

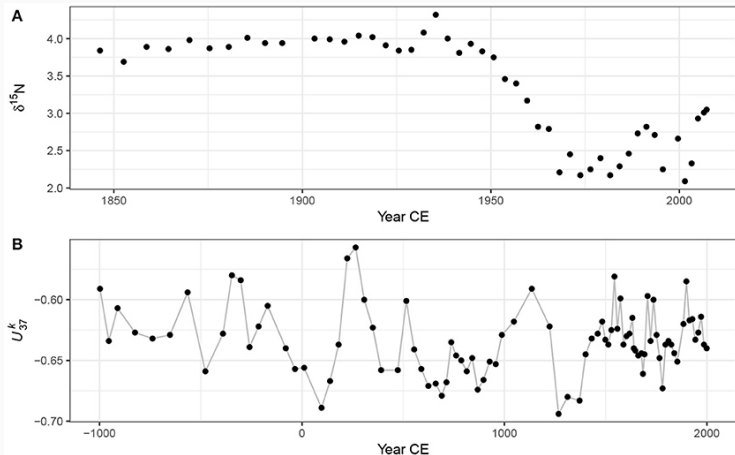
# Generalised Additive Models

---

Francisco Rodríguez-Sánchez

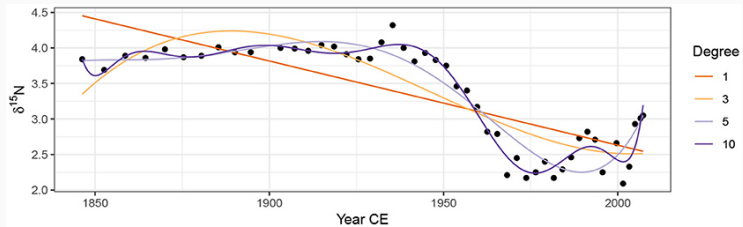
<https://frodriguezsanchez.net>

# How do we model these time series?



Simpson 2018

# How do we model these time series?



Simpson 2018

# GAMs allow us to model non-linear relationships using smooths

Generalised Linear Model (GLM):

$$y = a + bx$$

Generalised Additive Model (GAM):

$$y = a + s(x)$$

## Modelling non-linear time series with GAM

```
isotopes <- readRDS("data/isotope.rds")
```

	Depth	d13C	TotalC	d15N	TotalN	DryWeight	Year
1	0.2	-27.57	806.49	3.05	64.21	8.2	2007.254
2	0.4	-27.67	949.33	3.01	73.26	7.6	2006.510
3	0.8	-27.63	1305.52	2.93	93.25	11.6	2004.941
4	1.2	-27.62	1136.04	2.33	86.09	9.6	2003.269
5	1.6	-27.48	1028.27	2.09	93.80	10.9	2001.496
6	2.0	-27.39	809.91	2.66	79.98	9.9	1999.626

# Modelling non-linear time series with GAM

```
library("mgcv")  
m <- gam(d15N ~ s(Year, k = 15), data = isotopes, method = "REML")
```

Family: gaussian

Link function: identity

Formula:

d15N ~ s(Year, k = 15)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.30958	0.02622	126.2	<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)	9.282	11.07	61.33	<2e-16 ***

---

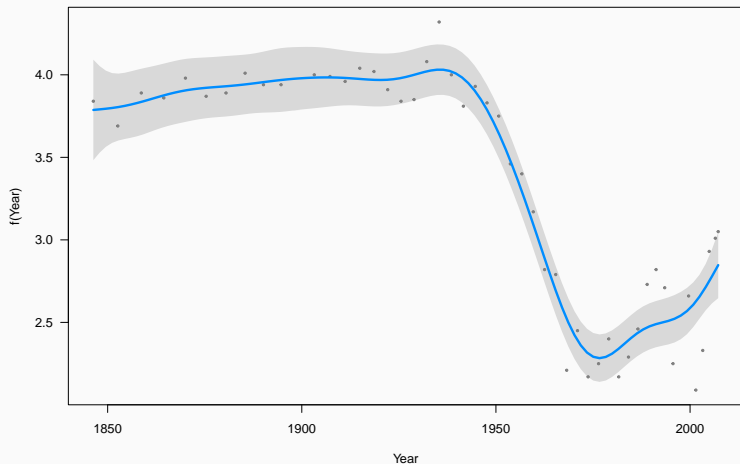
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.935 Deviance explained = 94.8%

