EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Time series data and autocorrelation

Jacolien van Rij and Martijn Wieling
LSA 2015, Chicago | July 14

Time series data

- Examples:
  - weather & environment: temperature, precipitation, CO_2
  - financial: stock market analysis
  - psycholinguistic data:
    - EEG, eye tracking, articulography
    - RTs

Time series data

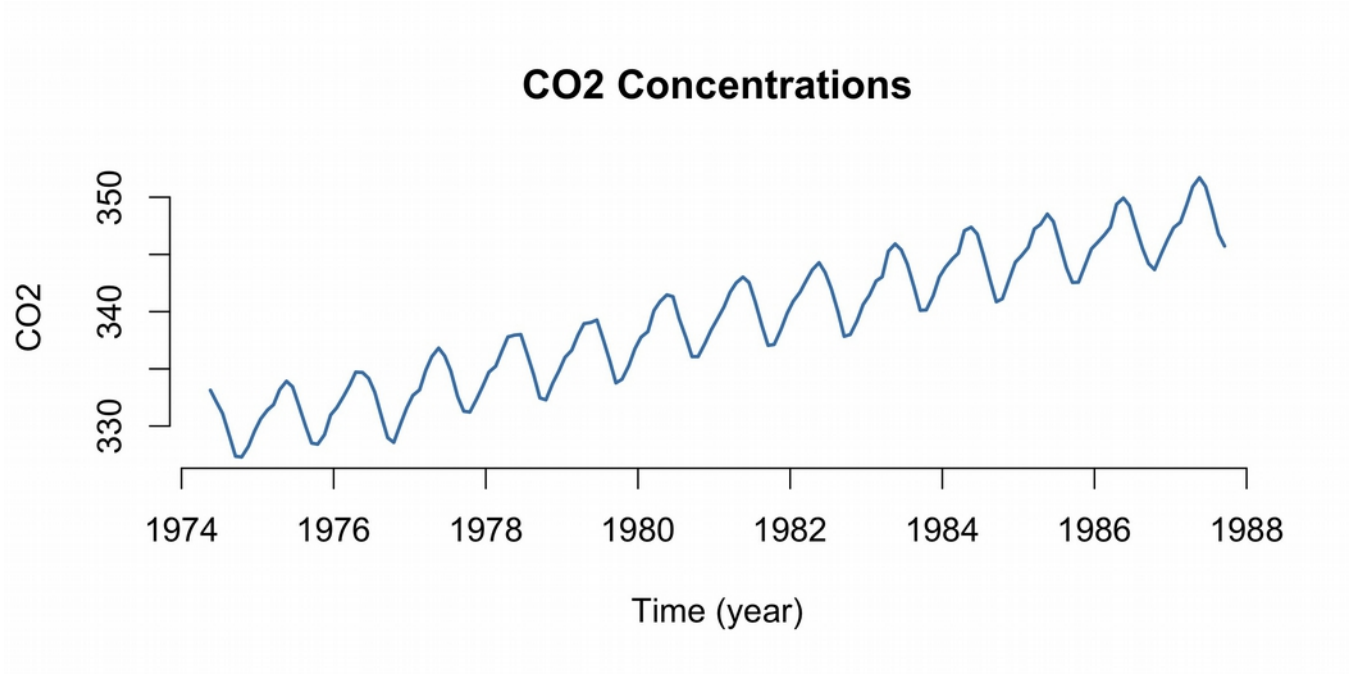
- Examples:
  - weather & environment: temperature, precipitation, CO_2
  - financial: stock market analysis
  - psycholinguistic data:
    - EEG, eye tracking, articulography
    - RTs
- The value of each measure is (partly) determined by the previous samples

CO2 example

```
> dat <- read.table('co2.txt', header=T, sep='\t')
> head(dat)
      CO2 YearMonth Year Month
1 333.13   1974.38 1974     5
2 332.09   1974.46 1974     6
3 331.10   1974.54 1974     7
4 329.14   1974.63 1974     8
5 327.36   1974.71 1974     9
6 327.29   1974.79 1974    10
```

CO2 example

```
> plot(dat$YearMonth, dat$CO2, type='l',
    xlab='Time (year)', ylab='CO2',
    main='CO2 Concentrations')
```



CO2 example

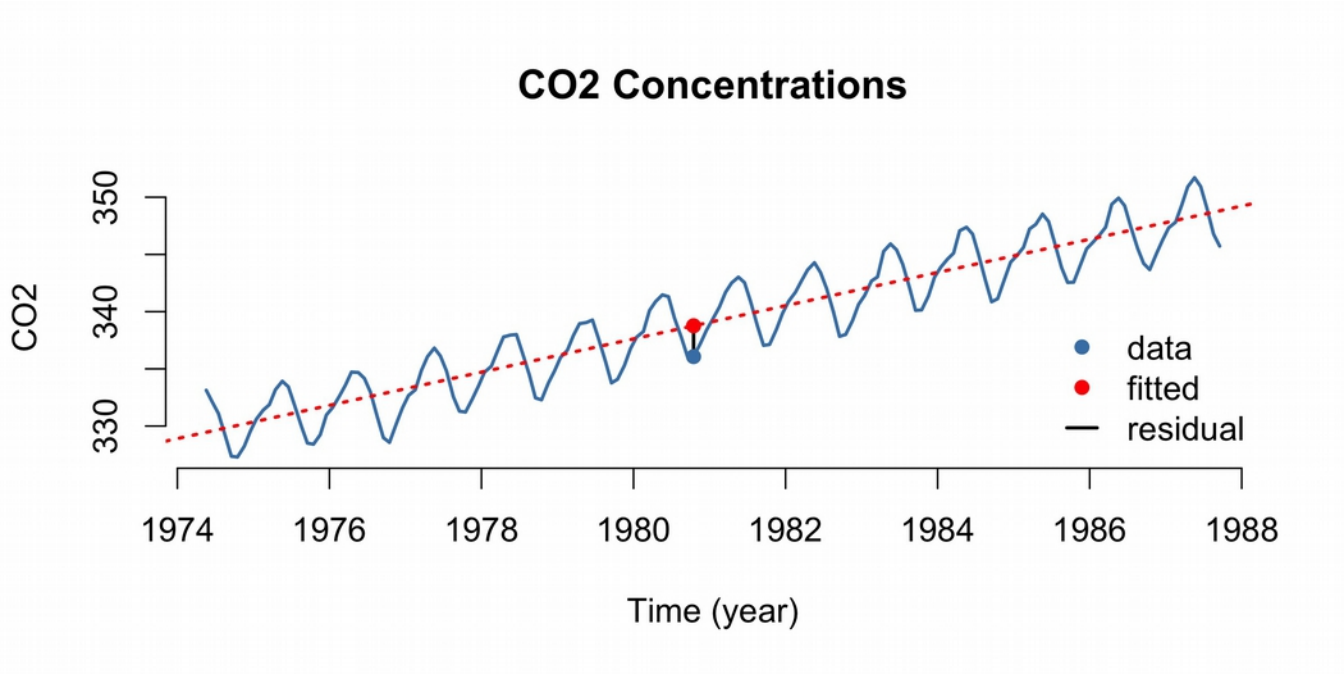
```
# linear regression:
> lm1 <- lm(CO2 ~ YearMonth, data=dat)

> round( summary(lm1)$coefficient, 3)

             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2537.189     88.117 -28.793        0
YearMonth       1.452      0.044  32.642        0
```

