GAM Example Milan

```
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-31. For overview type 'help("mgcv-package")'.
```

```
library(nlme)
```

Here, we describe fitting a GAM to COVID-19 and environmental data from Milan, Italy. First, we read in the .csv file while omitting null values.

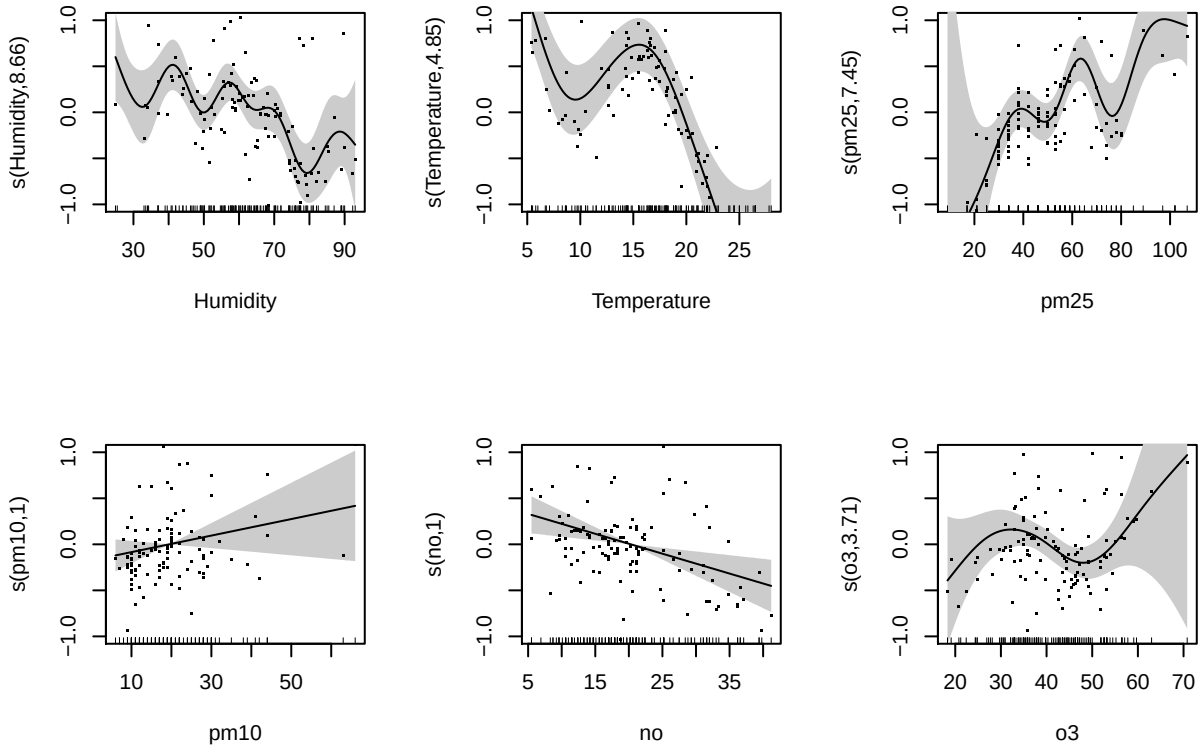
```
milan_full <- read.csv("milan_full.csv")
milan_full <- na.omit(milan_full)
```

Then, we create the GAM. RR relative rate of infection (the actual rate of infection divided by the average rate of infection), and we have created a three-day lag behind reported data. We want to see how humidity, temperature, PM2.5, PM10, NO2, and O3 influence the relative rate. We use a gaussian distribution for the GAM with a log link function.

```
infections.gam <- gam(LAGRR ~ s(Humidity) + s(Temperature) + s(pm25) + s(pm10) + s(no) + s(o3), family=g
```

We plot each of the individual splines, which show how each variable influences relative rate if all other variables are held constant.

```
plot(infections.gam, ylim=c(-1,1),scale=0,se=2, residuals=TRUE, shade=TRUE,pages=1)
```



Finally, we print a summary of the GAM's results.

```
summary(infections.gam)
```

```
##
## Family: gaussian
## Link function: log
##
## Formula:
## LAGRR ~ s(Humidity) + s(Temperature) + s(pm25) + s(pm10) + s(no) +
##     s(o3)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.3552     0.1277  -2.781  0.00646 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                  edf Ref.df      F  p-value
## s(Humidity)    8.662  8.949  4.091 0.000176 ***
## s(Temperature) 4.855  5.886 10.565 3.65e-09 ***
## s(pm25)        7.450  8.104  7.943 1.21e-08 ***
## s(pm10)        1.000  1.000  1.940 0.166663
## s(no)          1.000  1.000 10.133 0.001921 **
## s(o3)          3.714  4.637  2.388 0.052733 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.761   Deviance explained =   81%
## GCV = 0.30609  Scale est. = 0.24041   n = 129
```