

Latex document

Reto Zihlmann

November 2019

1 Model fitting and prediction

1.1 Statistical Model

The statistical model I use is a GAM which stand for generalized additive model. The model can be written as

$$g(E(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m)$$

1.2 Properties

There are two properties we might know from other models

Generalized The $g(\cdot)$ is the link function we know from generalised linear models

Additive The predictor is a linear combination of smooth functions $f_x()$ The smooth function makes GAM more general than generalized linear models but they do not allow automatically for interactions (no curse of dimensionality).

1.3 Frequency Data

Applying the model to the frequency data leads to

$$E(\log(Frequency)) = \beta_0 + f_{time}(time)$$

which means our $g(\cdot)$ is the identity function and we have only one smooth spline function for our predictor *time*.

1.4 Fitting

The model is fitted in R with the following comands

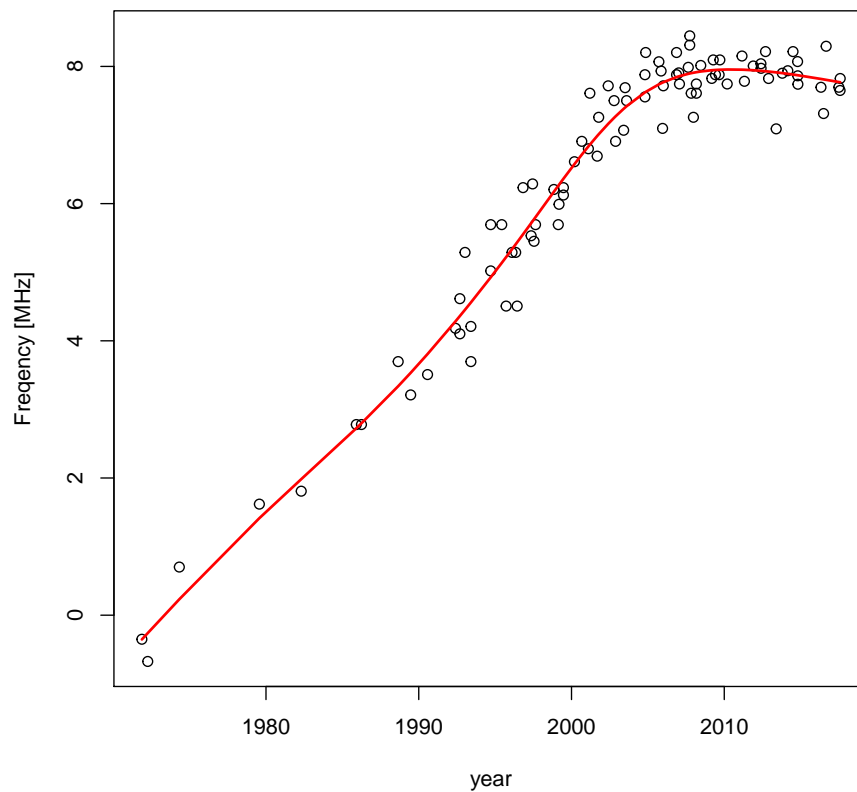
```
data <- read.table("frequency.dat")
names(data) <- c("year", "freq")
library(mgcv)
```

```
## Loading required package: nlme
## This is mgcv 1.8-29. For overview type 'help("mgcv-package")'.
fit <- gam(log(freq) ~ s(year), data = data)
```

1.5 Prediction

We can plot the fitted values and compare them to the observed values

```
plot(log(freq) ~ year, data = data, ylab = "Frequency [MHz]")
lines(data$year, predict(fit), col = "red", lwd = 2)
```



The summary function shows that the effect of year on Frequency is highly significant. The smoothing parameter is determined via cross validation. The smoothed line has 5.26 estimated degrees of freedom which is more than a simple linear regression.

```
summary(fit)

##
## Family: gaussian
## Link function: identity
##
## Formula:
## log(freq) ~ s(year)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.45383    0.03913    165    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df      F p-value
## s(year)  5.266  6.378 413.3  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.967   Deviance explained = 96.9%
## GCV = 0.14808   Scale est. = 0.13777    n = 90
```