CO2 example

```
# linear regression:
> dat$Time <- dat$YearMonth - 1974

> lm1 <- lm(CO2 ~ Time, data=dat)

> round( summary(lm1)$coefficient, 3)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  328.906      0.358 919.941        0
Time           1.452      0.044  32.642        0

# CO2 = 328.906 + 1.452*(YearMonth - 1974)
```

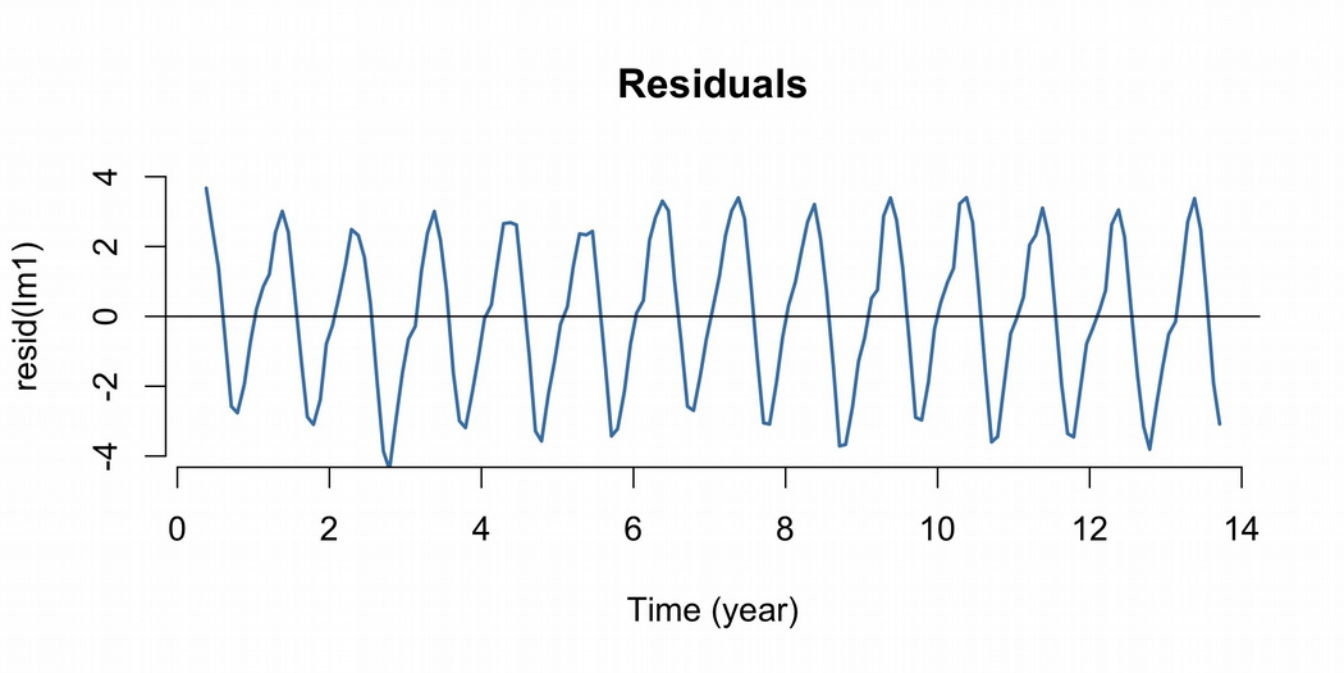
CO2 example

```
# CO2 = 328.906 + 1.452*(YearMonth - 1974)
```



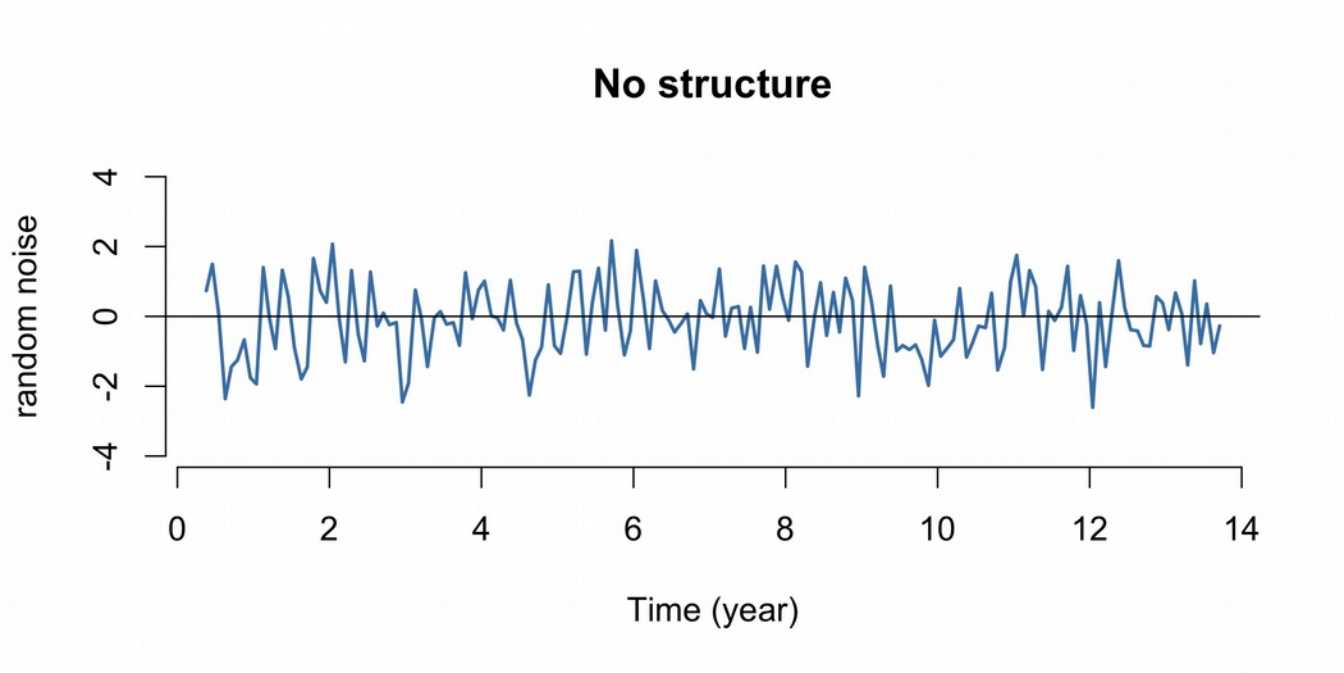
CO2 example

```
> plot(dat$Time, resid(lm1), type='l',
    xlab='Time (year)', ylab='resid(lm1)',
    main='Residuals')
```



