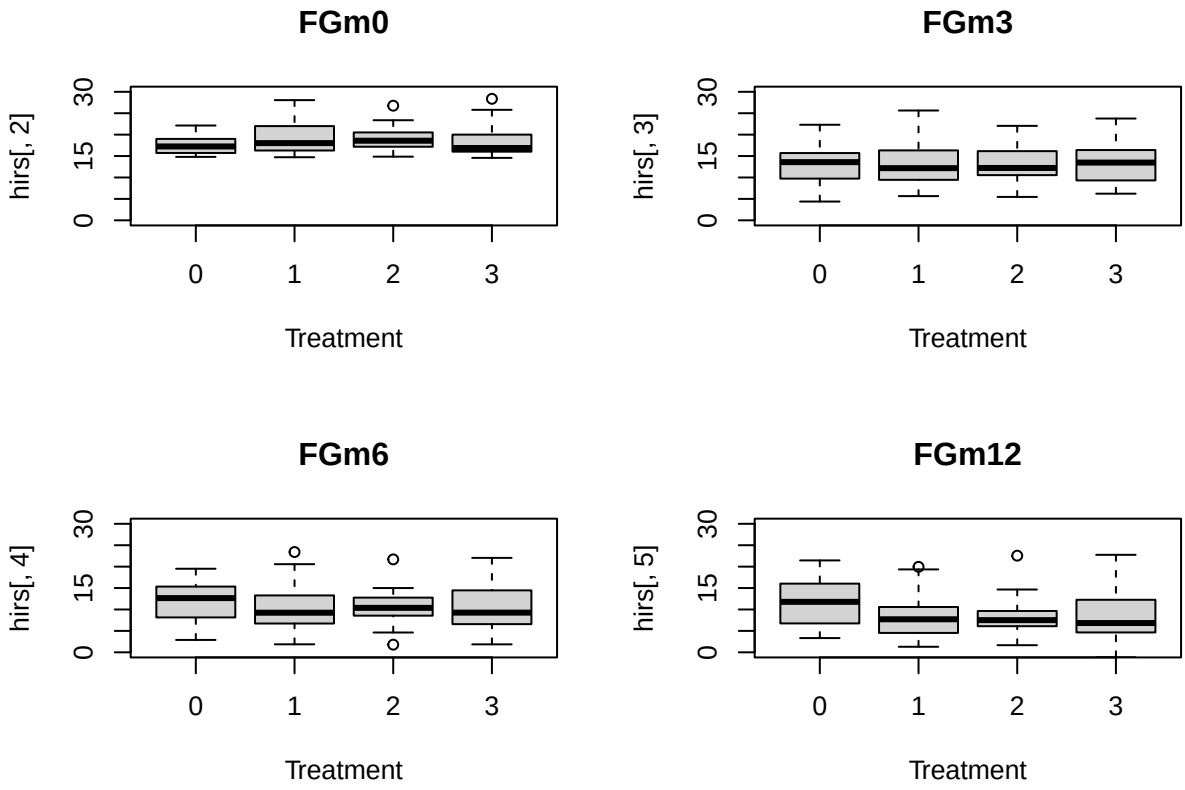
```
##  Treatment      FGm0            FGm3             FGm6            FGm12
##  0:23      Min.   :14.57   Min.   : 4.381   Min.   : 1.786   Min.   :-1.163
##  1:26      1st Qu.:16.23   1st Qu.: 9.557   1st Qu.: 7.202   1st Qu.: 5.093
##  2:24      Median :17.65   Median :12.643   Median :10.286   Median : 7.524
##  3:26      Mean   :18.57   Mean   :13.084   Mean   :10.853   Mean   : 8.911
##            3rd Qu.:20.17   3rd Qu.:16.219   3rd Qu.:14.204   3rd Qu.:12.101
##            Max.   :28.36   Max.   :25.637   Max.   :23.411   Max.   :22.759
##
##     SysPres         DiaPres          weight           height
##  Min.   : 88.0   Min.   :46.00   Min.   : 41.00   Min.   :1.480
##  1st Qu.:110.0   1st Qu.:65.00   1st Qu.: 57.00   1st Qu.:1.580
##  Median :115.0   Median :70.00   Median : 64.00   Median :1.610
##  Mean   :115.9   Mean   :70.04   Mean   : 68.06   Mean   :1.613
##  3rd Qu.:120.0   3rd Qu.:75.00   3rd Qu.: 74.50   3rd Qu.:1.650
##  Max.   :162.0   Max.   :95.00   Max.   :113.00   Max.   :1.800
##  NA's   :8       NA's   :8       NA's   :8        NA's   :8
```

```
attach(hirs)

boxplot(hirs[,2:5])
```

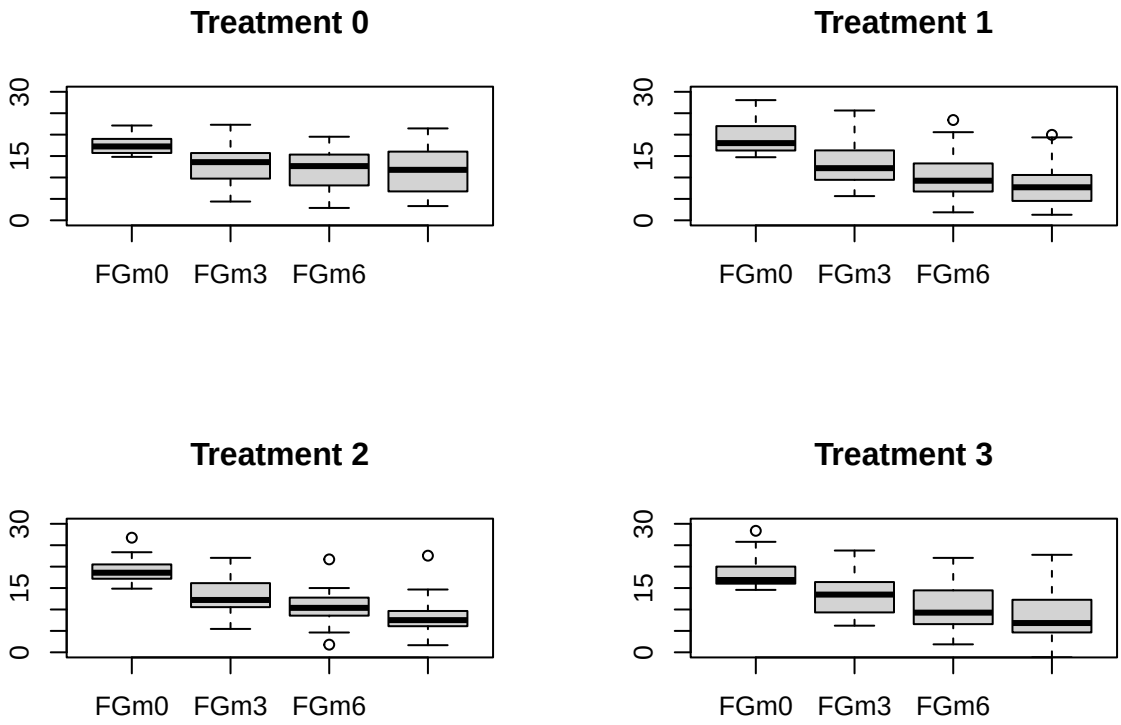



```
par(mfrow=c(2,2))
boxplot(hirs[,2]~Treatment,ylim=c(0,30), main=names(hirs)[2], xlab="Treatment")
boxplot(hirs[,3]~Treatment,ylim=c(0,30), main=names(hirs)[3], xlab="Treatment")
boxplot(hirs[,4]~Treatment,ylim=c(0,30), main=names(hirs)[4], xlab="Treatment")
boxplot(hirs[,5]~Treatment,ylim=c(0,30), main=names(hirs)[5], xlab="Treatment")
```



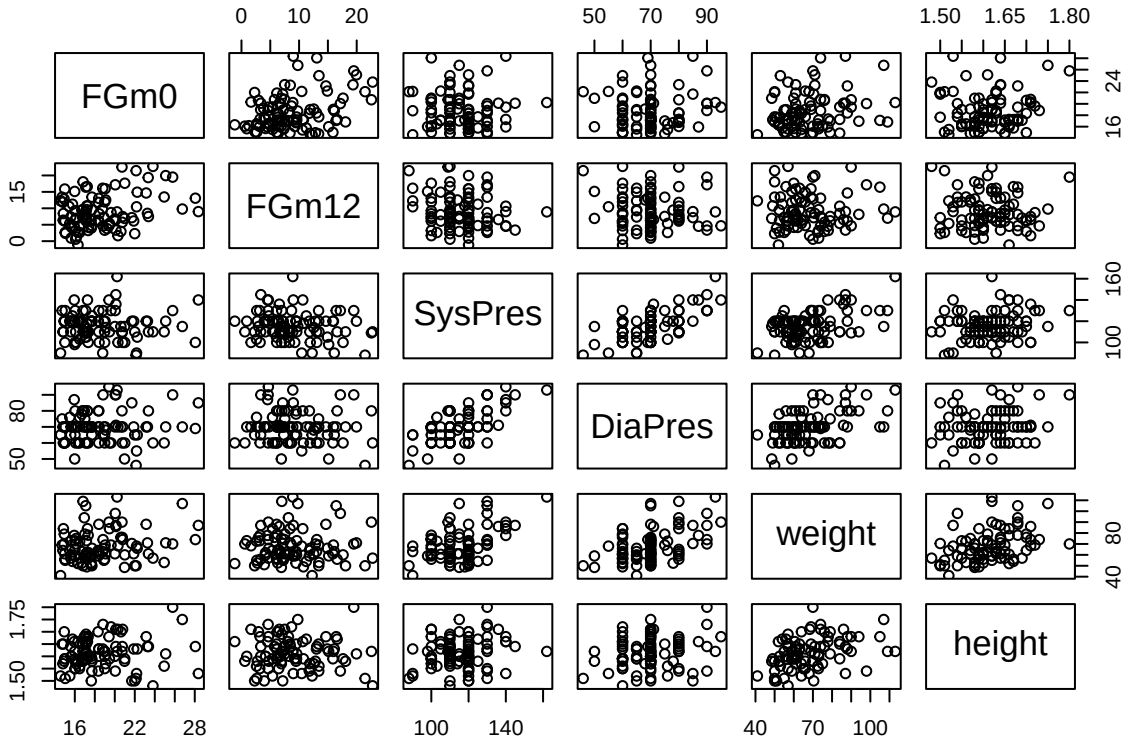

```
par(mfrow=c(1,1))

par(mfrow=c(2,2))
boxplot(hirs[Treatment==0,2:5],ylim=c(0,30), main="Treatment 0")
boxplot(hirs[Treatment==1,2:5],ylim=c(0,30), main="Treatment 1")
boxplot(hirs[Treatment==2,2:5],ylim=c(0,30), main="Treatment 2")
boxplot(hirs[Treatment==3,2:5],ylim=c(0,30), main="Treatment 3")
```




```
par(mfrow=c(1,1))

plot(hirs[,c(-1,-3,-4)])
```




```
library(mgcv)
```

Model 1.0

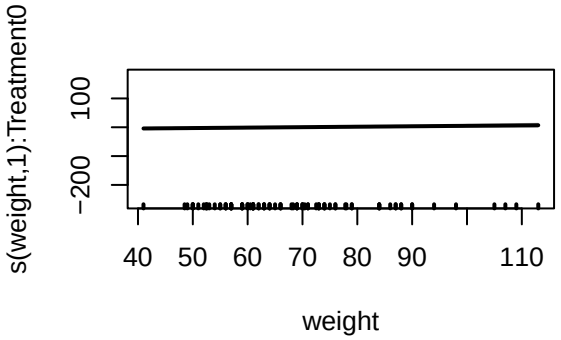
First we fit an additive model with Gaussian family using all the variables measured at the beginning of the clinical trial, smoothed

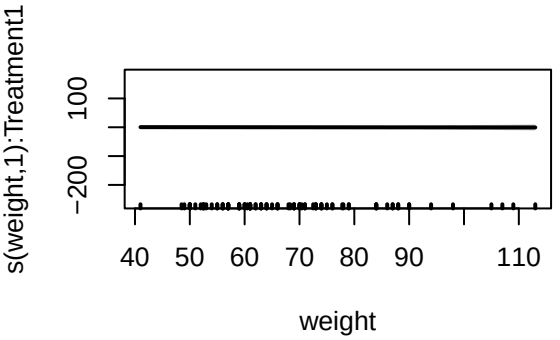
```
gam1.0 <- gam(FGm12 ~ FGm0 + Treatment+s(weight, by=Treatment) + s(height, by=Treatment)+
                s(SysPres, by=Treatment) + s(DiaPres, by=Treatment), data=hirs)
summary(gam1.0)
```