-REML = 3.9734 Scale est. = 0.03299 n = 48

# Visualising fitted GAM

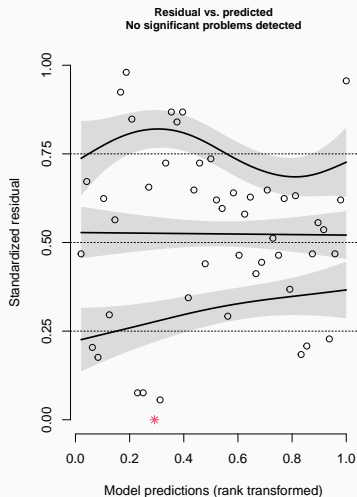
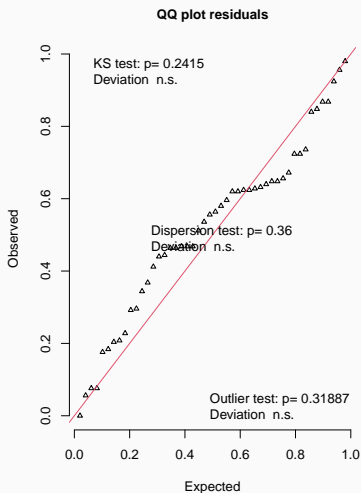
```
visreg(m)
```



# Checking fitted GAM

```
library("DHARMa")  
simulateResiduals(m, plot = TRUE)
```

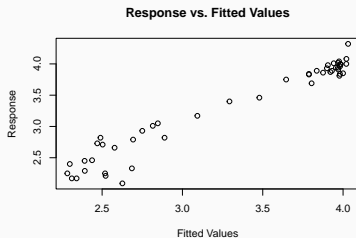
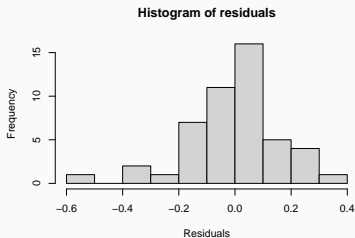
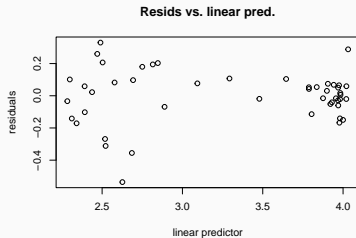
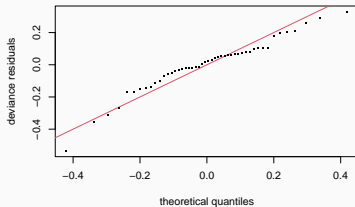
DHARMa residual diagnostics





# Checking fitted GAM

```
gam.check(m)
```



## Including temporal autocorrelation

```
mod <- gamm(d15N ~ s(Year, k = 15), data = isotopes,  
            correlation = corCAR1(form = ~ Year), method = "REML")
```

Family: gaussian

Link function: identity

Formula:

d15N ~ s(Year, k = 15)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.30909	0.03489	94.84	<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)	7.954	7.954	47.44	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.929

## Modelling infant mortality

---

# Modelling infant mortality

```
mort <- read.csv("data/UN_GDP_infantmortality.csv")
```

	country	infant.mortality	gdp
1	Afghanistan	154	2848
2	Albania	32	863
3	Algeria	44	1531
4	American.Samoa	11	NA
5	Andorra	NA	NA
6	Angola	124	355

# Modelling infant mortality with a GLM

```
library("MASS")  
mort.glm <- glm.nb(infant.mortality ~ gdp, data = mort)
```