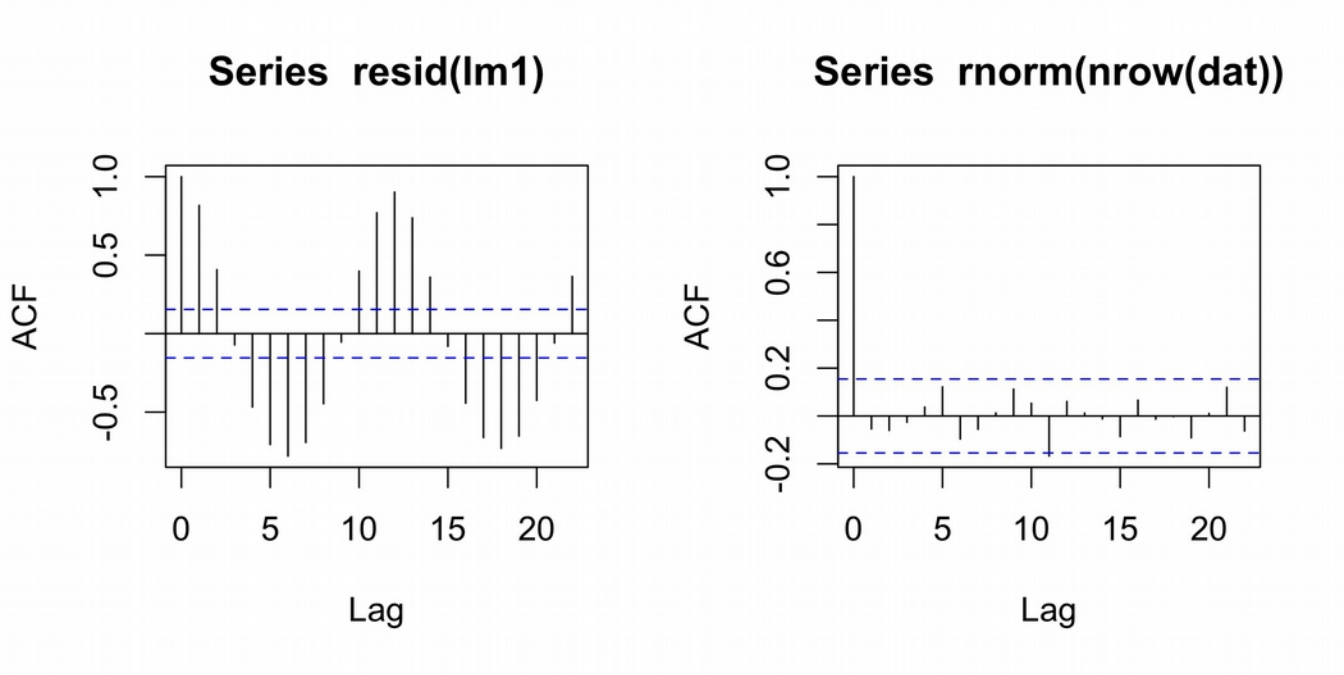
CO2 example

```
> plot(dat$Time, rnorm(nrow(dat)), type='l',
    xlab='Time (year)', ylab='Random noise',
    main='No structure')
```



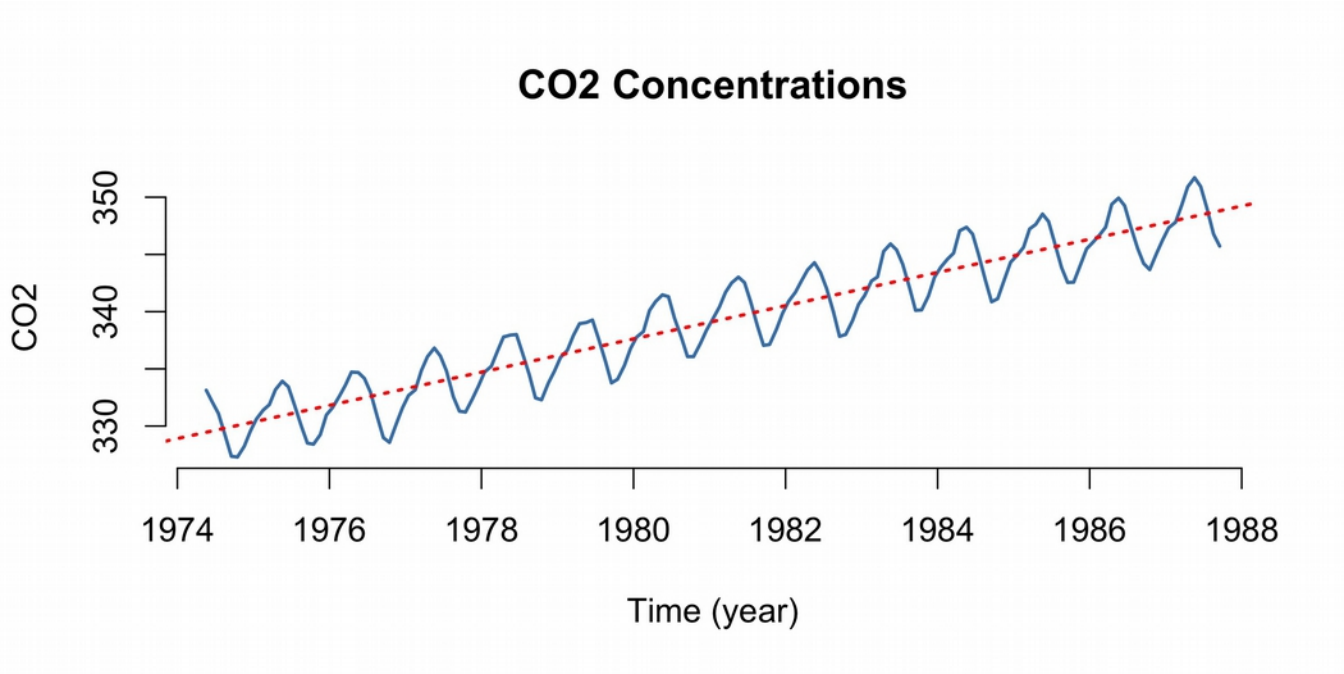
ACF plot

```
> par(mfrow=c(1,2), cex=1.2)
    acf(resid(lm1))
    acf(rnorm(nrow(dat)))
```