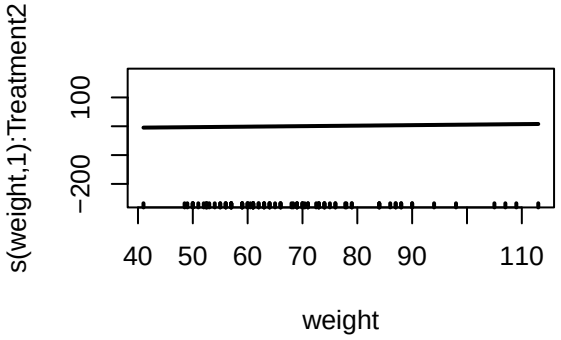
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ FGm0 + Treatment + s(weight, by = Treatment) + s(height,
##     by = Treatment) + s(SysPres, by = Treatment) + s(DiaPres,
##     by = Treatment)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1847     4.5893  -0.040   0.9680
## FGm0          0.8063     0.1740   4.635 2.47e-05 ***
```

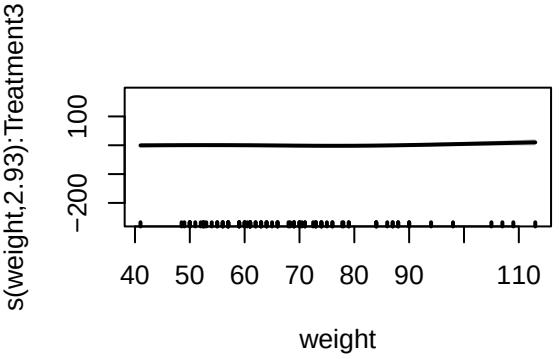
```
## Treatment1   -7.1000     3.4040  -2.086   0.0420 *
## Treatment2   -6.5543     3.3796  -1.939   0.0579 .
## Treatment3   -6.9275     3.3975  -2.039   0.0466 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                          edf Ref.df     F p-value
## s(weight):Treatment0  1.000  1.000 2.499  0.1200
## s(weight):Treatment1  1.000  1.000 0.014  0.9049
## s(weight):Treatment2  1.000  1.000 6.548  0.0134 *
## s(weight):Treatment3  2.930  3.562 1.850  0.1510
## s(height):Treatment0  5.819  6.326 3.673  0.0107 *
## s(height):Treatment1  2.094  2.571 2.106  0.1242
## s(height):Treatment2  2.331  2.868 4.540  0.0104 *
## s(height):Treatment3  2.424  2.989 2.871  0.0386 *
## s(SysPres):Treatment0 1.000  1.000 0.441  0.5098
## s(SysPres):Treatment1 1.000  1.000 0.493  0.4857
## s(SysPres):Treatment2 1.000  1.000 0.806  0.3733
## s(SysPres):Treatment3 1.000  1.000 7.662  0.0078 **
## s(DiaPres):Treatment0 3.576  4.030 0.501  0.7344
## s(DiaPres):Treatment1 3.039  3.699 1.134  0.4183
## s(DiaPres):Treatment2 1.000  1.000 4.501  0.0387 *
## s(DiaPres):Treatment3 4.254  4.851 1.997  0.0796 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.552   Deviance explained = 74.4%
## GCV = 21.586  Scale est. = 12.224    n = 91
```

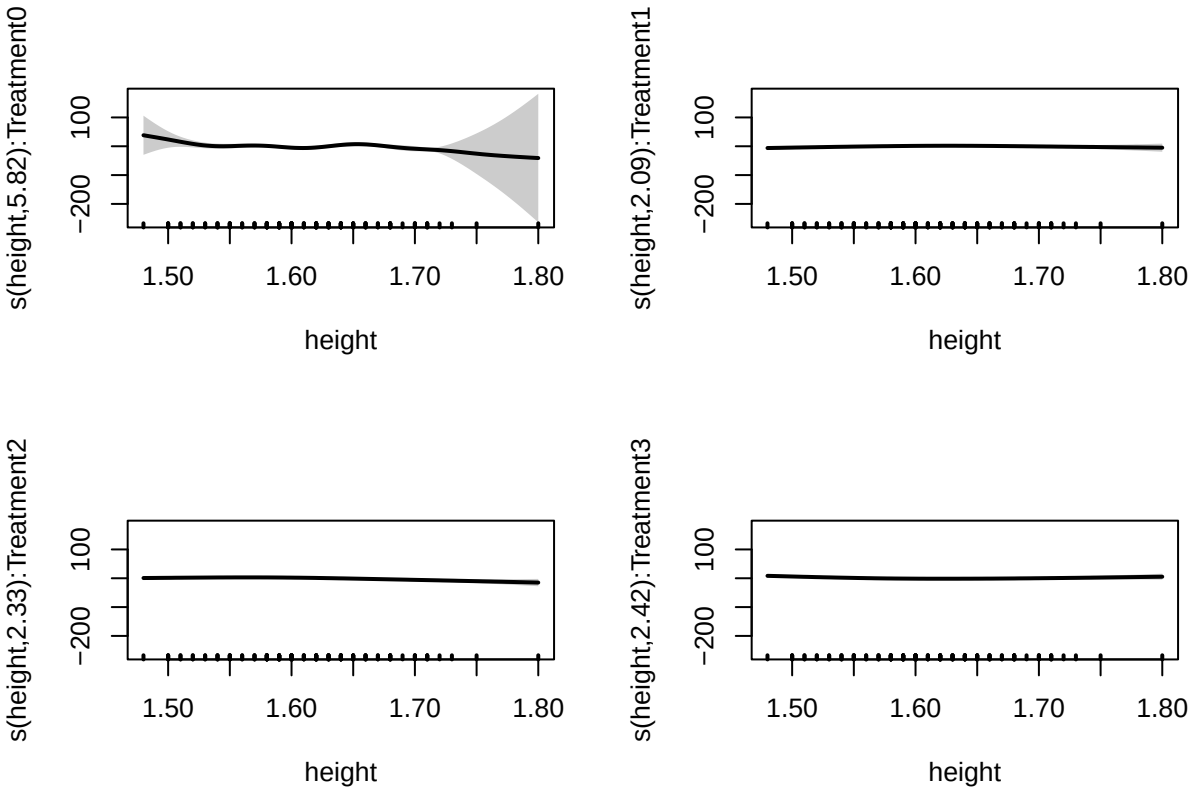
```
plot(gam1.0,pages=4, residuals=TRUE, shade=TRUE, cex=2, lwd=2)
```

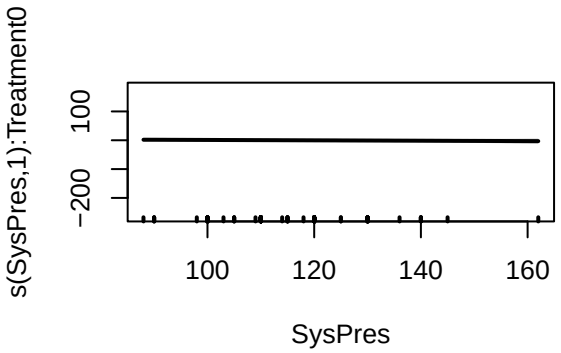

s(weight,1):Treatment0
100
−200
40 50 60 70 80 90 110
weight

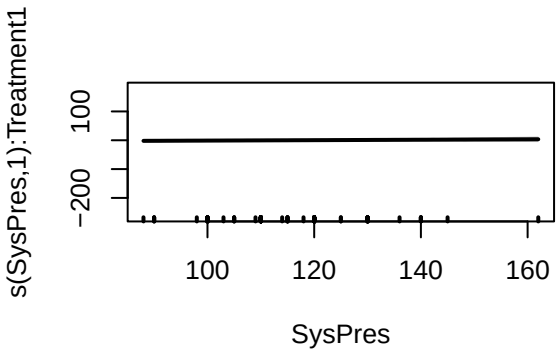

s(weight,1):Treatment1
100
−200
40 50 60 70 80 90 110
weight

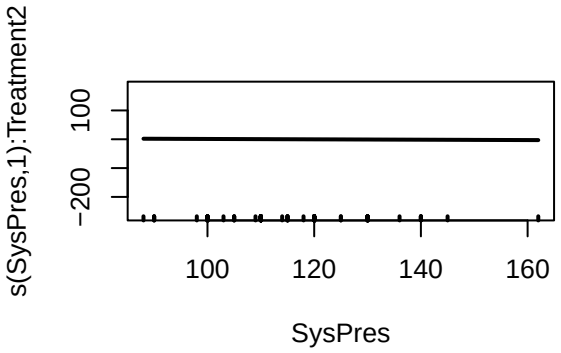

s(weight,1):Treatment2
100
−200
40 50 60 70 80 90 110
weight

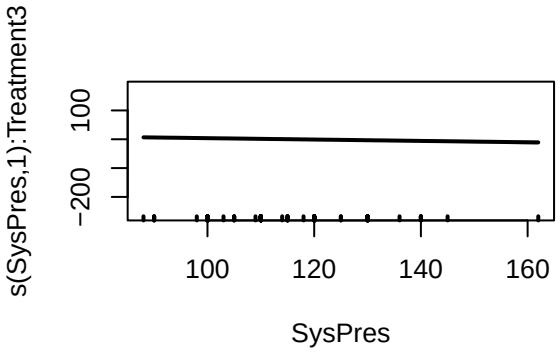

s(weight,2.93):Treatment3
100
−200
40 50 60 70 80 90 110
weight

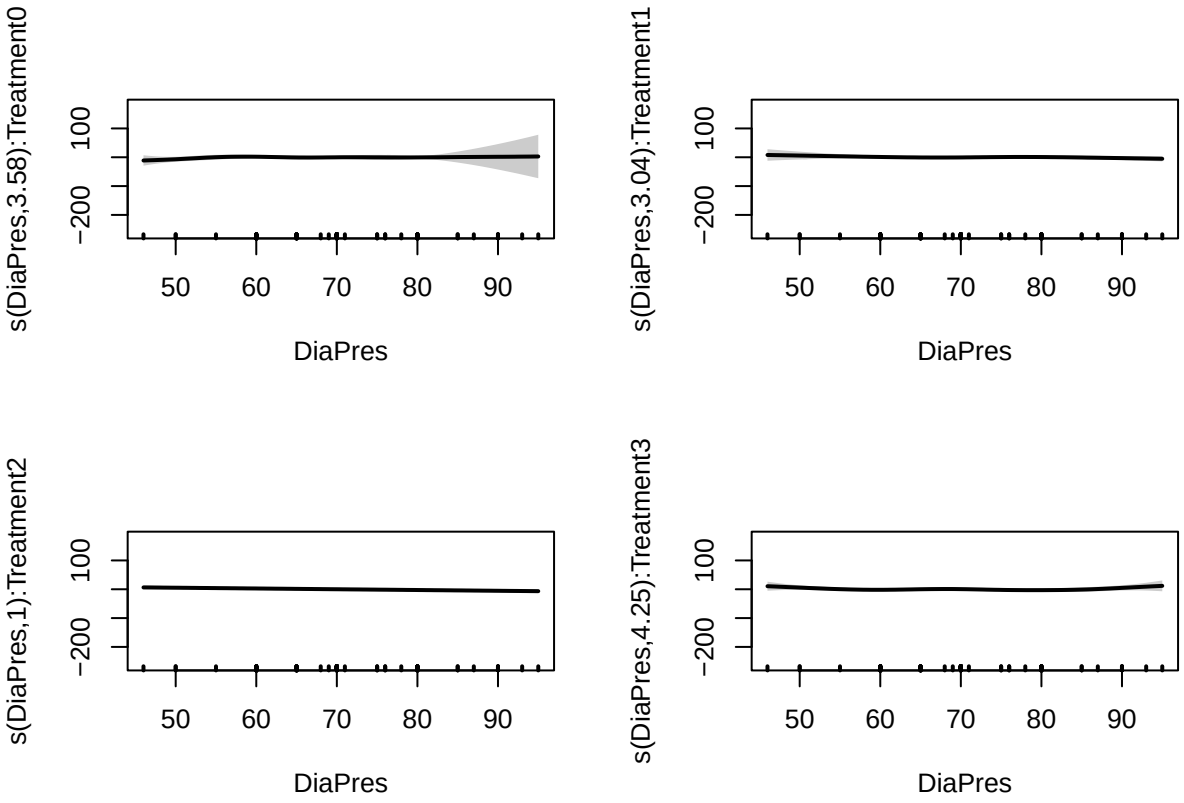

s(height,5.82):Treatment0
s(height,2.09):Treatment1
s(height,2.33):Treatment2
s(height,2.42):Treatment3
−200
100
1.50
1.60
1.70
1.80
height


s(SysPres,1):Treatment0
−200
100
100
120
140
160
SysPres

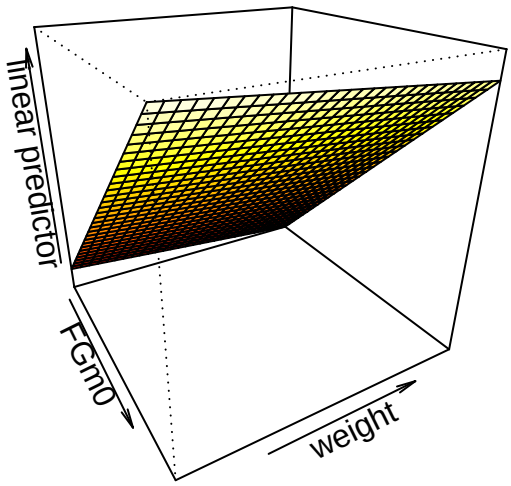

s(SysPres,1):Treatment1
−200
100
100
120
140
160
SysPres


s(SysPres,1):Treatment2
−200
100
100
120
140
160
SysPres

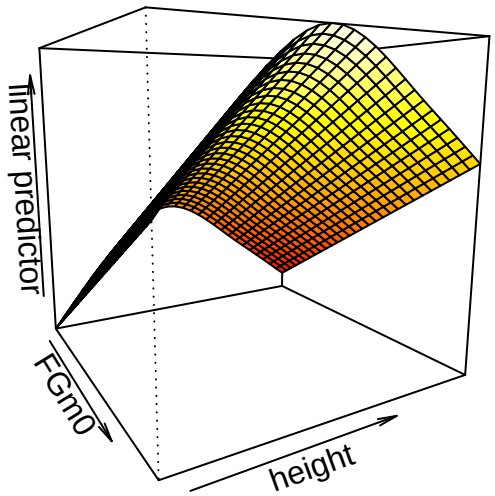

s(SysPres,1):Treatment3
−200
100
100
120
140
160
SysPres



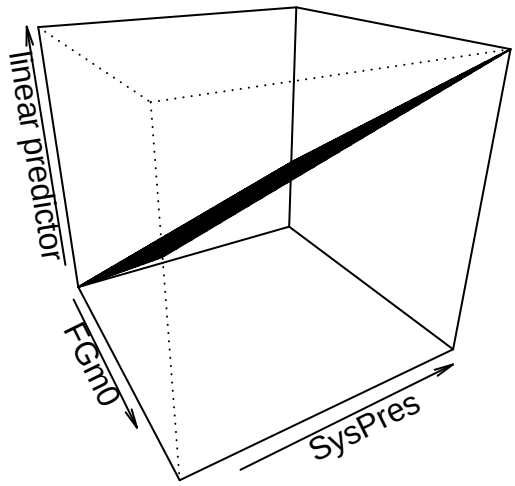

```
vis.gam(gam1.0,view=c("FGm0","weight"),
        theta = 60, phi = 25, r = sqrt(3), d = 1,)
```



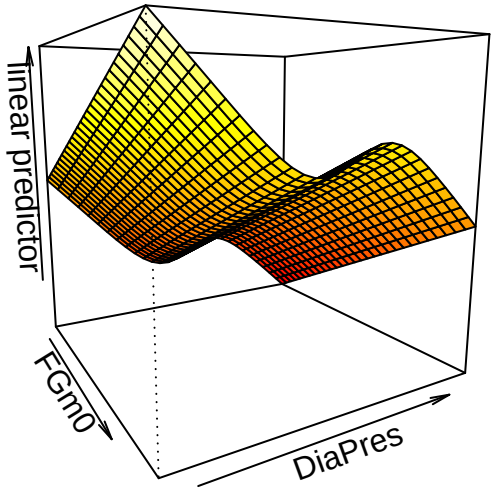