Call:

```
glm.nb(formula = infant.mortality ~ gdp, data = mort, init.theta = 2.460991808,  
       link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.8024	-1.0447	-0.3650	0.5232	2.9116

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.072e+00	5.727e-02	71.11	<2e-16 ***
gdp	-8.675e-05	6.221e-06	-13.95	<2e-16 ***

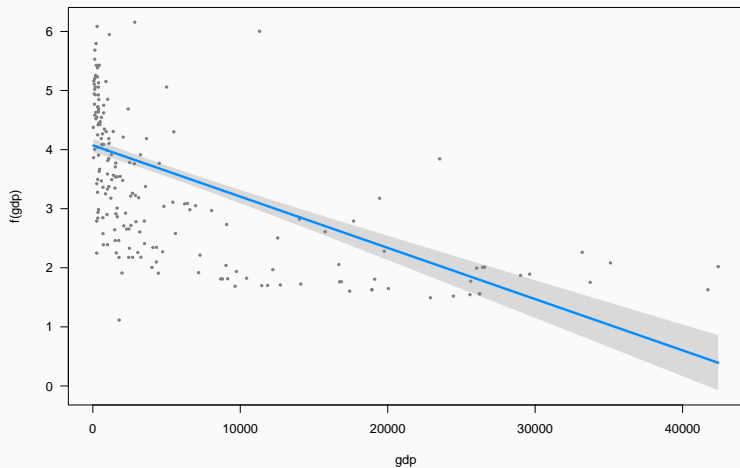
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(2.461) family taken to be 1)

Null deviance: 385.83 on 192 degrees of freedom  
Residual deviance: 202.51 on 191 degrees of freedom

# Modelling infant mortality with a GLM



# Modelling infant mortality with a GLM (log.gdp)

```
mort$log.gdp <- log(mort$gdp)
mort.glm.log <- glm.nb(infant.mortality ~ log.gdp, data = mort)
```

Call:

```
glm.nb(formula = infant.mortality ~ log.gdp, data = mort, init.theta = 3.119314453,
link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7517	-0.8692	-0.3575	0.3090	4.5063

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	7.07818	0.20045	35.31	<2e-16 ***
log.gdp	-0.47238	0.02647	-17.85	<2e-16 ***

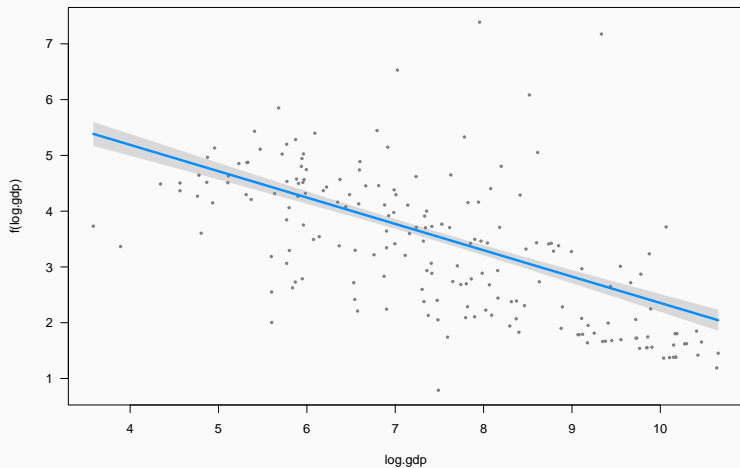
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(3.1193) family taken to be 1)

Null deviance: 478.54 on 192 degrees of freedom  
Residual deviance: 198.03 on 191 degrees of freedom

## Modelling infant mortality with a GLM (log.gdp)





# Modelling infant mortality with a GAM

```
library("mgcv")  
mort.gam <- gam(infant.mortality ~ s(log.gdp), family = nb, data = mort)
```

Family: gaussian

Link function: identity

Formula:

d15N ~ s(Year, k = 15)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.30958	0.02622	126.2	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

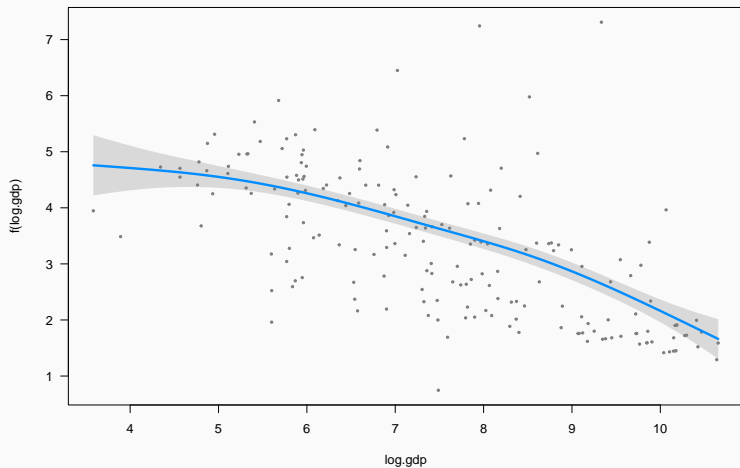
	edf	Ref.df	F	p-value
s(Year)	9.282	11.07	61.33	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