Reducing autocorrelation

- **Method 1:** Improving model fit

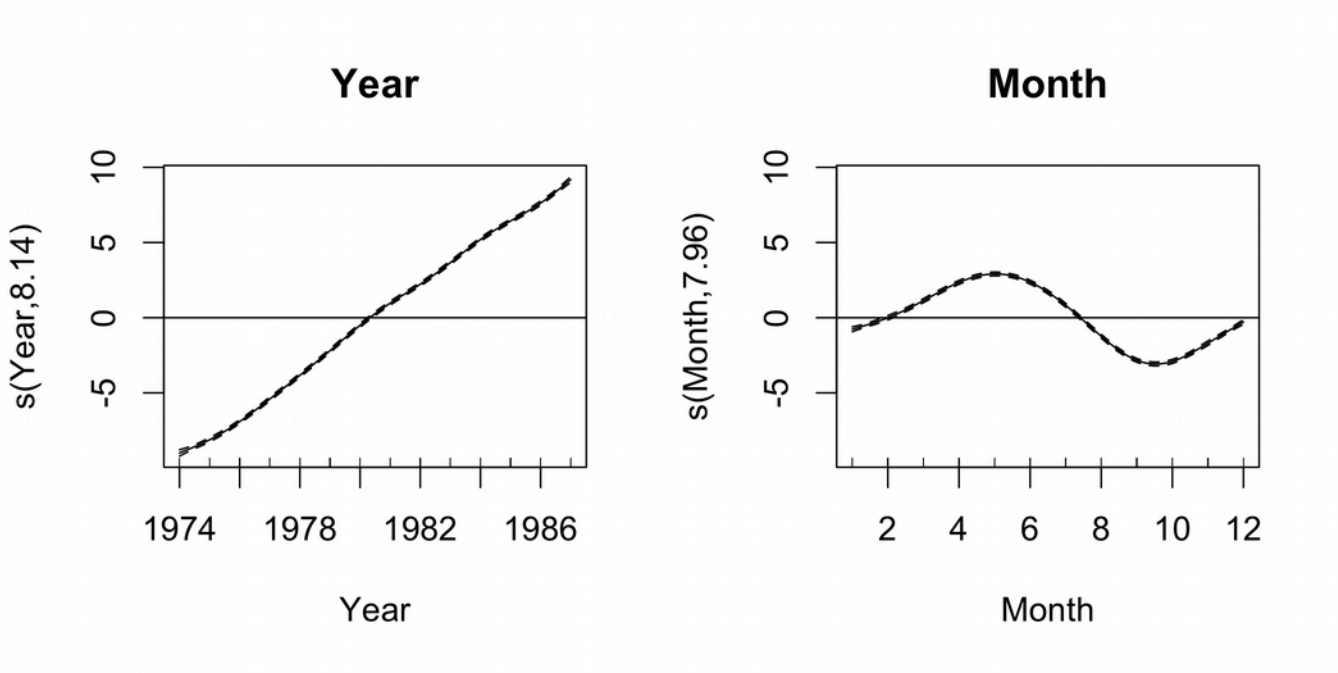


1. Improving model fit

```
> head(dat)
      CO2 YearMonth Year Month
1 333.13   1974.38 1974     5
2 332.09   1974.46 1974     6
3 331.10   1974.54 1974     7
4 329.14   1974.63 1974     8
5 327.36   1974.71 1974     9
6 327.29   1974.79 1974    10

> gam1 <- gam(CO2 ~ s(Year) + s(Month), data=dat)
```

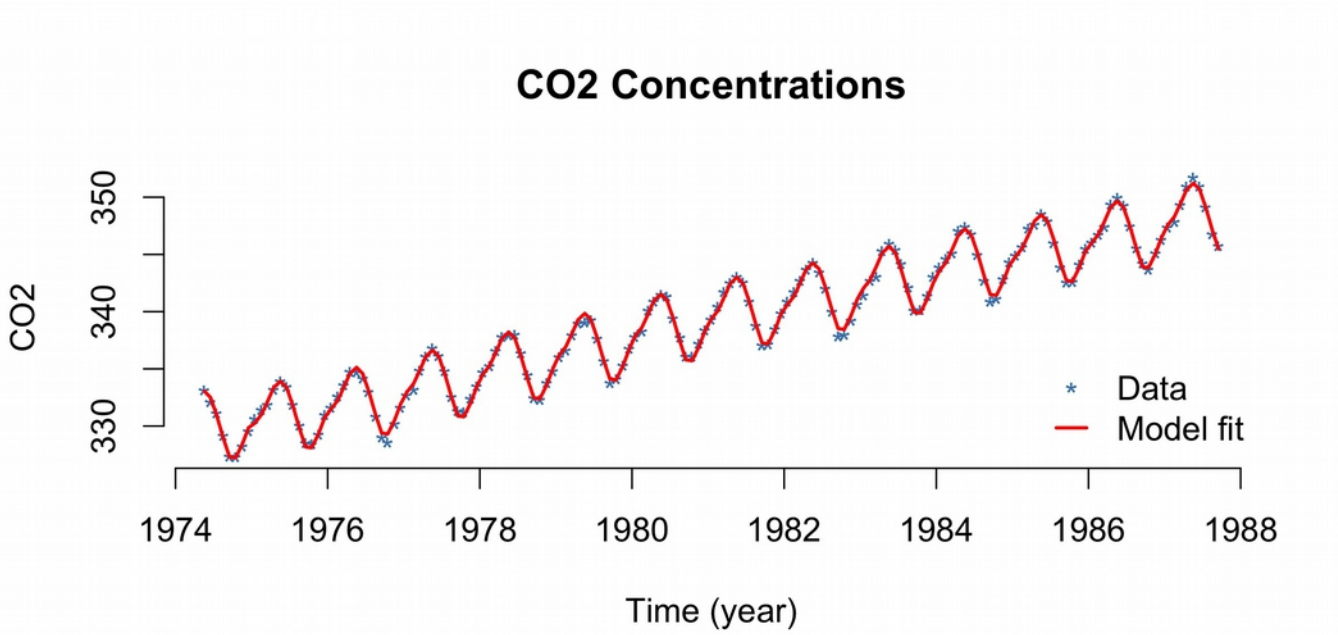
1. Improving model fit

```
plot( gam1 )
```