```
vis.gam(gam1.0,view=c("FGm0","height"),
        theta = 60, phi = 10, r = sqrt(3), d = 1,)
```



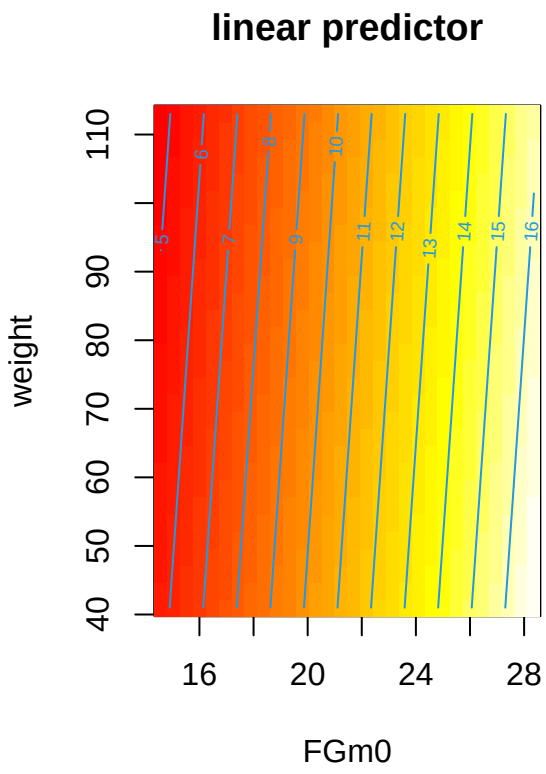

```
vis.gam(gam1.0,view=c("FGm0","SysPres"),
        theta = 60, phi = 25, r = sqrt(3), d = 1,)
```

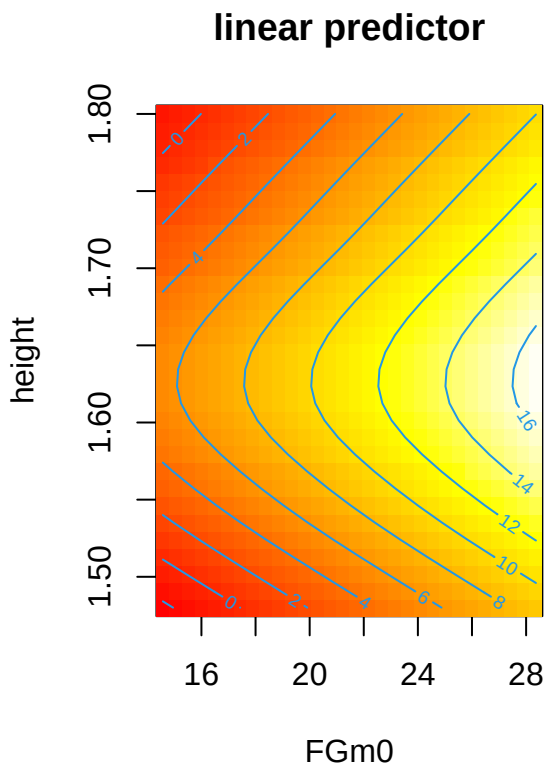



```
vis.gam(gam1.0,view=c("FGm0","DiaPres"),
        theta = 60, phi = 10, r = sqrt(3), d = 1,)
```

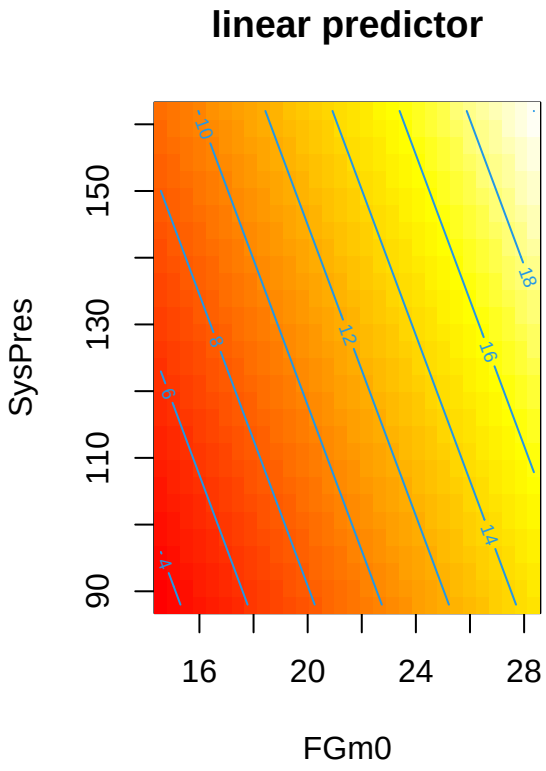


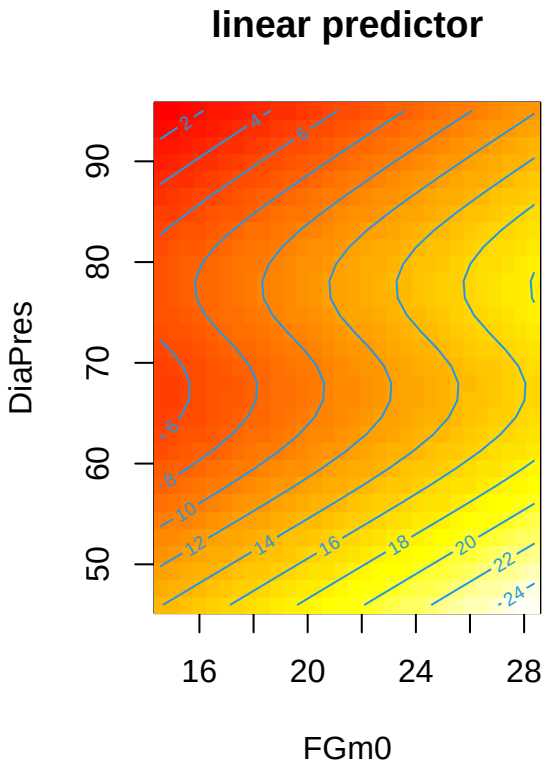

```
op <- par(mfrow=c(1,2))
vis.gam(gam1.0, view=c("FGm0","weight"), plot.type = "contour")
vis.gam(gam1.0, view=c("FGm0","height"), plot.type = "contour")
```





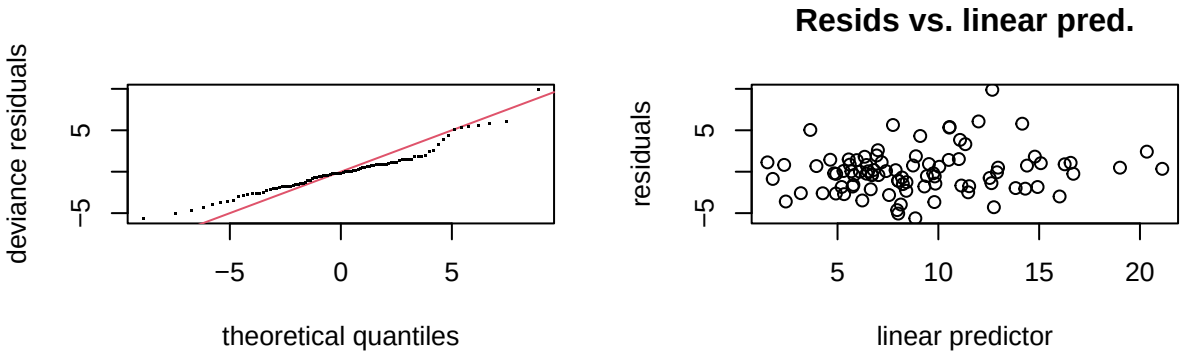

```
par(op)
op <- par(mfrow=c(1,2))
vis.gam(gam1.0, view=c("FGm0","SysPres"), plot.type = "contour")
vis.gam(gam1.0, view=c("FGm0","DiaPres"), plot.type = "contour")
```

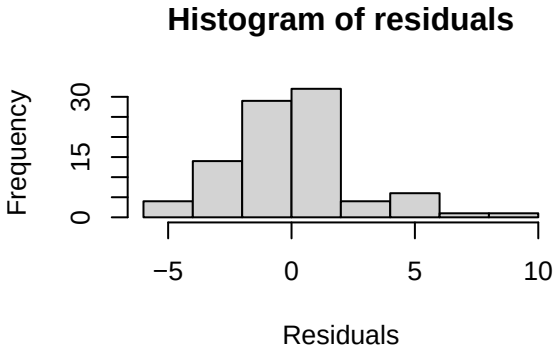


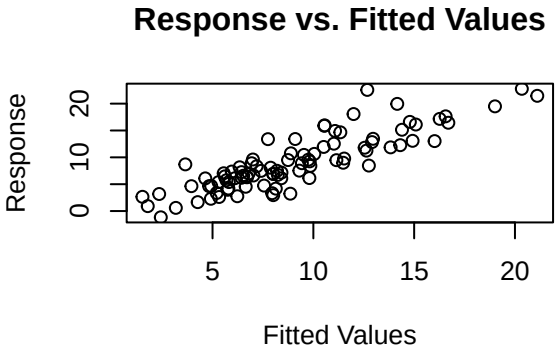



```
par(op)
```

```
gam.check(gam1.0)
```








```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 22 iterations.
## The RMS GCV score gradient at convergence was 2.28132e-07 .
## The Hessian was positive definite.
## Model rank =  149 / 149
##
## 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(weight):Treatment0  9.00 1.00    0.93    0.26
## s(weight):Treatment1  9.00 1.00    0.93    0.24
## s(weight):Treatment2  9.00 1.00    0.93    0.19
## s(weight):Treatment3  9.00 2.93    0.93    0.18
## s(height):Treatment0  9.00 5.82    1.06    0.71
## s(height):Treatment1  9.00 2.09    1.06    0.63
## s(height):Treatment2  9.00 2.33    1.06    0.70
## s(height):Treatment3  9.00 2.42    1.06    0.66
## s(SysPres):Treatment0 9.00 1.00    1.02    0.56
## s(SysPres):Treatment1 9.00 1.00    1.02    0.50
## s(SysPres):Treatment2 9.00 1.00    1.02    0.52
## s(SysPres):Treatment3 9.00 1.00    1.02    0.51
## s(DiaPres):Treatment0 9.00 3.58    1.03    0.58
## s(DiaPres):Treatment1 9.00 3.04    1.03    0.58
## s(DiaPres):Treatment2 9.00 1.00    1.03    0.55
```

```
## s(DiaPres):Treatment3 9.00 4.25    1.03    0.56
```

We see that for SysPres all edfs are close to 1 which means that the smoothing is non-significant. The same happens with weight so we will consider a model with those variables with no smoothing. The fact that the smoothing does nothing can be seen in the following plots.

Regarding performance, we could expect the model to behave much better. We will try to improve it by gradually adding complexity.

We see that the residuals are not normally distributed, the distribution is not symmetrical. However, their behavior is not alarming when compared with the theoretical normal quantiles.

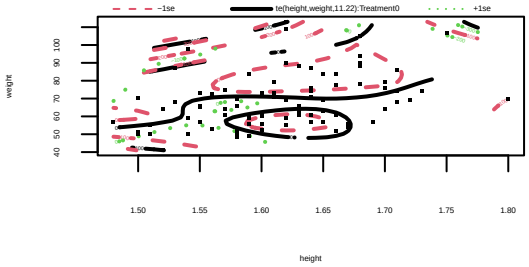
Model 2.0

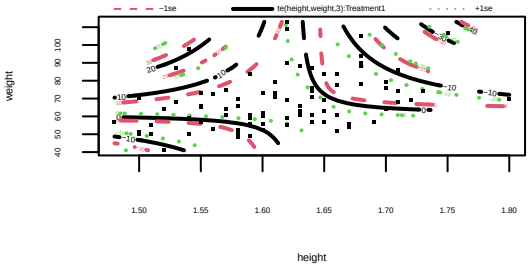
The second model we propose is to model weight and height; and the pressures as tensor products.By adding this complexity we could expect an increase of the model performance.

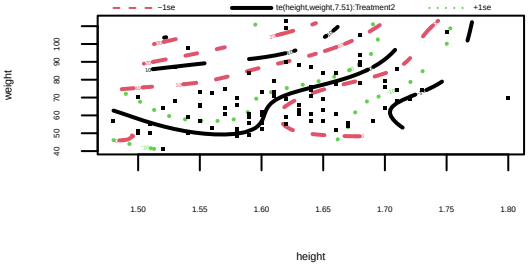
```
gam2.0 <- gam(FGm12 ~ FGm0 + te(height, weight, by=Treatment)+te(DiaPres, SysPres, by=Treatment), data=
summary(gam2.0)
```

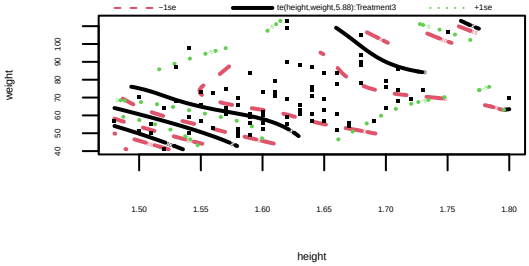
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ FGm0 + te(height, weight, by = Treatment) + te(DiaPres,
##     SysPres, by = Treatment)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.6207     3.1019  -3.424  0.00169 **
## FGm0          0.9509     0.1572   6.051 8.97e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                    edf Ref.df     F  p-value
## te(height,weight):Treatment0    11.224 12.023 3.819 0.001101 **
## te(height,weight):Treatment1     3.000  3.000 6.323 0.001721 **
## te(height,weight):Treatment2     7.509  8.749 4.156 0.001761 **
## te(height,weight):Treatment3     5.884  7.221 3.421 0.006313 **
## te(DiaPres,SysPres):Treatment0  7.885  8.452 3.783 0.003060 **
## te(DiaPres,SysPres):Treatment1  8.471  9.846 1.781 0.091233 .
## te(DiaPres,SysPres):Treatment2  5.821  6.783 5.117 0.000885 ***
## te(DiaPres,SysPres):Treatment3  6.847  7.421 5.916 0.000240 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.768   Deviance explained = 91.7%
## GCV = 17.815  Scale est. = 6.335     n = 91
```

