# Estimating rates and magnitudes of temporal change

## Using generalized additive models with stratigraphic records

Gavin Simpson

June 19th, 2018

# Funding

NSERC
CRSNG

# Palaeo data *are* time series

**Interest in changes in the data over time — implies the estimation of trends in data**

Commonly, trend detection involves eye-balling the data

Fundamentally irreproducible — poor science

Conflates **signal** and **noise**

Victor Freitas

# Spurious correlation

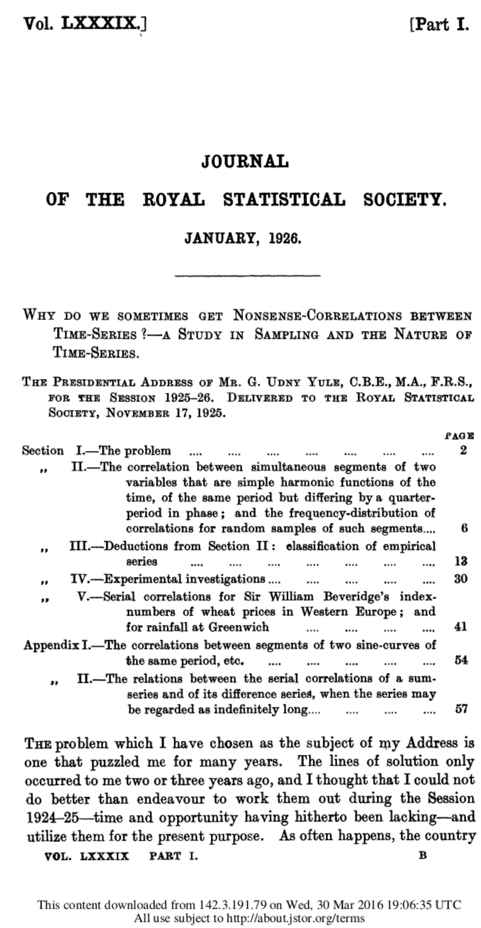
## Don't correlate time series

Known for almost a century this is bad

Non-parametric doesn't mean no assumptions — still needs i.i.d. data

Only testing for restricted range of trend shapes

- linear (Pearson correlation)
- monotonic (rank correlation)


Vol. LXXXIX.] [Part I.

JOURNAL

OF THE ROYAL STATISTICAL SOCIETY.

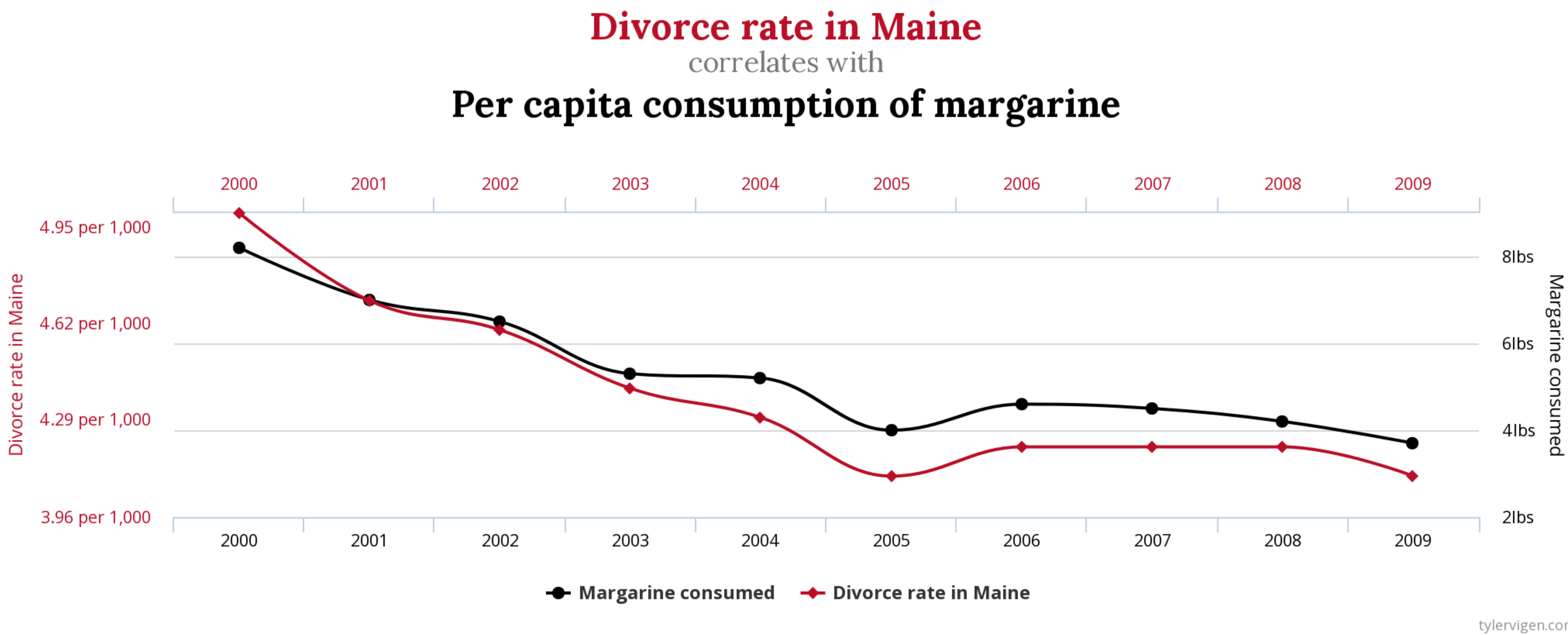
JANUARY, 1926.