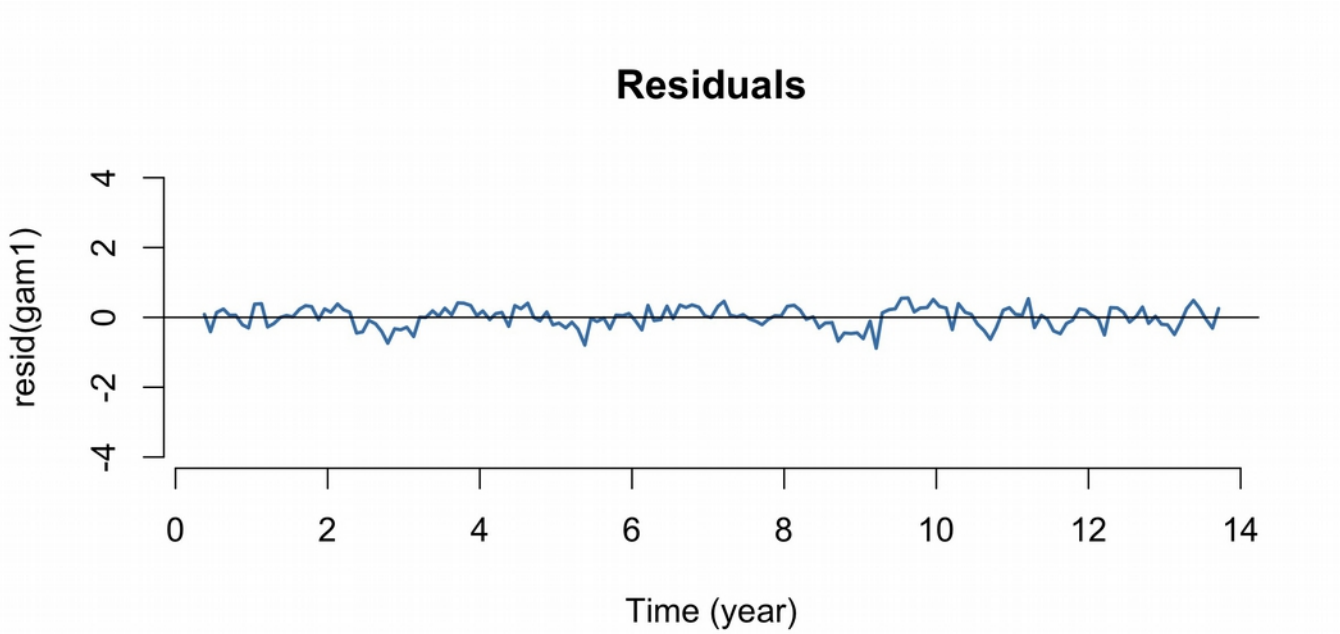
1. Improving model fit

```
> dat$fit <- fitted( gam1 )
> plot(dat$YearMonth, dat$fit, type='l')
```



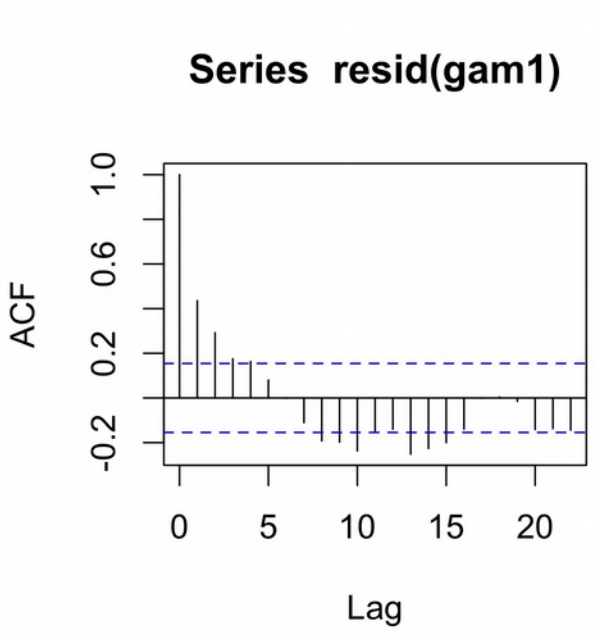
1. Improving model fit

```
> plot(dat$YearMonth, resid( gam1 ), type='l')
```

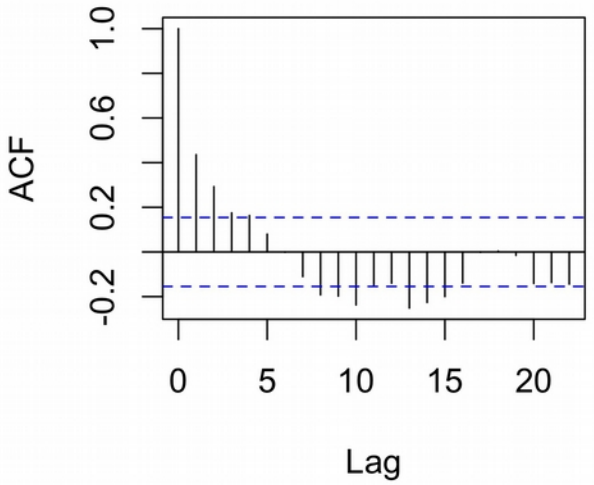


1. Improving model fit

```
> acf( resid( gam1 ) )
```



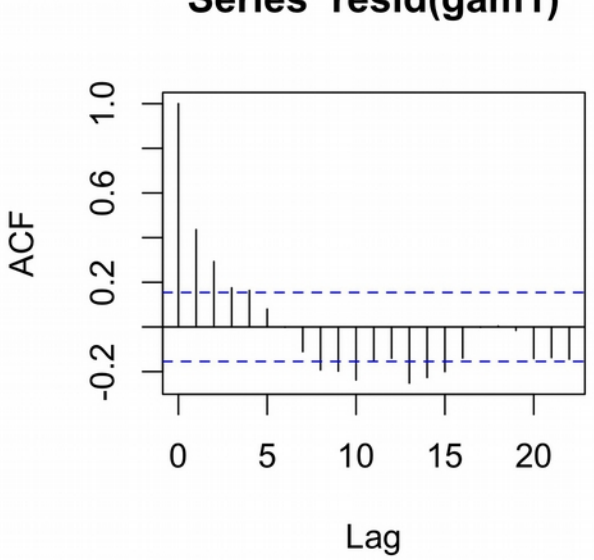
Reducing autocorrelation



- **Method 2:** Including AR(1) model
  - Taking into account that the residuals are correlated

- AR(ρ) model: autoregressive model of order ρ
  - AR(1) = $X_t = \varphi X_{t-1} + \epsilon_t$
  - AR(ρ) = $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_\rho X_{t-\rho} + \epsilon_t$

Reducing autocorrelation



- **Method 2:** Including AR(1) model
  - Taking into account that the residuals are correlated

- AR(ρ) model: autoregressive model of order ρ
  - AR(1) = $X_t = \varphi X_{t-1} + \epsilon_t$