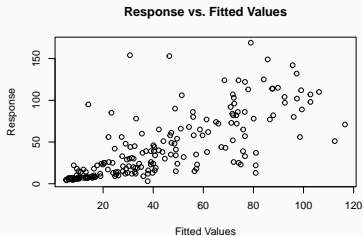
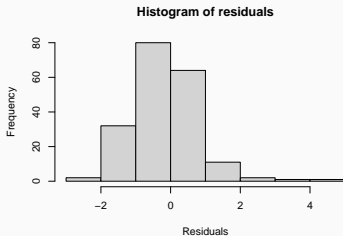
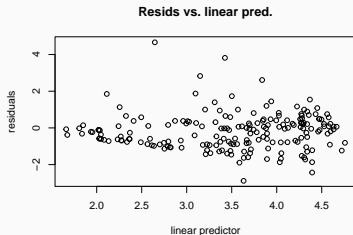
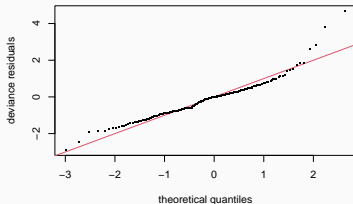
R-sq.(adj) = 0.935 Deviance explained = 94.8%

# Modelling infant mortality with a GAM



# Checking GAM

```
gam.check(mort.gam)
```



Method: REML    Optimizer: outer newton

# Comparing models

```
library("performance")  
compare_performance(mort.glm, mort.glm.log, mort.gam)
```

# Comparison of Model Performance Indices

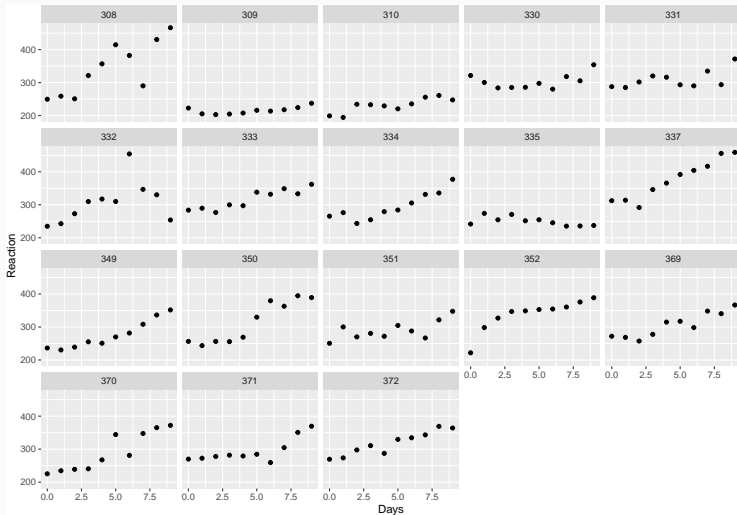
Name	Model	AIC	BIC	RMSE	Sigma	Score
mort.glm	negbin	1714.957	1724.745	31.089	1.030	-
mort.glm.log	negbin	1667.750	1677.538	30.034	1.018	-
mort.gam	gam	1661.141	1680.512	26.249	1.027	-

## Generalised Additive Mixed Models (GAMM)

---

# Reaction time with sleep deprivation

```
library("lme4")  
data("sleepstudy")
```



# Modelling reaction time with sleep deprivation (GAMM)

```
sgamm <- gam(Reaction ~ s(Days, Subject, k = 3, bs = "fs"),  
             data = sleepstudy, method = "REML")
```

Family: gaussian

Link function: identity

Formula:

Reaction ~ s(Days, Subject, k = 3, bs = "fs")

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	295.22	10.49	28.15	<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(Days,Subject)	42.2	53	16.05	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.826 Deviance explained = 86.7%

-REML = 886.71 Scale est. = 551.61 n = 180

# Modelling reaction time with sleep deprivation (GAMM)

```
visreg(sgamma, xvar = "Days", by = "Subject")
```