WHY DO WE SOMETIMES GET NONSENSE-CORRELATIONS BETWEEN TIME-SERIES?—A STUDY IN SAMPLING AND THE NATURE OF TIME-SERIES.

THE PRESIDENTIAL ADDRESS OF MR. G. UDNY YULE, C.B.E., M.A., F.R.S., FOR THE SESSION 1925–26. DELIVERED TO THE ROYAL STATISTICAL SOCIETY, NOVEMBER 17, 1925.



THE problem which I have chosen as the subject of my Address is one that puzzled me for many years. The lines of solution only occurred to me two or three years ago, and I thought that I could not do better than endeavour to work them out during the Session 1924–25—time and opportunity having hitherto been lacking—and utilize them for the present purpose. As often happens, the country




Source: Tyler Viglen http://www.tylervigen.com/spurious-correlations

# Loess must die

Loess is a simple scatterplot smoother
— fit is controlled by the *span*

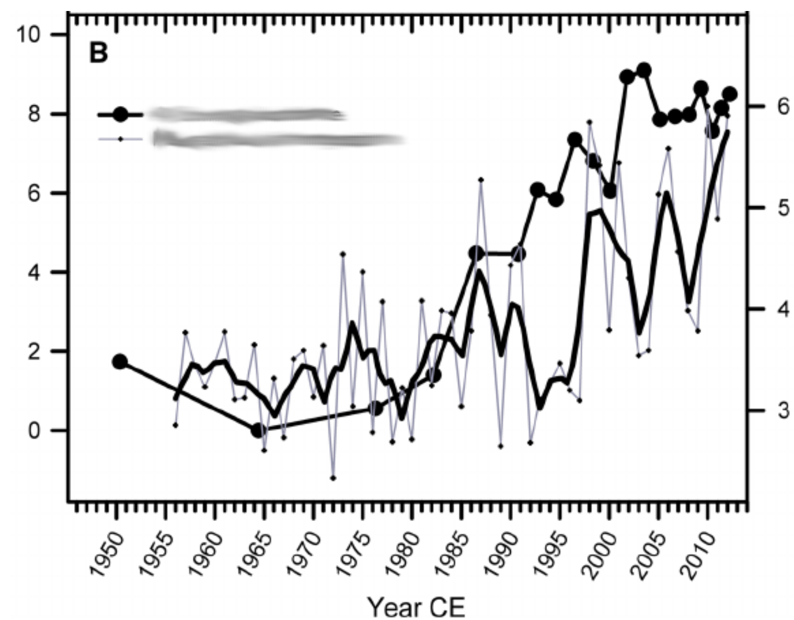
Must choose *span* appropriately

Difficult (impossible?) to do properly...

As a result often chosen subjectively!

Plain wrong — don't do it!