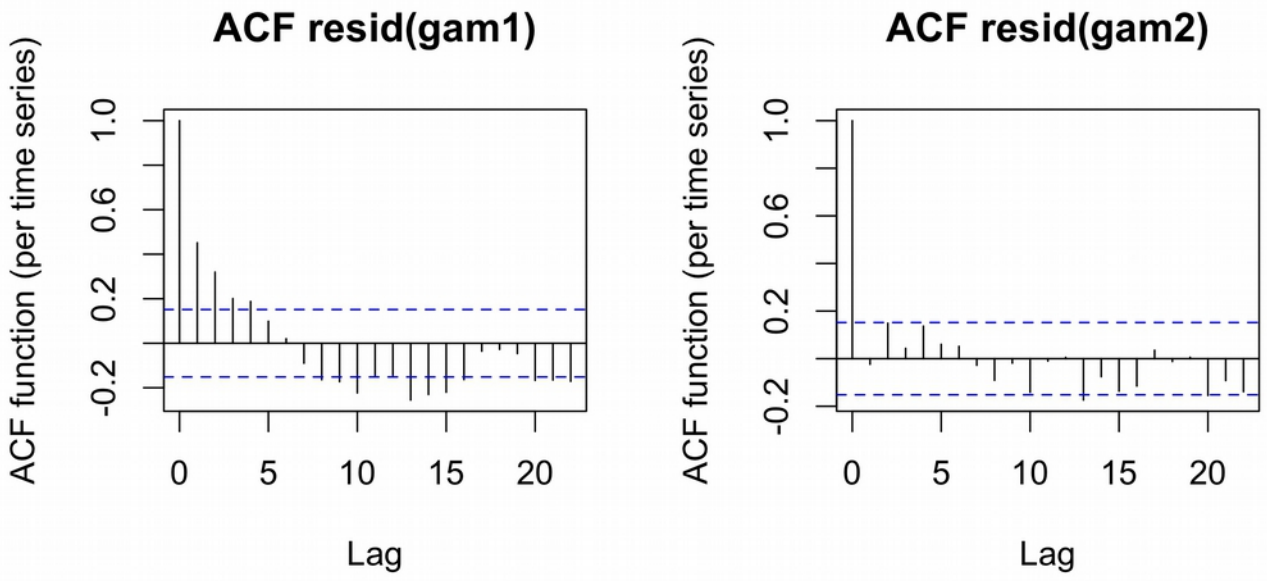
- Including AR(1) model in GAMM
  - $y_t = f(Time_t) + \epsilon_t \longrightarrow \epsilon_t = \rho \cdot \epsilon_{t-1} + noise$

2. Including AR(1) model

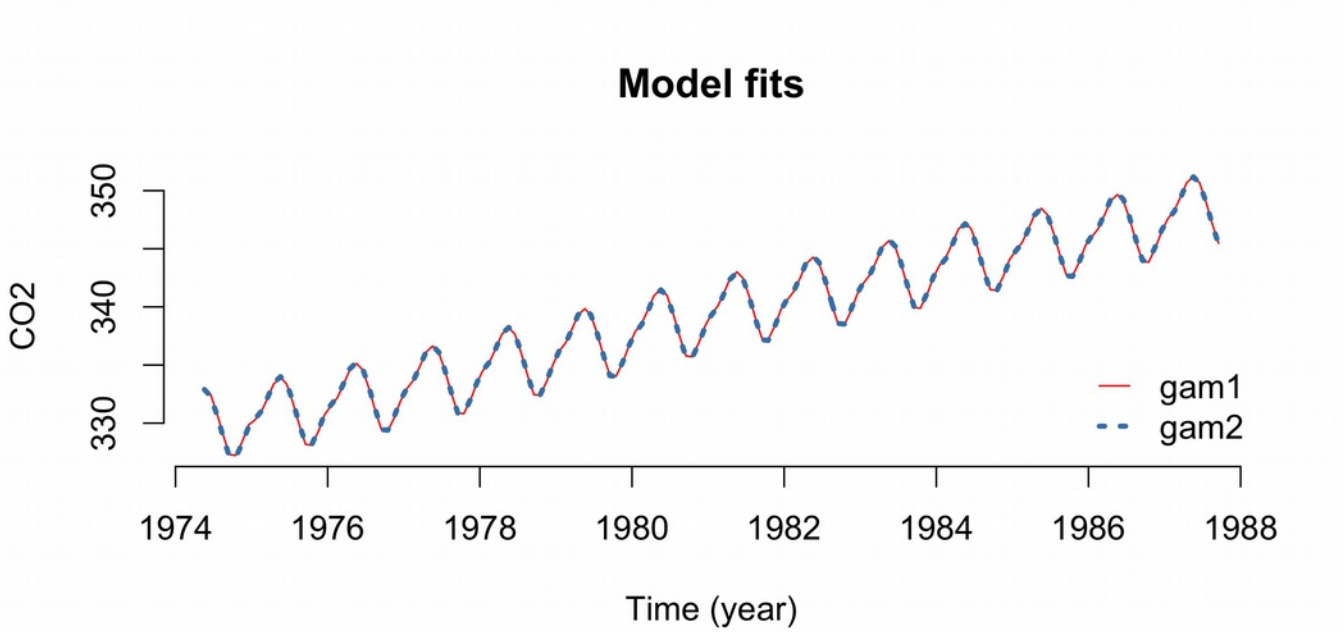
```
gam1 <- gam(CO2 ~ s(Year) + s(Month), data=dat)

# find the autocorrelation value at Lag 1:
> acf(resid(gam1), plot=FALSE)$acf[2]
[1] 0.4515055

# run new model with autocorrelation parameter:
> dat$start <- c(TRUE, rep(FALSE, nrow(dat)-1) )

> gam3 <- bam(CO2 ~ s(Year) + s(Month), data=dat,
     rho=0.45, AR.start=dat$start )
```

2. Including AR(1) model

```
> library(itsadug)
> acf_resid(gam1, main='ACF resid(gam1)')
> acf_resid(gam2, main='ACF resid(gam2)')
```