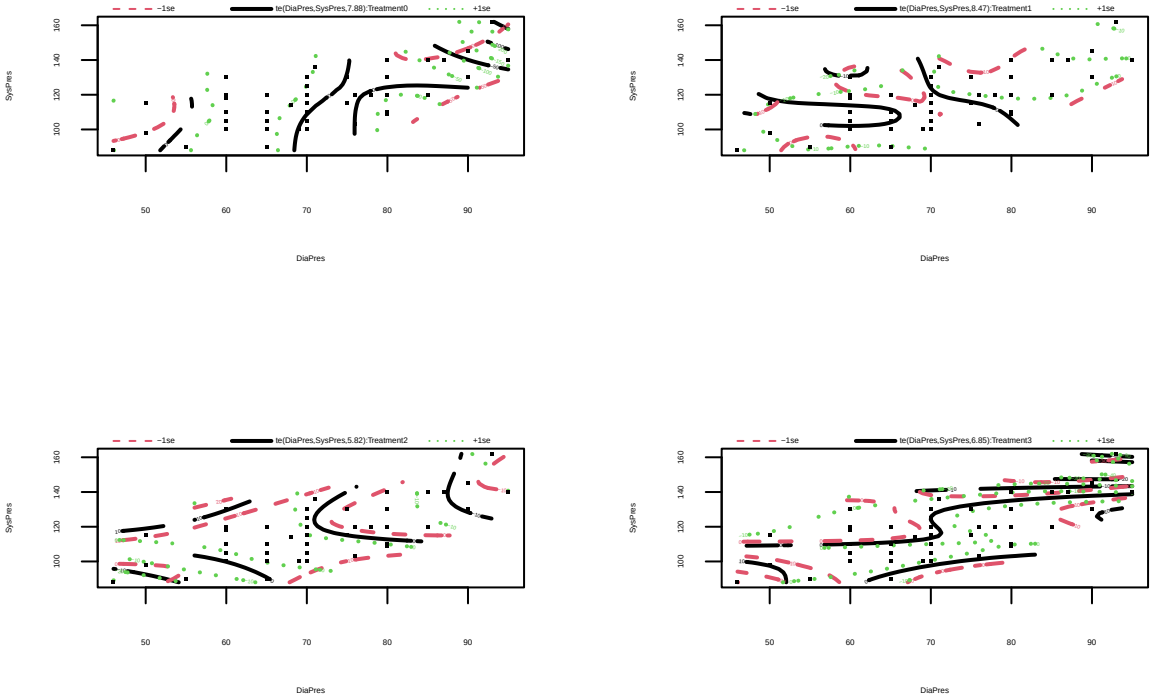
```
par(mfrow=c(4,2))
plot(gam2.0,pages=2, residuals=TRUE, shade=TRUE, cex=2, lwd=2)
```


−1se
te(height,weight,11.22):Treatment0
+1se
weight
height
1.50
1.55
1.60
1.65
1.70
1.75
1.80
40
50
60
70
80
90
100

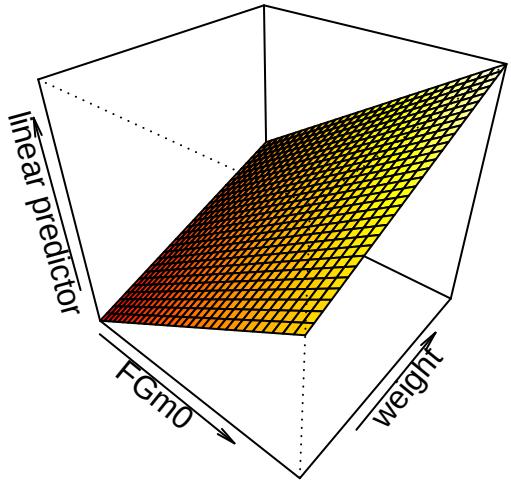

−1se
te(height,weight,3):Treatment1
+1se
weight
height
1.50
1.55
1.60
1.65
1.70
1.75
1.80
40
50
60
70
80
90
100


−1se
te(height,weight,7.51):Treatment2
+1se
weight
height
1.50
1.55
1.60
1.65
1.70
1.75
1.80
40
50
60
70
80
90
100

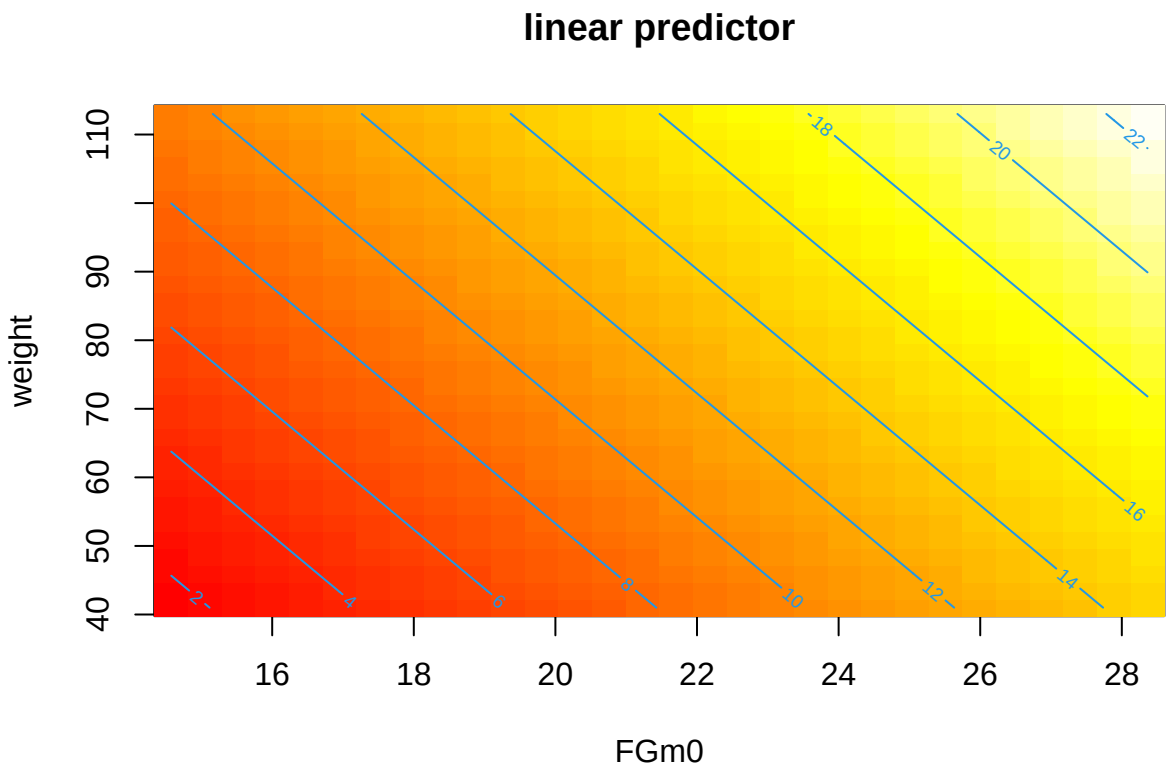

−1se
te(height,weight,5.88):Treatment3
+1se
weight
height
1.50
1.55
1.60
1.65
1.70
1.75
1.80
40
50
60
70
80
90
100



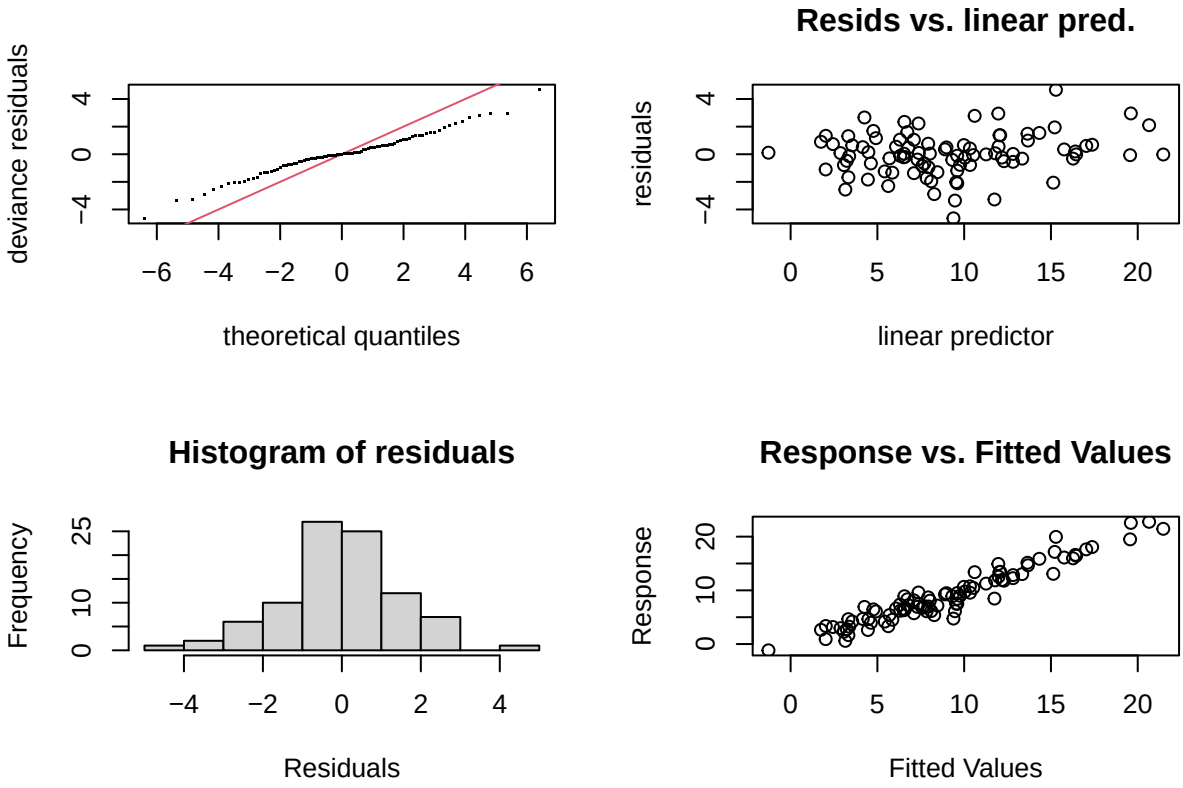

```
vis.gam(gam2.0,view=c("FGm0","weight"),
        theta = 40, phi = 35, r = sqrt(4), d = 1,)
```




```
vis.gam(gam2.0, view=c("FGm0","weight"), plot.type = "contour")
```




```
gam.check(gam2.0)
```




```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 22 iterations.
## The RMS GCV score gradient at convergence was 2.413232e-07 .
## The Hessian was positive definite.
## Model rank =  194 / 194
##
## 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
## te(height,weight):Treatment0   24.00 11.22    0.96    0.29
## te(height,weight):Treatment1   24.00  3.00    0.97    0.29
## te(height,weight):Treatment2   24.00  7.51    0.96    0.23
## te(height,weight):Treatment3   24.00  5.88    0.96    0.27
## te(DiaPres,SysPres):Treatment0 24.00  7.88    1.24    0.98
## te(DiaPres,SysPres):Treatment1 24.00  8.47    1.22    0.99
## te(DiaPres,SysPres):Treatment2 24.00  5.82    1.22    1.00
## te(DiaPres,SysPres):Treatment3 24.00  6.85    1.25    0.99
```

The behavior in residuals distribution is still asymmetrical, but random enough to justify the fit, as it can be seen in the second image of the model check.

The model performance has also noticeably increased, now to a 91% of variance explained. In addition, the number of significant p-values in variables has also been increased. When considering 0.1 of significance, all variables present relevant behavior in the model.

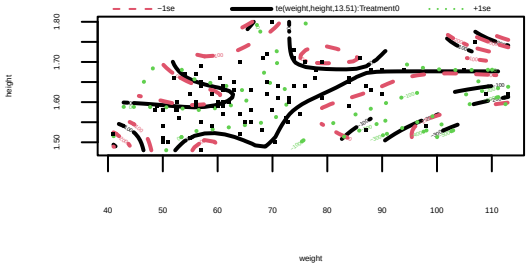
Model 3.0

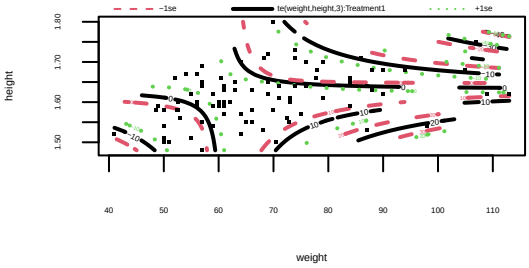
The third model is treating weight and height as a tensor product but keeping the pressures independently smoothed and vice versa.

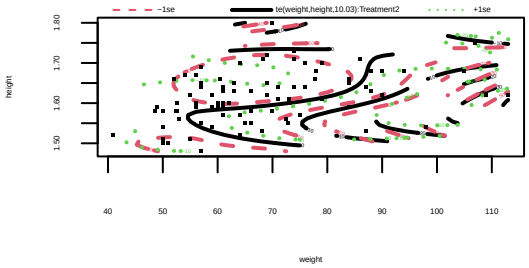
```
gam3.0 <- gam(FGm12 ~ FGm0 + te(weight, height, by=Treatment) +
                s(SysPres, by=Treatment) + s(DiaPres, by=Treatment), data=hirs)
summary(gam3.0)
```