Don't use Loess for inference ...
exploratory data analysis only

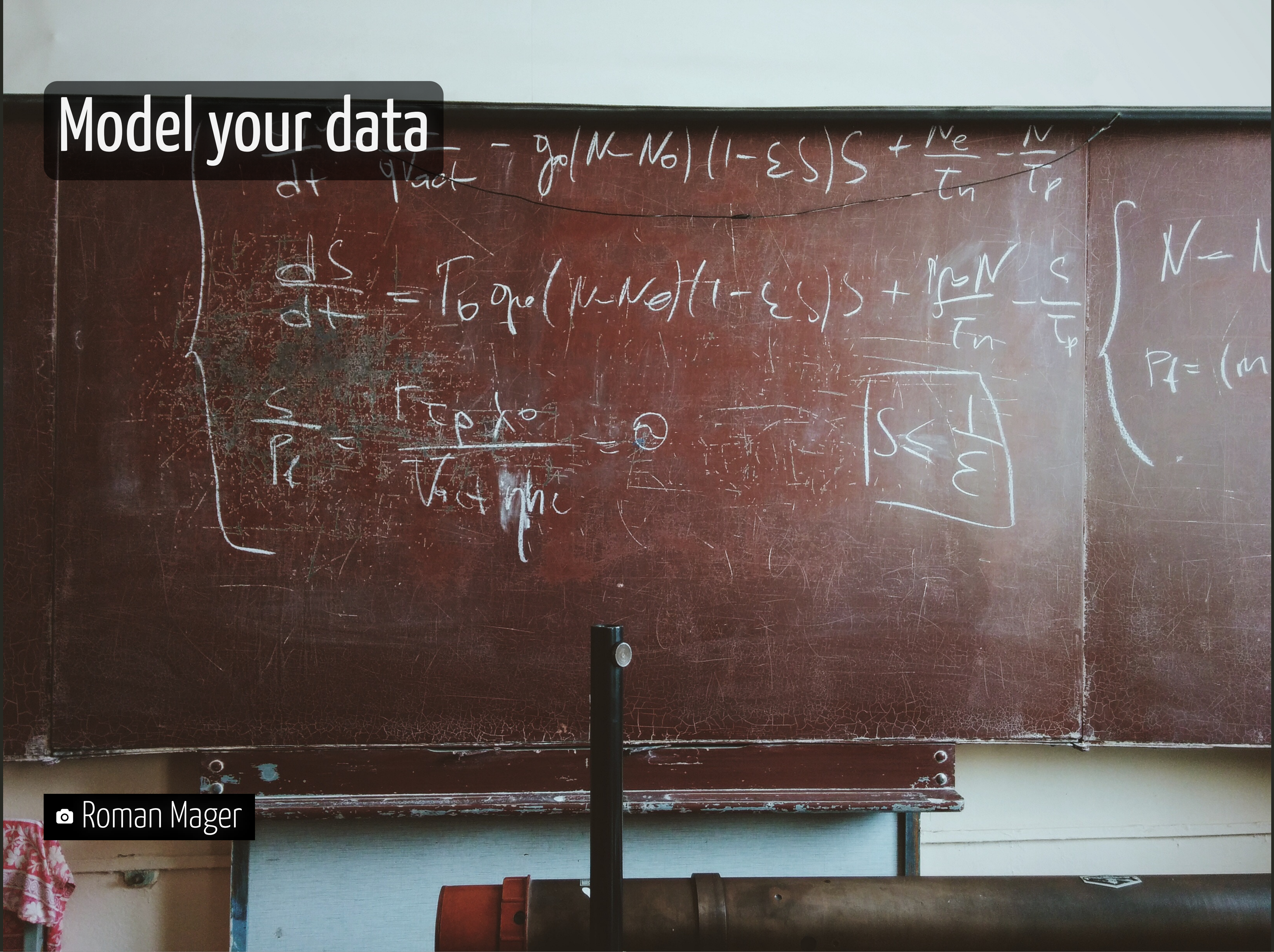



Source: doi: 10/f8gqjj

# What are we to do...?


Ross Findon

# Model your data



Roman Mager

# Model your data

Many models available for time series but palaeo data are often unhelpful

1. uneven spacing of observations in time (typically)
2. compaction, variable accumulation rates → non-constant variance