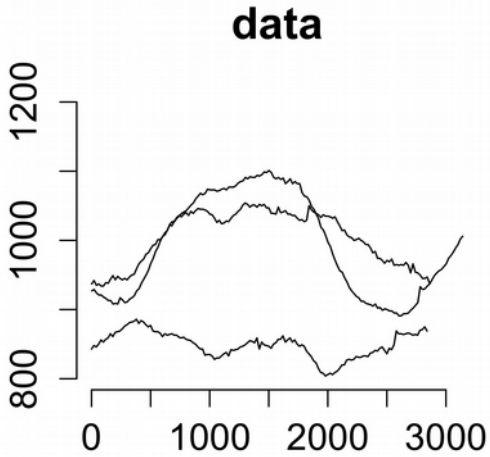


2. Including AR(1) model



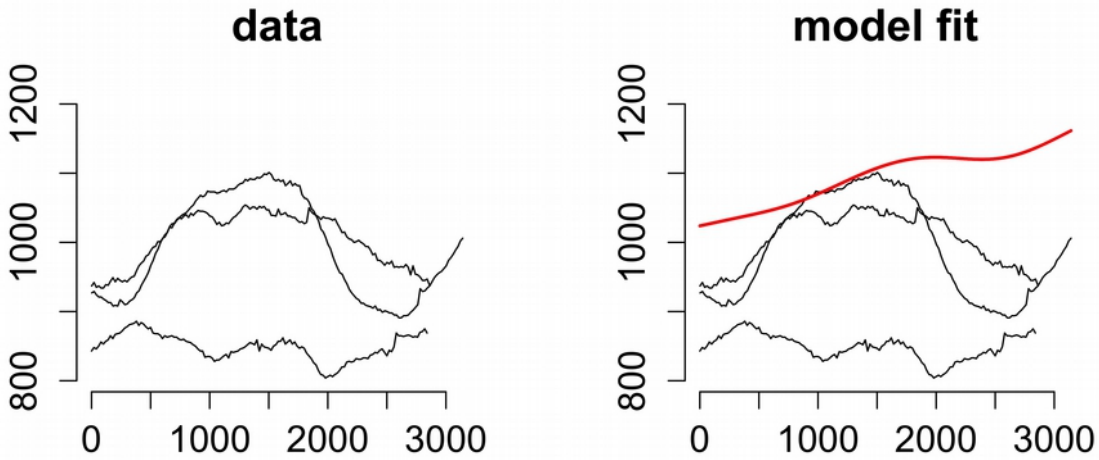
Autocorrelation in pupil data

```
m1 <- bam(Pupil ~ s(Time), data=dat)
```



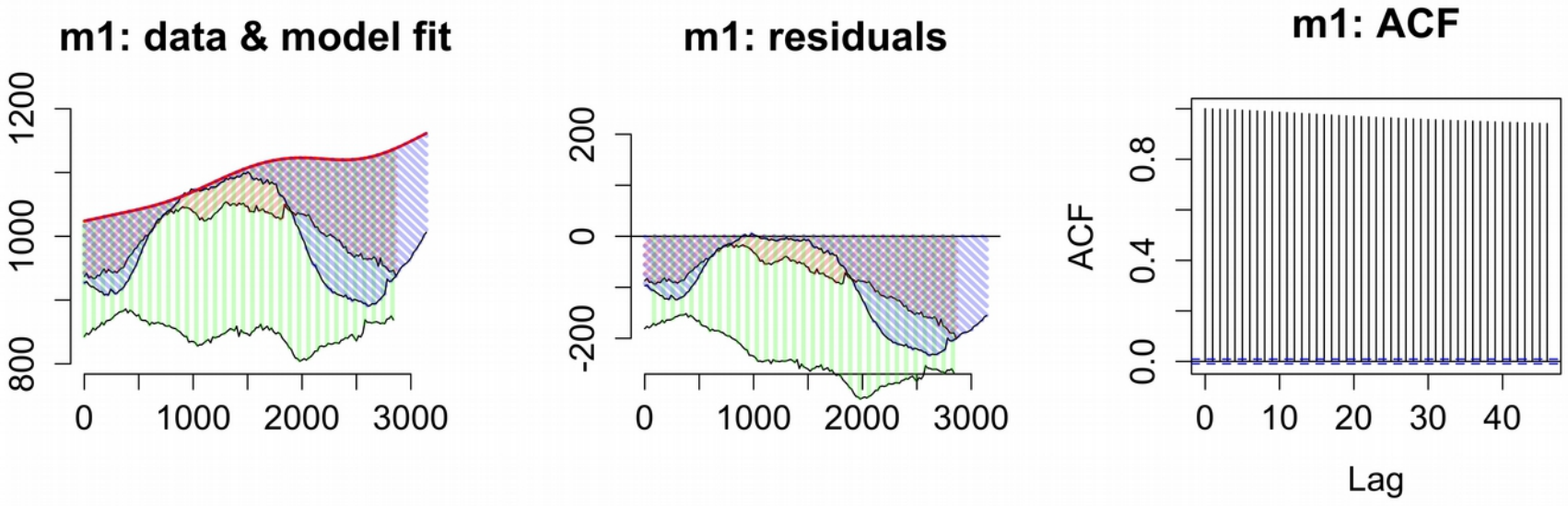
Autocorrelation in pupil data

```
m1 <- bam(Pupil ~ s(Time), data=dat)
```



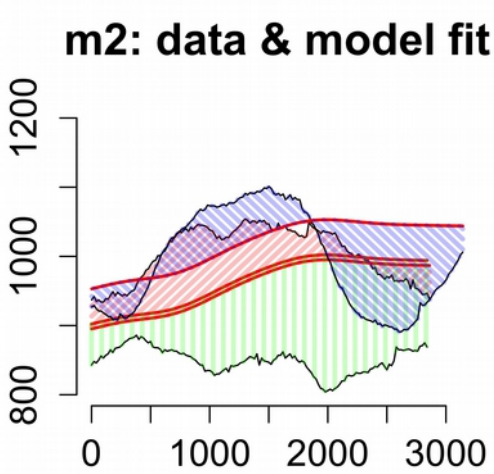
Autocorrelation in pupil data

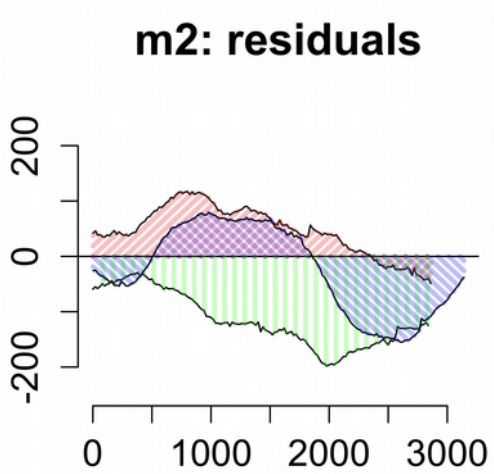
```
m1 <- bam(Pupil ~ s(Time), data=dat)
```