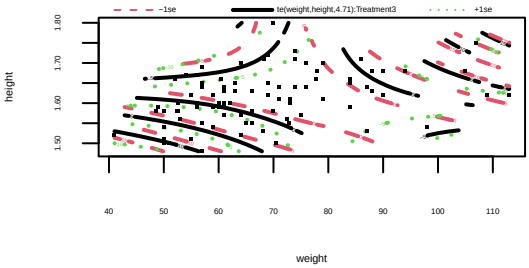
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ FGm0 + te(weight, height, by = Treatment) + s(SysPres,
##     by = Treatment) + s(DiaPres, by = Treatment)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.5522     2.7051  -3.901 0.000521 ***
## FGm0          0.9936     0.1377   7.213 5.96e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                  edf Ref.df     F  p-value
## te(weight,height):Treatment0 13.511 14.177 6.114 2.03e-05 ***
## te(weight,height):Treatment1  3.000  3.000 7.592 0.000681 ***
## te(weight,height):Treatment2 10.029 11.072 4.052 0.001124 **
## te(weight,height):Treatment3  4.705  5.619 7.099 0.000204 ***
## s(SysPres):Treatment0         2.241  2.474 0.600 0.486774
## s(SysPres):Treatment1         1.000  1.000 0.978 0.330785
## s(SysPres):Treatment2         5.699  6.238 3.448 0.010520 *
## s(SysPres):Treatment3         4.089  4.748 3.333 0.043120 *
## s(DiaPres):Treatment0         3.192  3.561 7.378 0.001022 **
## s(DiaPres):Treatment1         6.421  7.341 2.543 0.037944 *
## s(DiaPres):Treatment2         1.000  1.000 5.438 0.026856 *
## s(DiaPres):Treatment3         5.002  5.483 6.405 0.000336 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.824   Deviance explained = 94.3%
## GCV = 14.979  Scale est. = 4.7919    n = 91
```

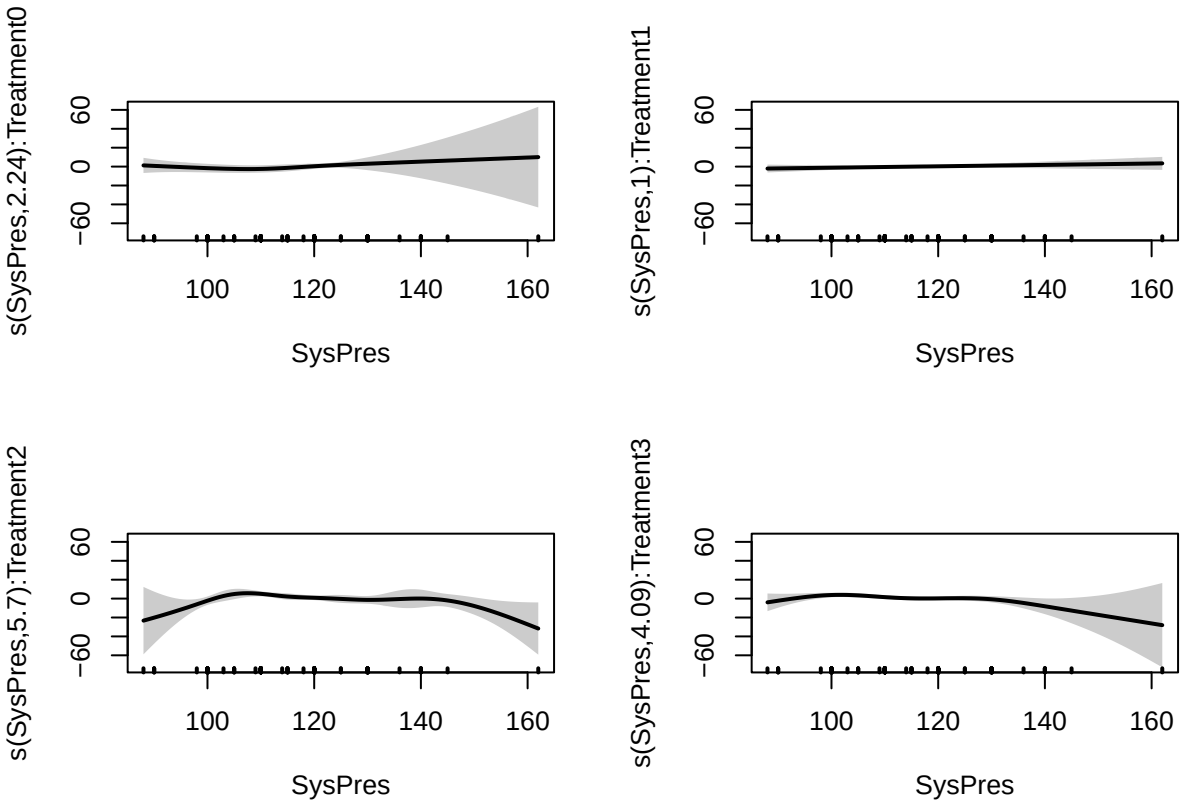
```
par(mfrow=c(4,2))
plot(gam3.0,pages=3, residuals=TRUE, shade=TRUE, cex=2, lwd=2)
```

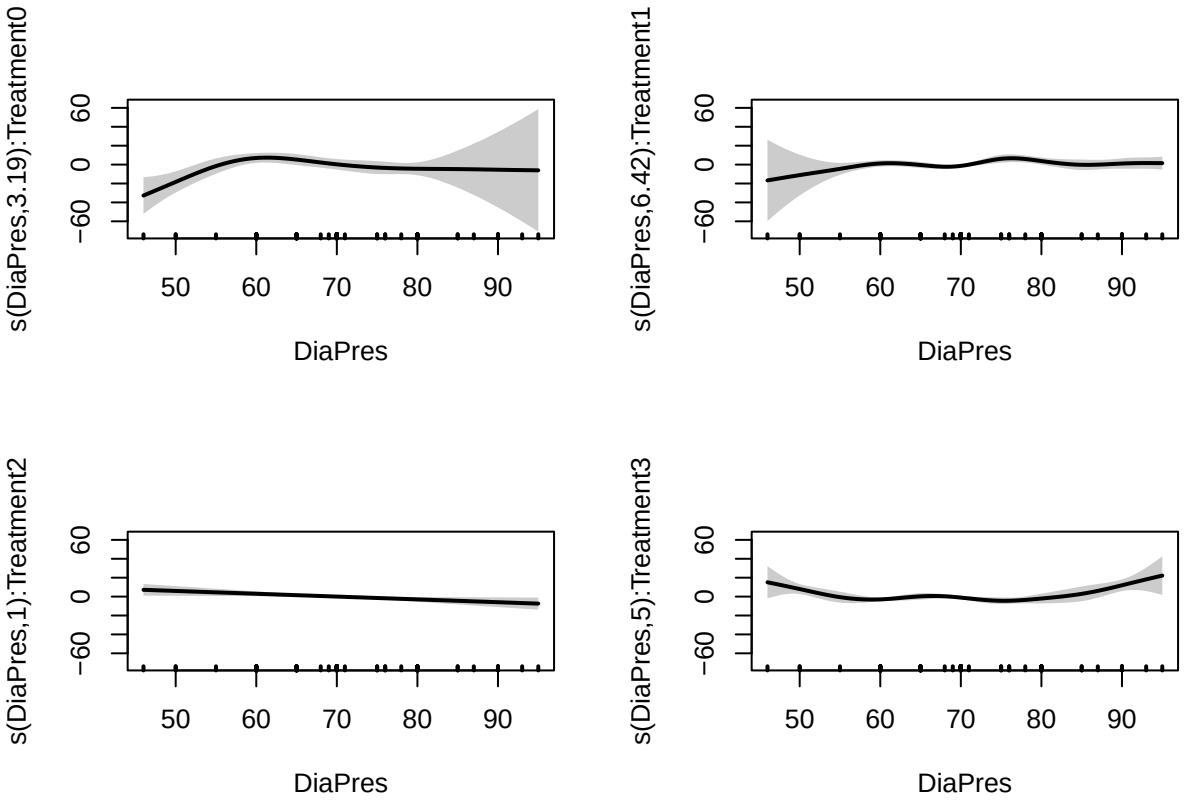

−1se
te(weight,height,13.51):Treatment0
+1se
height
weight


−1se
te(weight,height,3):Treatment1
+1se
height
weight

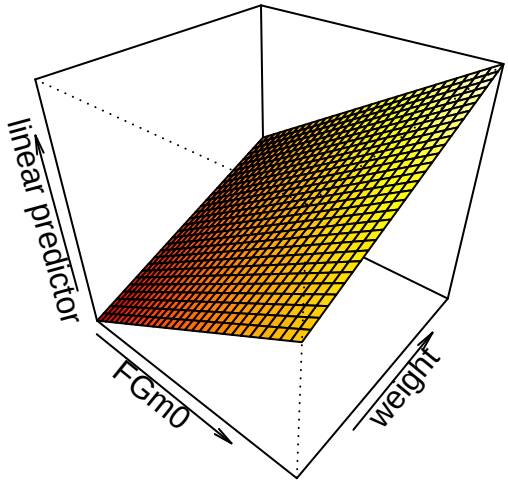

−1se
te(weight,height,10.03):Treatment2
+1se
height
weight


−1se
te(weight,height,4.71):Treatment3
+1se
height
weight

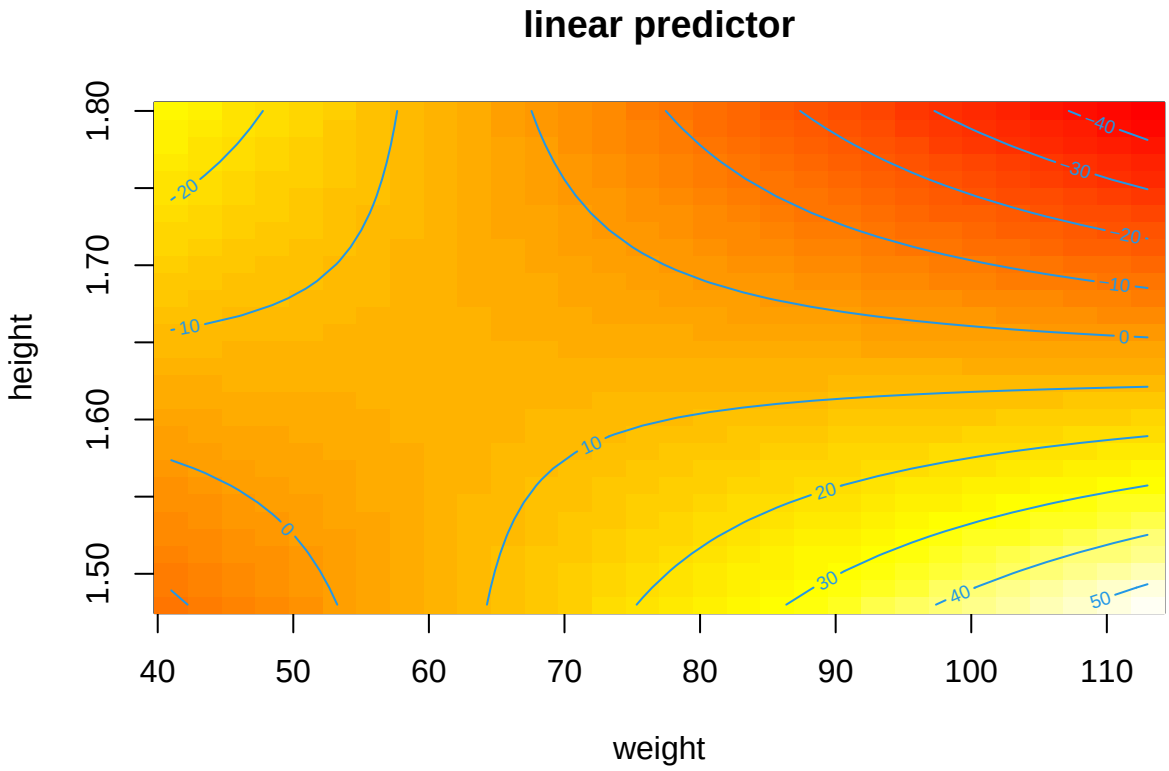

s(SysPres,2.24):Treatment0
s(SysPres,1):Treatment1
s(SysPres,5.7):Treatment2
s(SysPres,4.09):Treatment3
−60
0
60
100
120
140
160
SysPres



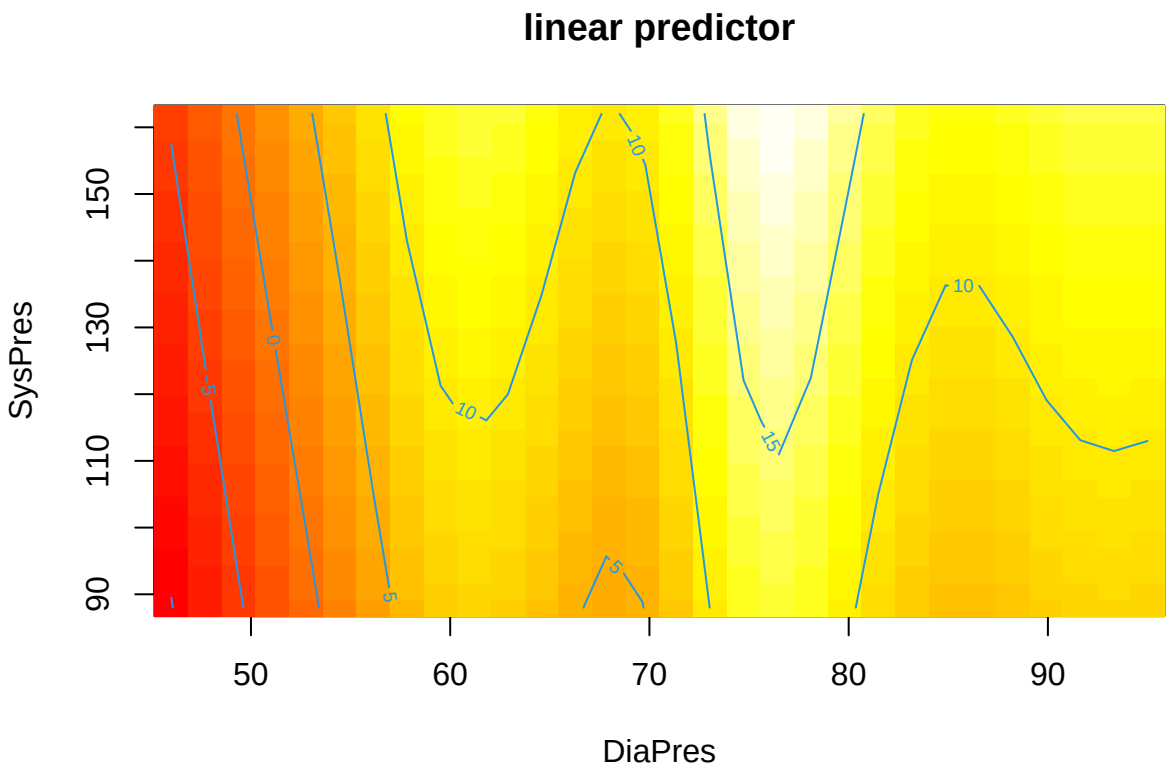

```
vis.gam(gam3.0,view=c("FGm0","weight"),
        theta = 40, phi = 35, r = sqrt(4), d = 1,)
```



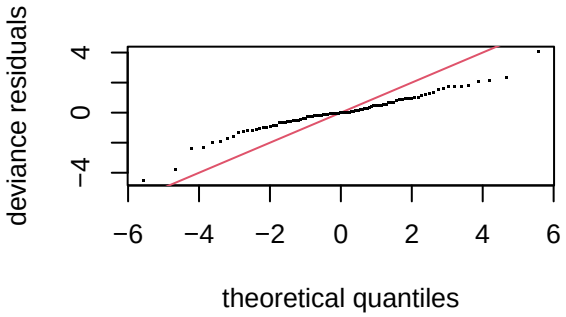

```
vis.gam(gam3.0, view=c("weight","height"), plot.type = "contour")
```

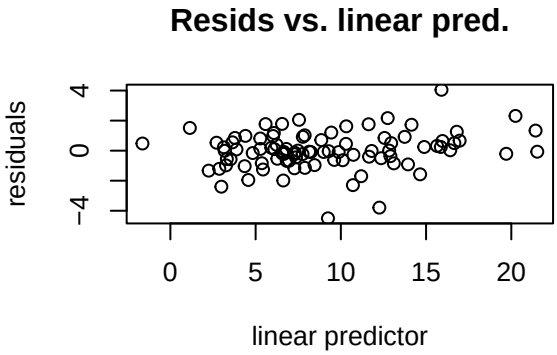


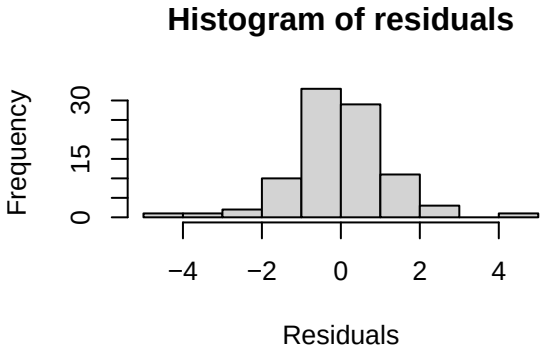

```
vis.gam(gam3.0, view=c("DiaPres","SysPres"), plot.type = "contour")
```

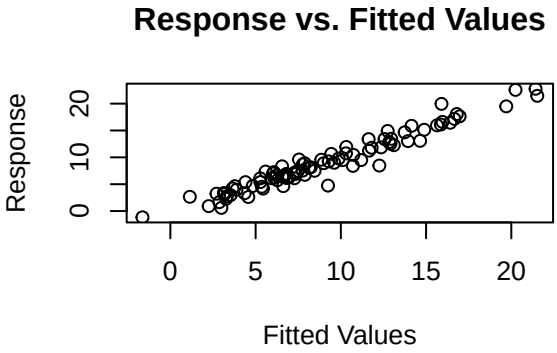



```
gam.check(gam3.0)
```










```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 30 iterations by steepest
## descent step failure.
## The RMS GCV score gradient at convergence was 3.487809e-05 .
## The Hessian was positive definite.
## Model rank =  170 / 170
##
## 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
## te(weight,height):Treatment0 24.00 13.51    0.98   0.350
## te(weight,height):Treatment1 24.00  3.00    0.99   0.345
## te(weight,height):Treatment2 24.00 10.03    0.99   0.440
## te(weight,height):Treatment3 24.00  4.71    0.98   0.370
## s(SysPres):Treatment0         9.00  2.24    1.06   0.655
## s(SysPres):Treatment1         9.00  1.00    1.06   0.685
## s(SysPres):Treatment2         9.00  5.70    1.06   0.700
## s(SysPres):Treatment3         9.00  4.09    1.06   0.690
## s(DiaPres):Treatment0         9.00  3.19    0.85   0.045 *
## s(DiaPres):Treatment1         9.00  6.42    0.85   0.060 .
## s(DiaPres):Treatment2         9.00  1.00    0.85   0.055 .
## s(DiaPres):Treatment3         9.00  5.00    0.85   0.085 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This is the best model so far. With a 94% of variance explained and 0.84 R adjusted coefficient it seems to be a satisfactory use of gams models, which rely on their simplification to deal with the curse of dimensionality.

When looking at the distributions of residuals, altogether they do not align with the theoretical quantiles, this is due to their high concentration around 0. This fact is however not alarming since, when looking at the distribution it seems to be no distinguishable pattern in the residuals. This model seems to perform satisfactory when explaining the data.

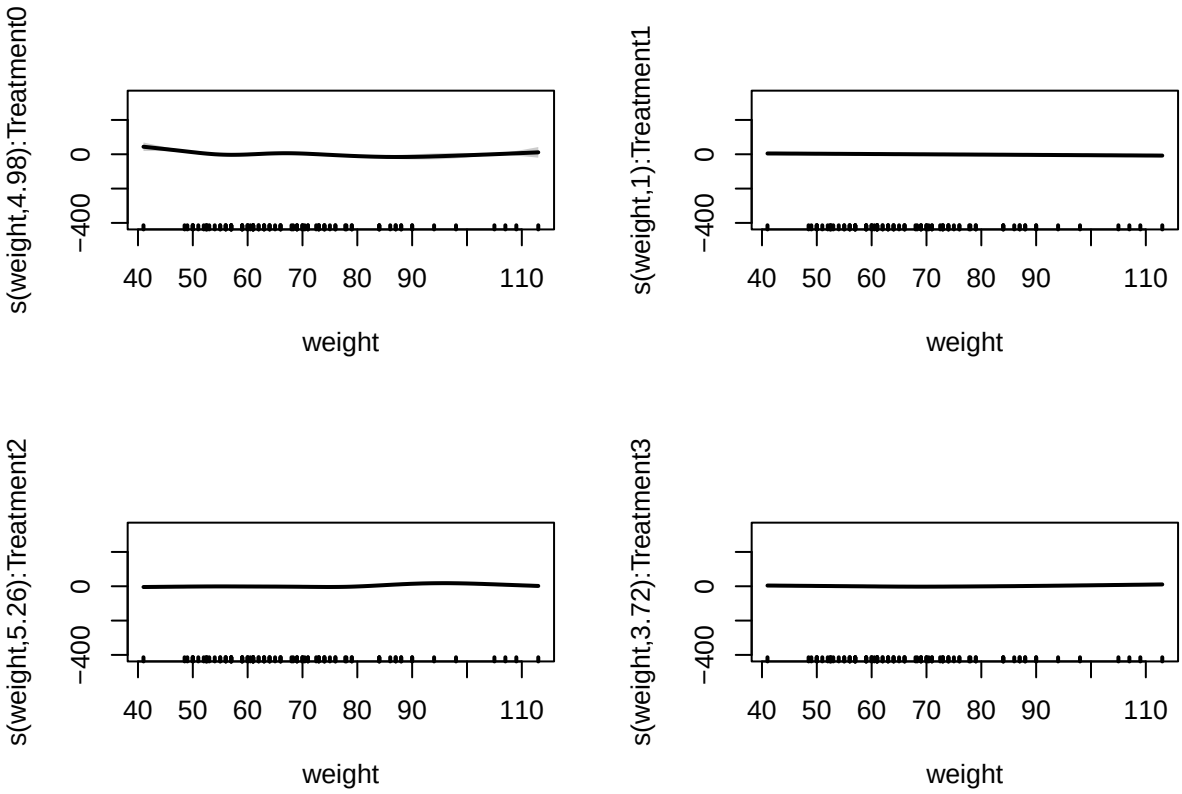
Model 4.0

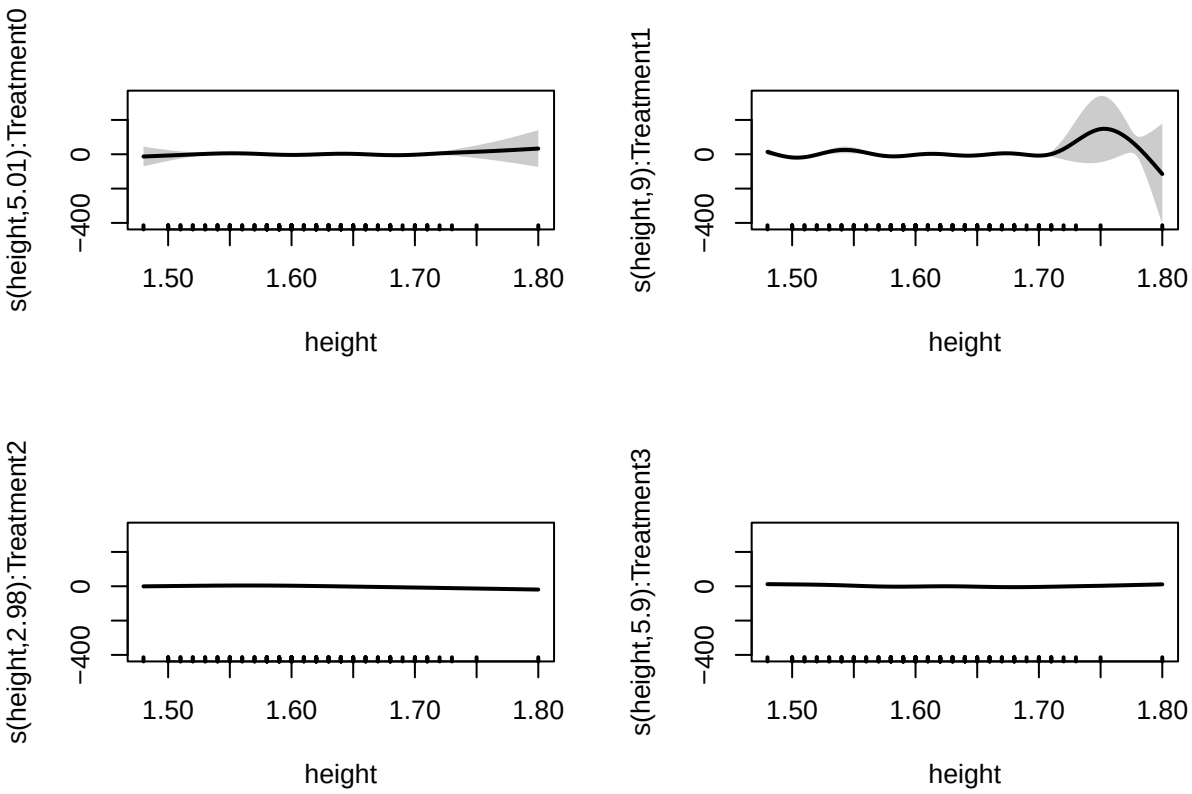
As a counterpart the fourth model is treating pressures as a tensor product but keeping weight and height independently smoothed.

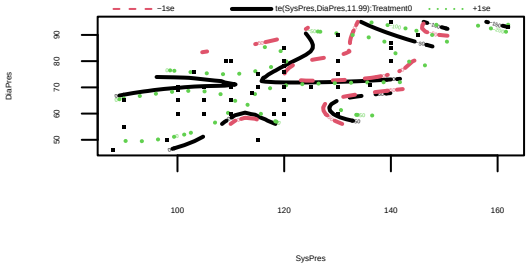
```
gam4.0 <- gam(FGm12 ~ FGm0 + s(weight, by=Treatment) + s(height, by=Treatment) +
                te(SysPres, DiaPres, by=Treatment), data=hirs)
summary(gam4.0)
```