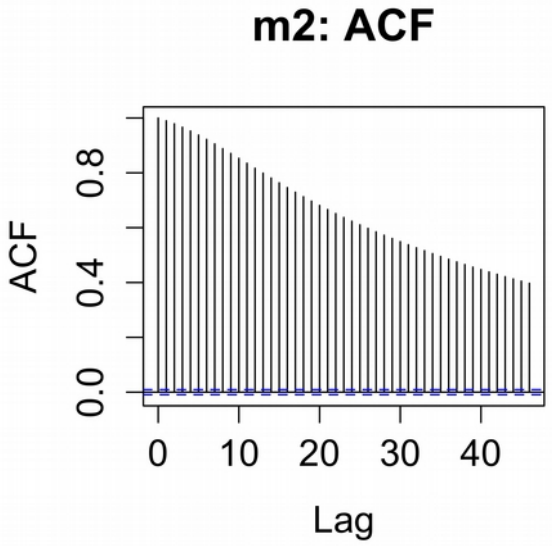


Autocorrelation in pupil data

```
m1 <- bam(Pupil ~ s(Time)
+ s(Subject, bs='re') + s(Item, bs='re'),
data=dat)
```

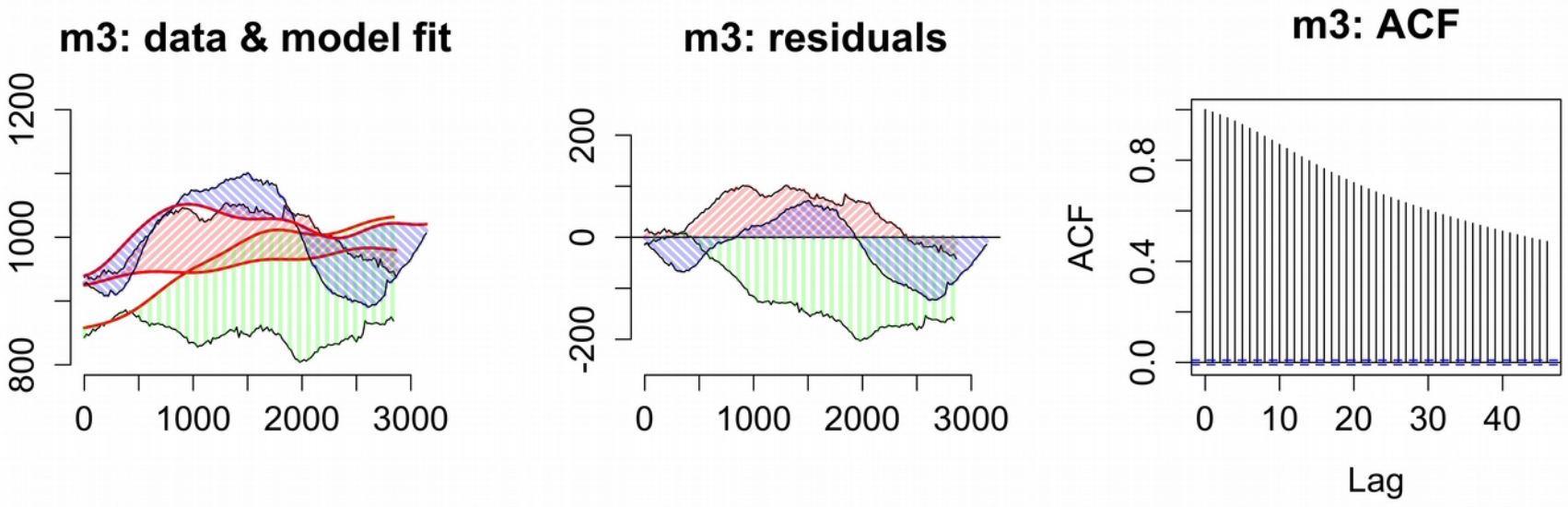






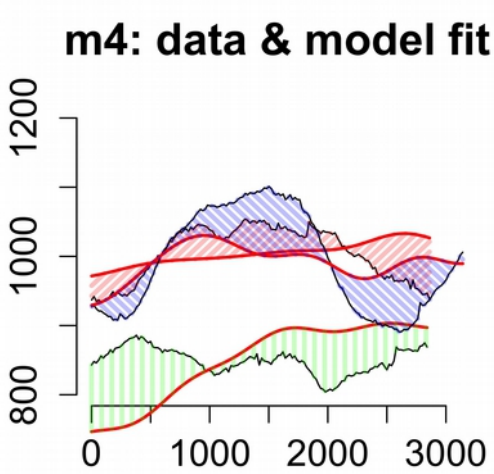
Autocorrelation in pupil data

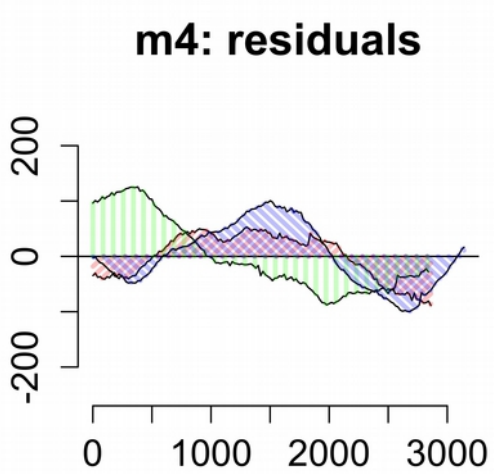
```
m3 <- bam(Pupil ~ s(Time)
+ s(Time, Subject, bs='fs', m=1)
+ s(Time, Item, bs='fs', m=1), data=dat)
```

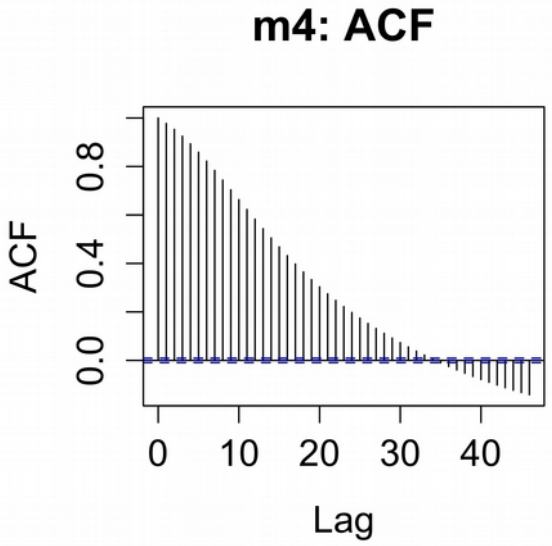


Autocorrelation in pupil data

```
m4 <- bam(Pupil ~ s(Time)
+ s(Time, Subject, bs='fs', m=1)
+ s(Time, Item, bs='fs', m=1)
+ s(Event, bs='re'), data=dat)
```







Autocorrelation in pupil data

```
m4 <- bam(Pupil ~ s(Time)
+ s(Time, Subject, bs='fs', m=1)
+ s(Time, Item, bs='fs', m=1)
+ s(Event, bs='re'), data=dat)
```