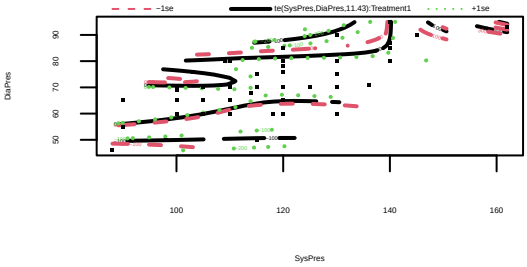
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ FGm0 + s(weight, by = Treatment) + s(height, by = Treatment) +
##     te(SysPres, DiaPres, by = Treatment)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.72506    1.11990  -5.112    0.994
## FGm0         0.68908    0.04641  14.847    0.993
##
## Approximate significance of smooth terms:
##                                    edf Ref.df         F  p-value
## s(weight):Treatment0             4.976  4.976 2.846e+01 0.141674
## s(weight):Treatment1             1.000  1.000 1.020e+02 0.062833 .
## s(weight):Treatment2             5.258  5.258 6.463e+01 0.112418
## s(weight):Treatment3             3.715  3.716 4.740e+00 0.327547
## s(height):Treatment0             5.012  5.012 1.559e+00 0.511180
## s(height):Treatment1             9.000  9.000 9.105e+05 0.000814 ***
## s(height):Treatment2             2.980  2.980 8.809e+01 0.080274 .
## s(height):Treatment3             5.902  5.902 1.369e+02 0.062499 .
## te(SysPres,DiaPres):Treatment0 11.993 11.993 5.149e+02 0.034370 *
## te(SysPres,DiaPres):Treatment1 11.425 11.425 3.805e+05 0.001220 **
## te(SysPres,DiaPres):Treatment2 12.552 12.552 7.778e+02 0.033758 *
## te(SysPres,DiaPres):Treatment3 15.186 15.186 1.179e+05 0.002729 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =      1   Deviance explained =  100%
## GCV = 1.5365  Scale est. = 1.5723e-05  n = 91
```

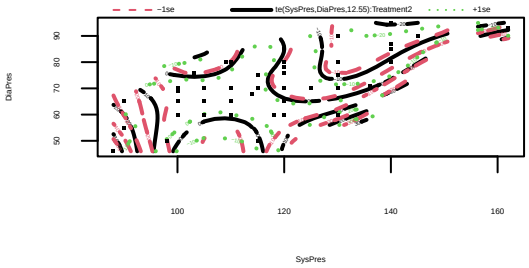
```
par(mfrow=c(4,2))
plot(gam4.0,pages=3, residuals=TRUE, shade=TRUE, cex=2, lwd=2)
```

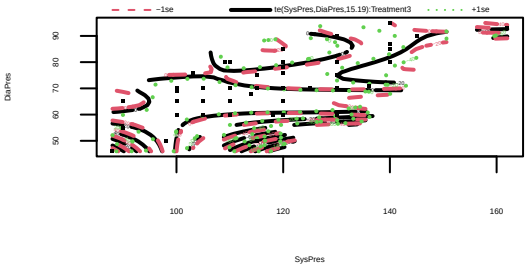

s(weight,4.98):Treatment0
s(weight,1):Treatment1
s(weight,5.26):Treatment2
s(weight,3.72):Treatment3
−400
0
40
50
60
70
80
90
110
weight


s(height,5.01):Treatment0
s(height,9):Treatment1
s(height,2.98):Treatment2
s(height,5.9):Treatment3
−400
0
1.50
1.60
1.70
1.80
height

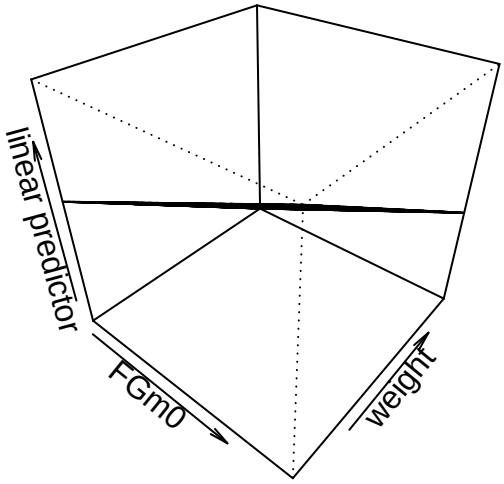




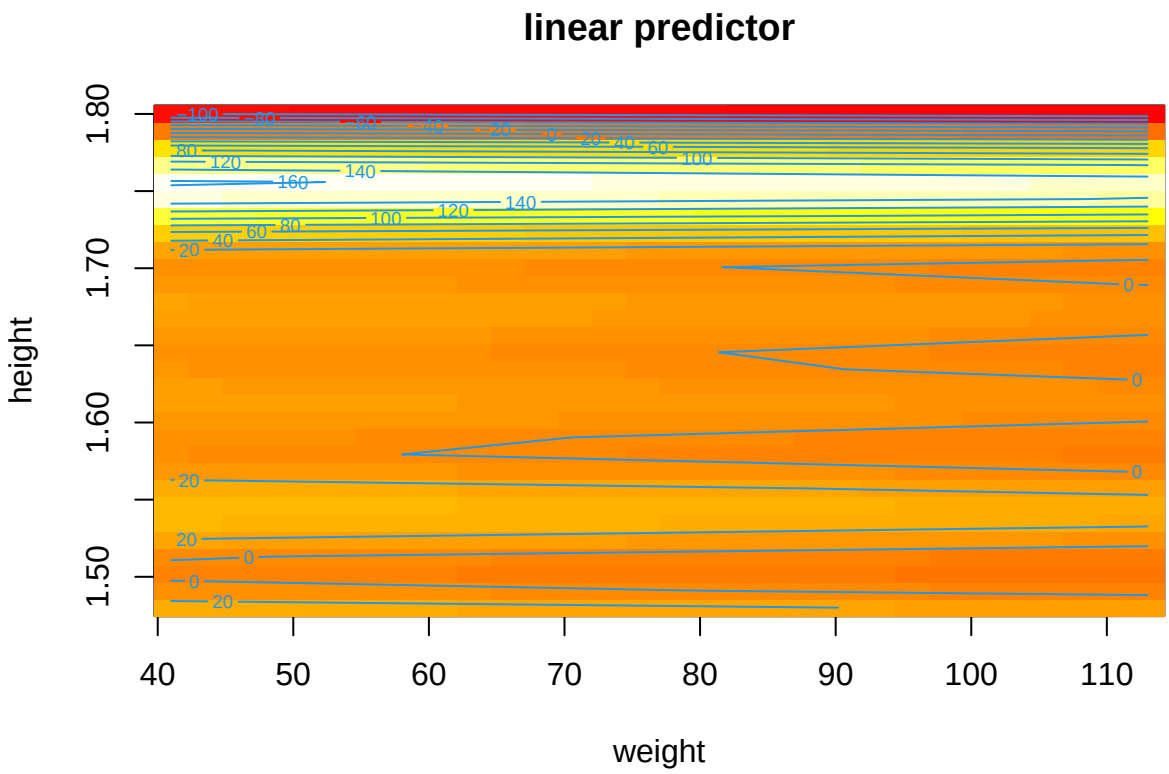




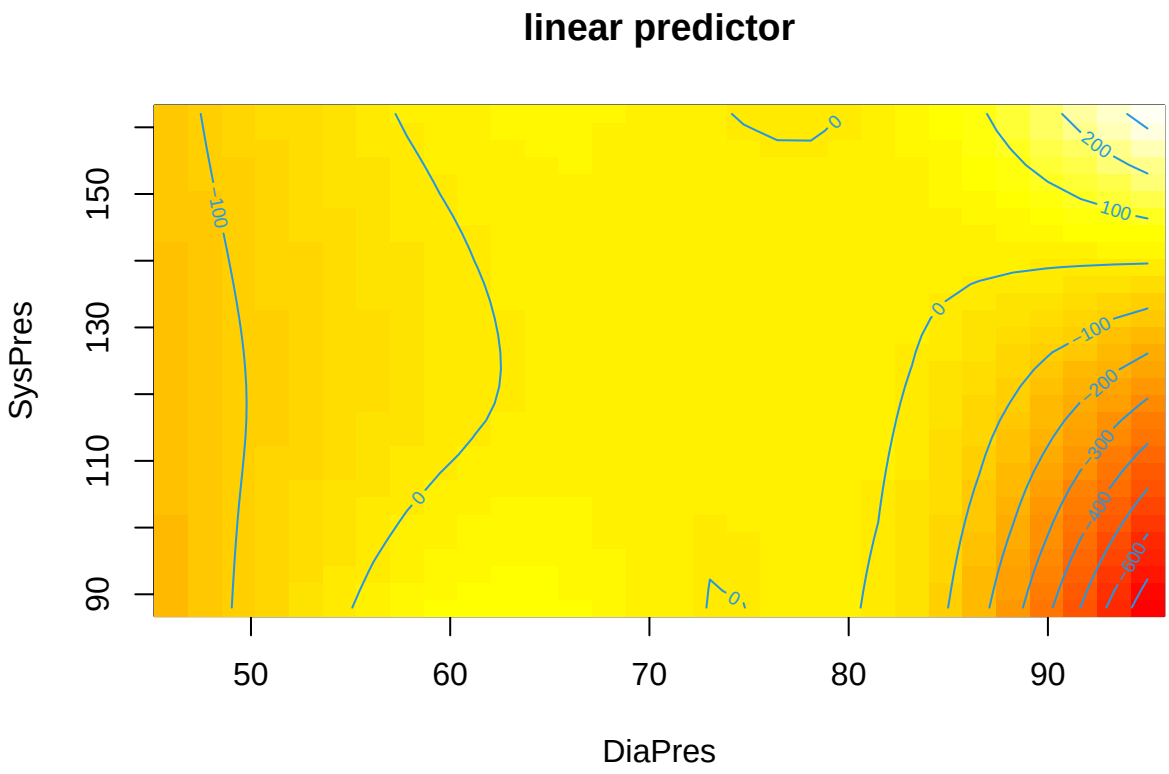

```
vis.gam(gam4.0,view=c("FGm0","weight"),
        theta = 40, phi = 35, r = sqrt(4), d = 1,)
```



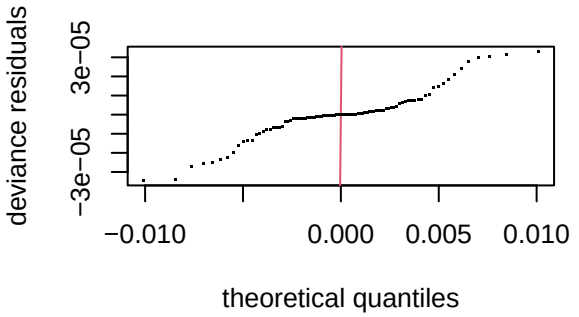

```
vis.gam(gam4.0, view=c("weight","height"), plot.type = "contour")
```

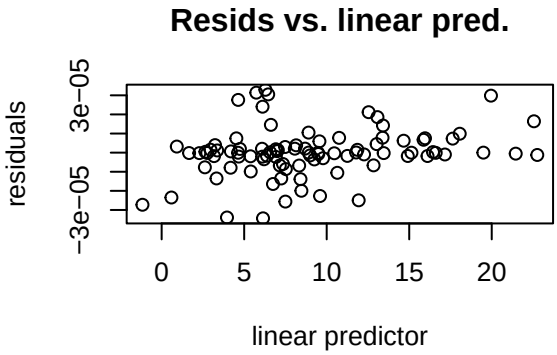


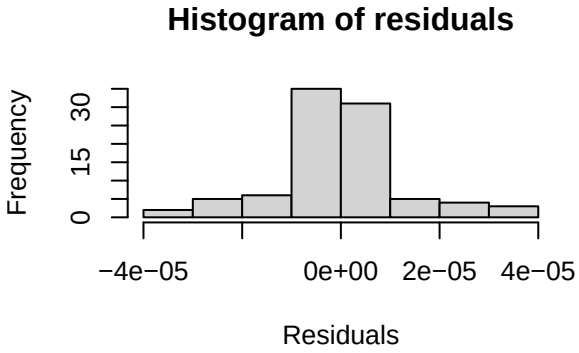

```
vis.gam(gam4.0, view=c("DiaPres","SysPres"), plot.type = "contour")
```

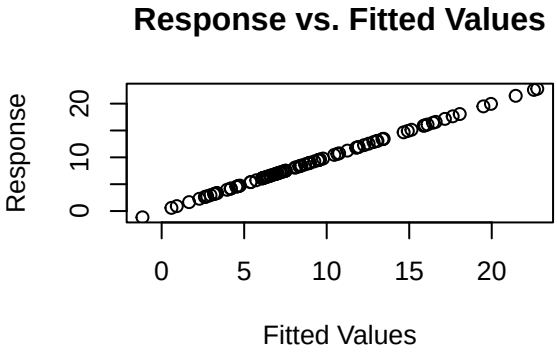



```
gam.check(gam4.0)
```










```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 241 iterations by steepest
## descent step failure.
## The RMS GCV score gradient at convergence was 0.001210238 .
## The Hessian was not positive definite.
## Model rank =  170 / 170
##
## 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(weight):Treatment0            9.00  4.98    0.93    0.24
## s(weight):Treatment1            9.00  1.00    0.93    0.23
## s(weight):Treatment2            9.00  5.26    0.93    0.18
## s(weight):Treatment3            9.00  3.72    0.93    0.29
## s(height):Treatment0            9.00  5.01    1.17    0.94
## s(height):Treatment1            9.00  9.00    1.17    0.94
## s(height):Treatment2            9.00  2.98    1.17    0.97
## s(height):Treatment3            9.00  5.90    1.17    0.95
## te(SysPres,DiaPres):Treatment0 24.00 11.99    1.21    0.99
## te(SysPres,DiaPres):Treatment1 24.00 11.43    1.24    0.99
## te(SysPres,DiaPres):Treatment2 24.00 12.55    1.27    0.99
## te(SysPres,DiaPres):Treatment3 24.00 15.19    1.23    0.99
```

This model is behaving in an unexpected way. By giving a 100% of variance explanation and and an adjusted

R coefficient of 1 we should expect overfitting. When looking at the check exposed before, it can be clearly seen that the residuals scale is outstanding by its order of magnitude. When we should expect some noise in these residuals, they behave almost constant near zero which is a sign of almost interpolation.

Anova Model Selection

For the selection of models, different ANOVA analysis will be performed. Since the predicted variable is continuous, the test chosen through this section will be the F-test. The ANOVA test is perfectly suited for this ocasion in which the models are one contained by the other.

```
anova(gam1.0, gam2.0, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ FGm0 + Treatment + s(weight, by = Treatment) + s(height,
##     by = Treatment) + s(SysPres, by = Treatment) + s(DiaPres,
##     by = Treatment)
## Model 2: FGm12 ~ FGm0 + te(height, weight, by = Treatment) + te(DiaPres,
##     SysPres, by = Treatment)
##   Resid. Df Resid. Dev   Df Deviance      F   Pr(>F)
## 1    47.104     629.95
## 2    25.504     205.00 21.6   424.95 3.1055 0.003518 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(gam1.0, gam3.0, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ FGm0 + Treatment + s(weight, by = Treatment) + s(height,
##     by = Treatment) + s(SysPres, by = Treatment) + s(DiaPres,
##     by = Treatment)
## Model 2: FGm12 ~ FGm0 + te(weight, height, by = Treatment) + s(SysPres,
##     by = Treatment) + s(DiaPres, by = Treatment)
##   Resid. Df Resid. Dev     Df Deviance      F    Pr(>F)
## 1    47.104     629.95
## 2    23.286     139.50 23.818   490.44 4.2971 0.0003996 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(gam1.0, gam4.0, test="F")
```

```
## Warning in pf(Fvalue, abs(dfs), df.scale, lower.tail = FALSE): NaNs produced
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ FGm0 + Treatment + s(weight, by = Treatment) + s(height,
##     by = Treatment) + s(SysPres, by = Treatment) + s(DiaPres,
##     by = Treatment)
## Model 2: FGm12 ~ FGm0 + s(weight, by = Treatment) + s(height, by = Treatment) +
##     te(SysPres, DiaPres, by = Treatment)
```

```
##     Resid. Df Resid. Dev     Df Deviance      F Pr(>F)
## 1  4.7104e+01     629.95
## 2 -7.0647e-05       0.00 47.104   629.95 850567    NaN
```

When compared with the first model (baseline) all of the additive models seem to add significant variance difference except for the fourth of them. This, together with the extreme behavior of the residuals, is a good sign for it to be discarded.

```
anova(gam2.0, gam3.0, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ FGm0 + te(height, weight, by = Treatment) + te(DiaPres,
##     SysPres, by = Treatment)
## Model 2: FGm12 ~ FGm0 + te(weight, height, by = Treatment) + s(SysPres,
##     by = Treatment) + s(DiaPres, by = Treatment)
##   Resid. Df Resid. Dev     Df Deviance      F   Pr(>F)
## 1    25.504      205.0
## 2    23.286      139.5 2.2182   65.495 6.1616 0.005825 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our two preferred models so far, are still different regarding ANOVA analysis, which is a good sign when deciding for a final one.

```
anova(gam1.0, gam2.0, gam3.0, gam4.0, test="F")
```

```
## Warning in pf(Fvalue, abs(dfs), df.scale, lower.tail = FALSE): NaNs produced

## Analysis of Deviance Table
##
## Model 1: FGm12 ~ FGm0 + Treatment + s(weight, by = Treatment) + s(height,
##     by = Treatment) + s(SysPres, by = Treatment) + s(DiaPres,
##     by = Treatment)
## Model 2: FGm12 ~ FGm0 + te(height, weight, by = Treatment) + te(DiaPres,
##     SysPres, by = Treatment)
## Model 3: FGm12 ~ FGm0 + te(weight, height, by = Treatment) + s(SysPres,
##     by = Treatment) + s(DiaPres, by = Treatment)
## Model 4: FGm12 ~ FGm0 + s(weight, by = Treatment) + s(height, by = Treatment) +
##     te(SysPres, DiaPres, by = Treatment)
##     Resid. Df Resid. Dev      Df Deviance       F Pr(>F)
## 1  4.7104e+01     629.95
## 2  2.5504e+01     205.00 21.6001   424.95 1251260    NaN
## 3  2.3286e+01     139.50  2.2182    65.50 1877871    NaN
## 4 -7.0647e-05       0.00 23.2861   139.50  381024    NaN
```

```
anova(gam1.0, gam2.0, gam3.0, test="F")
```

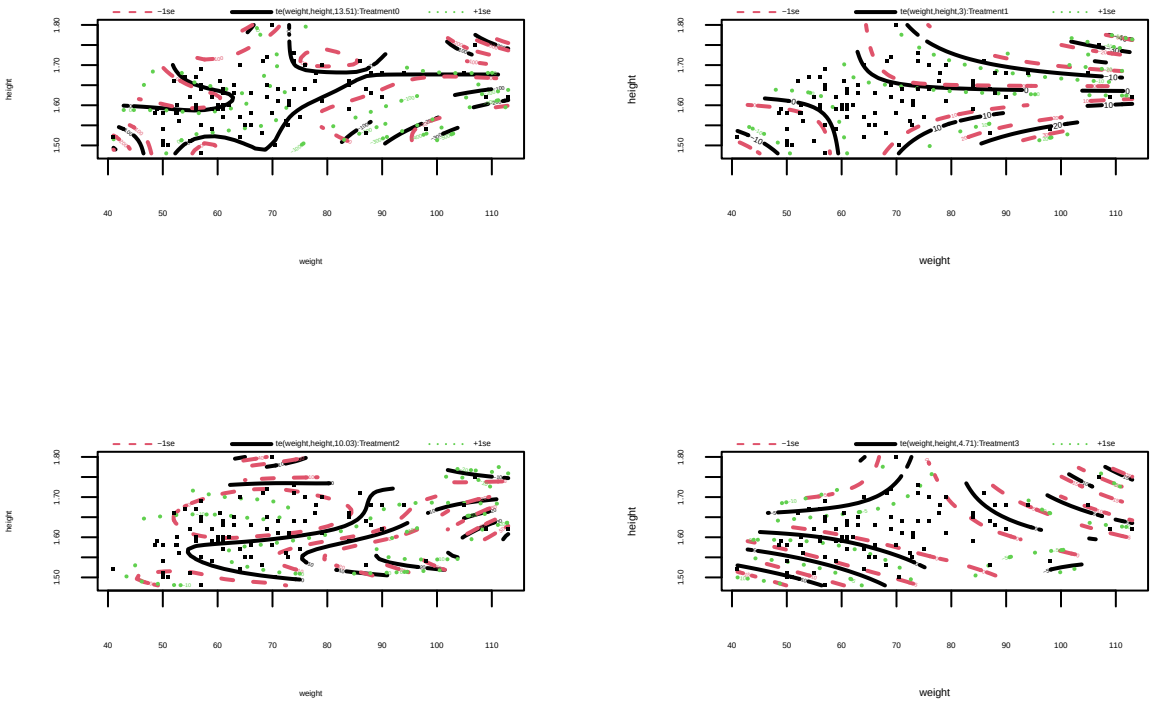
```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ FGm0 + Treatment + s(weight, by = Treatment) + s(height,
```

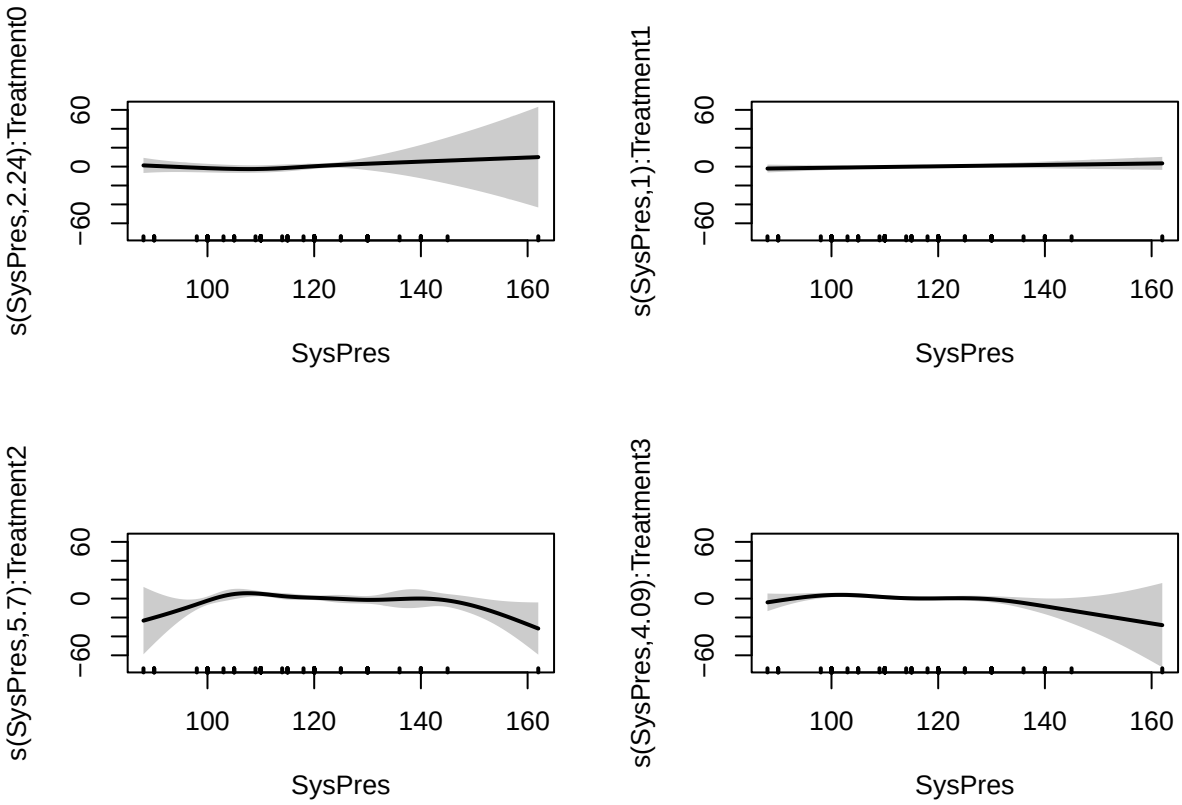
```
##     by = Treatment) + s(SysPres, by = Treatment) + s(DiaPres,
##     by = Treatment)
## Model 2: FGm12 ~ FGm0 + te(height, weight, by = Treatment) + te(DiaPres,
##     SysPres, by = Treatment)
## Model 3: FGm12 ~ FGm0 + te(weight, height, by = Treatment) + s(SysPres,
##     by = Treatment) + s(DiaPres, by = Treatment)
##   Resid. Df Resid. Dev      Df Deviance      F    Pr(>F)
## 1    47.104     629.95
## 2    25.504     205.00 21.6001   424.95 4.1056 0.0006534 ***
## 3    23.286     139.50  2.2182    65.50 6.1616 0.0058249 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

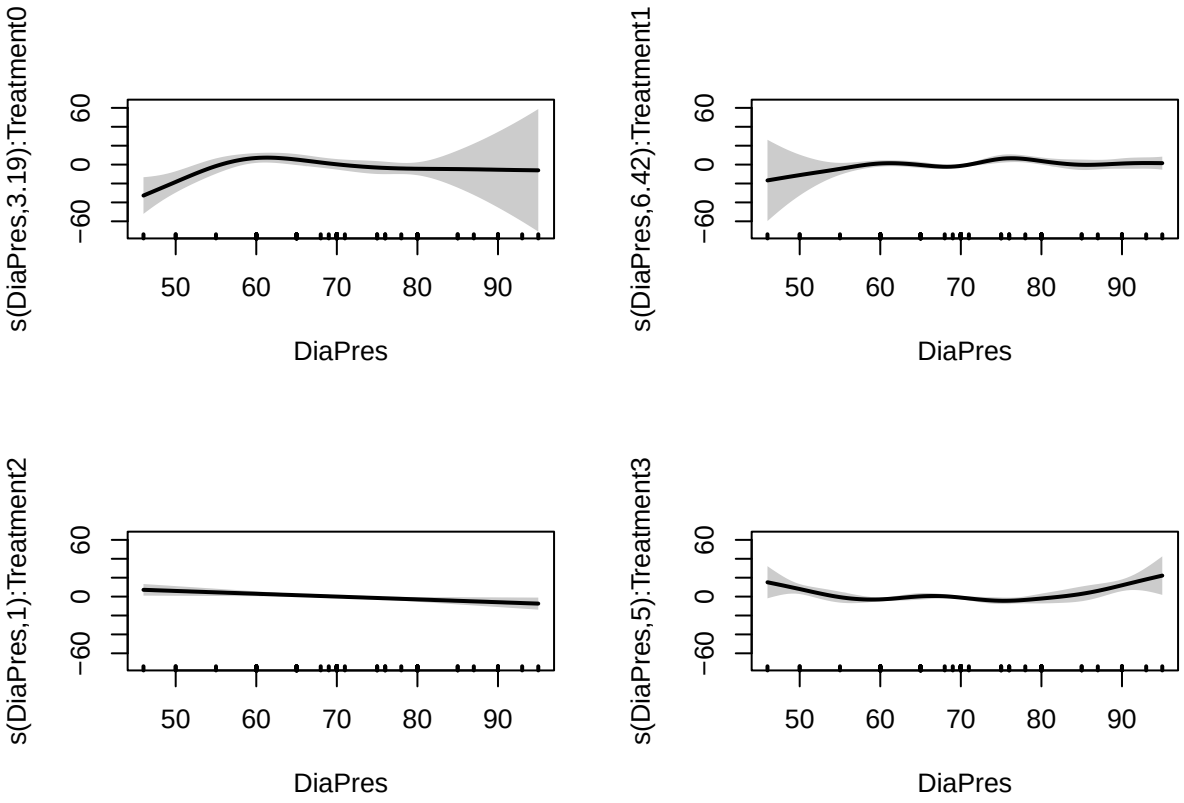
If the model 4 is included in the ANOVA analysis, the results still present a strange behavior. When it is subtracted, we can see that the results do coincide with the ones obtained so far. By taking these results into account together with the variance explanation and adjusted R coefficient of each model, our final selection will be the model3.

Model 3: Explanation

```
par(mfrow=c(2,2))
plot(gam3.0,pages=3, residuals=TRUE, shade=TRUE, cex=2, lwd=2)
```

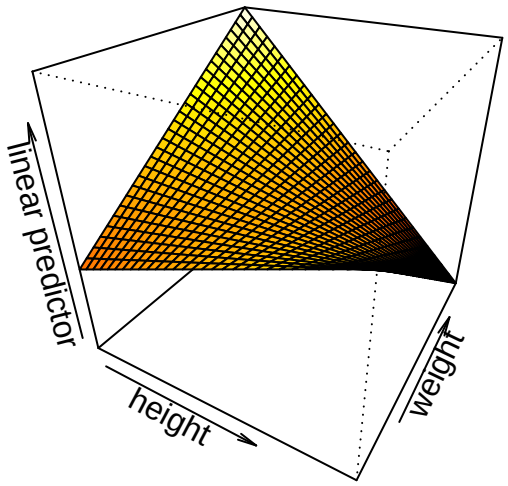



s(SysPres,2.24):Treatment0
s(SysPres,1):Treatment1
s(SysPres,5.7):Treatment2
s(SysPres,4.09):Treatment3
−60
0
60
100
120
140
160
SysPres



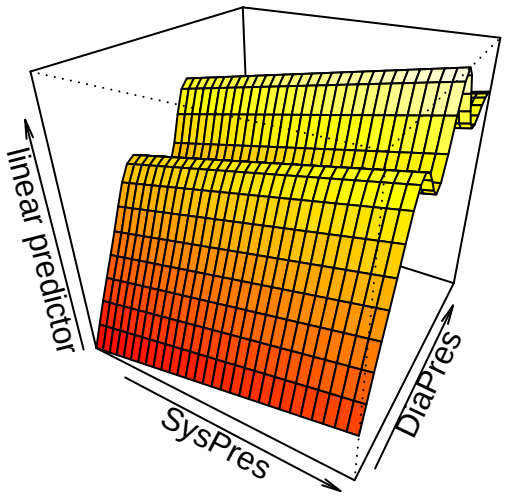
In the below representation of the model, it can be detected a clear relation between weight-height distribution and hirsutism. When high weights are distributed in less height (a sign that may correspond to overweight symptoms), a clear tendency to higher hirsutism values appears. This makes sense according to actual clinic researchments stating that obesity states can cause " **changes in the pattern of secretion or metabolism** *. On the contrary, hirsutism levels are lower when height and weight are more similarly distributed.

```
vis.gam(gam3.0, view=c("height","weight"), plot.type = "persp", theta=30, phi=30)
```



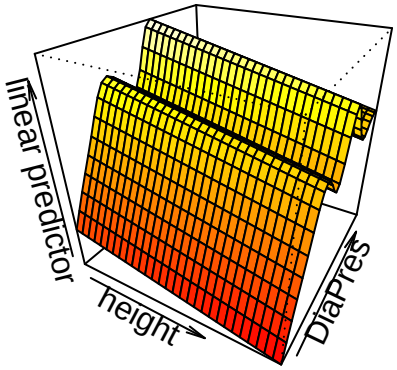
The below graphic representation presents an evident tendency for hirsutism regarding pressure variables. Systolis can Diastolic pressure do increase the levels of hirsutism, however, this relation is more complex than a simple linear increasment. It seems to present rapid increments, followed by little recess. It also can be observed that Diastolic pressure presents an stronger relation with hirsutism, since the slope of increase is more pronounced.

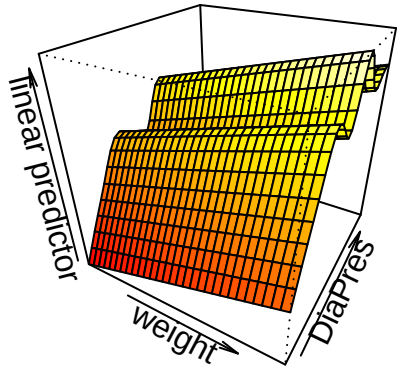
```
vis.gam(gam3.0, view=c("SysPres","DiaPres"), plot.type = "persp", theta=30, phi=30)
```



The relations between Diastolic pressure and height or weight present a behavior similar to the one described above, making it clear that this variable is relevant when considering hirsutism levels.

```
par(mfrow=c(1,2))
vis.gam(gam3.0, view=c("height","DiaPres"), plot.type = "persp", theta=30, phi=30)
vis.gam(gam3.0, view=c("weight","DiaPres"), plot.type = "persp", theta=30, phi=30)
```


linear predictor
height
DiaPres


linear predictor
weight
DiaPres