Can't use typical statistical time series models

**But we can use generalized additive models**


WIGGLY
THINGS

# GAMs

Linear trend model

$$\mathbb{E}(y_i) = \beta_0 + \beta_1 x_i$$

Generalized Additive Models

$$\mathbb{E}(y_i) = \beta_0 + f(x_i)$$

We assume trend is **smooth**

# Splines

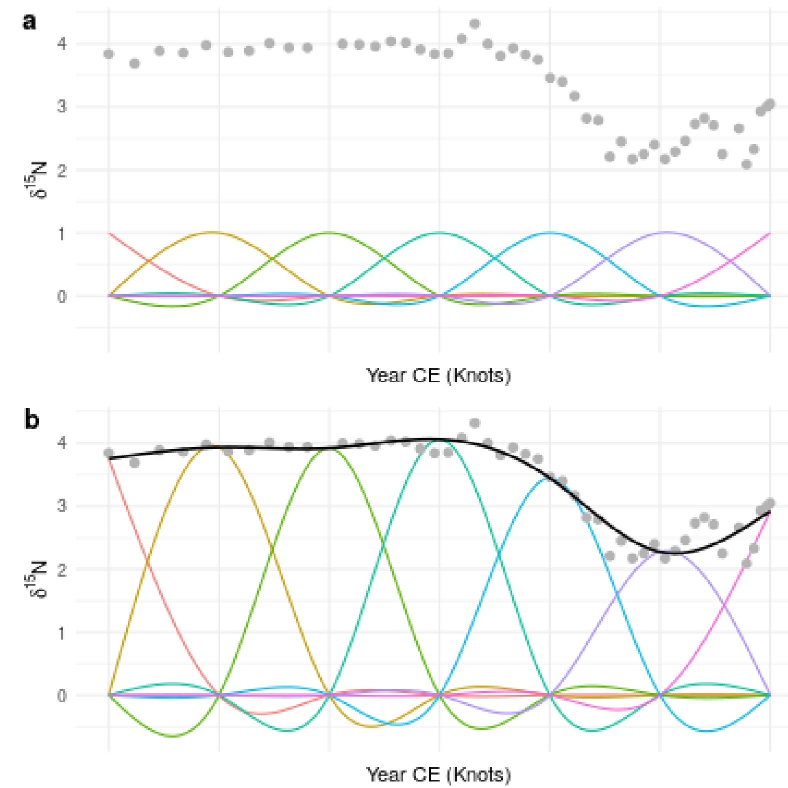
Splines are smooth functions

Made of little basis functions

Estimate $\beta_k$ for each basis function

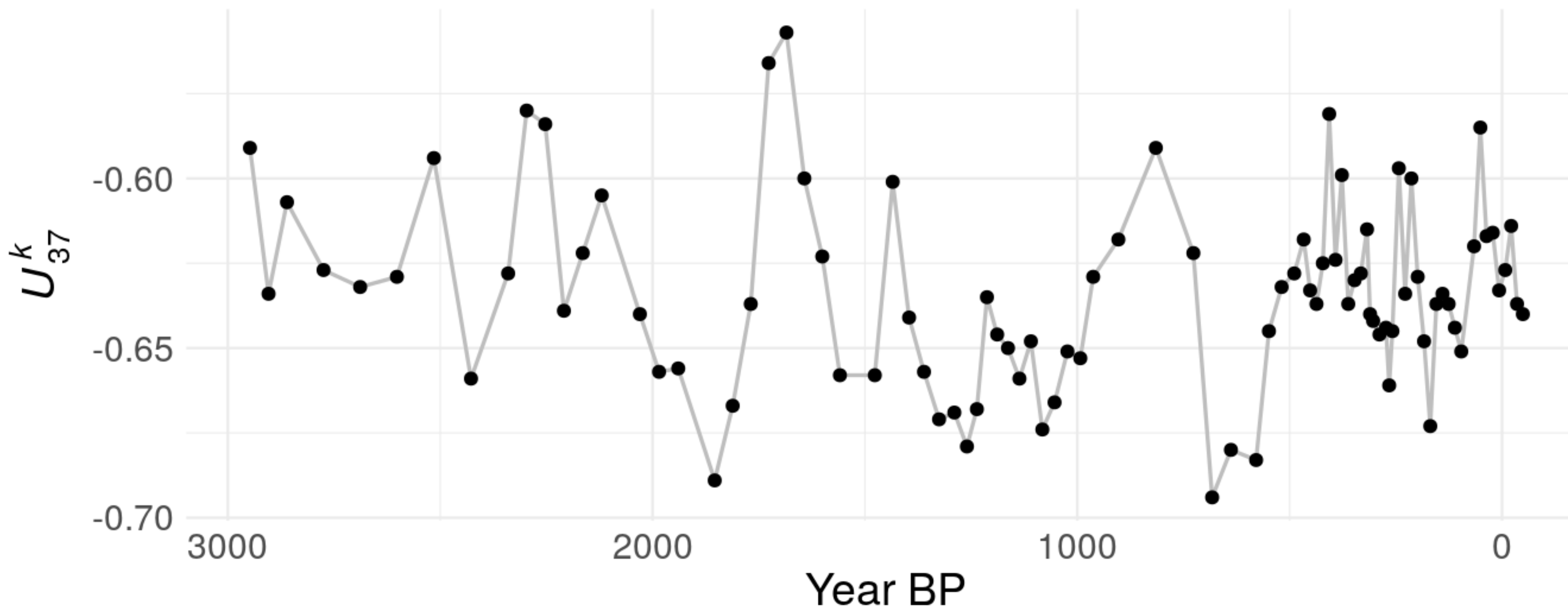
Sum the weighted basis functions at each time point

Use a penalty on **wiggliness** to avoid over-fitting



# Braya-Sø

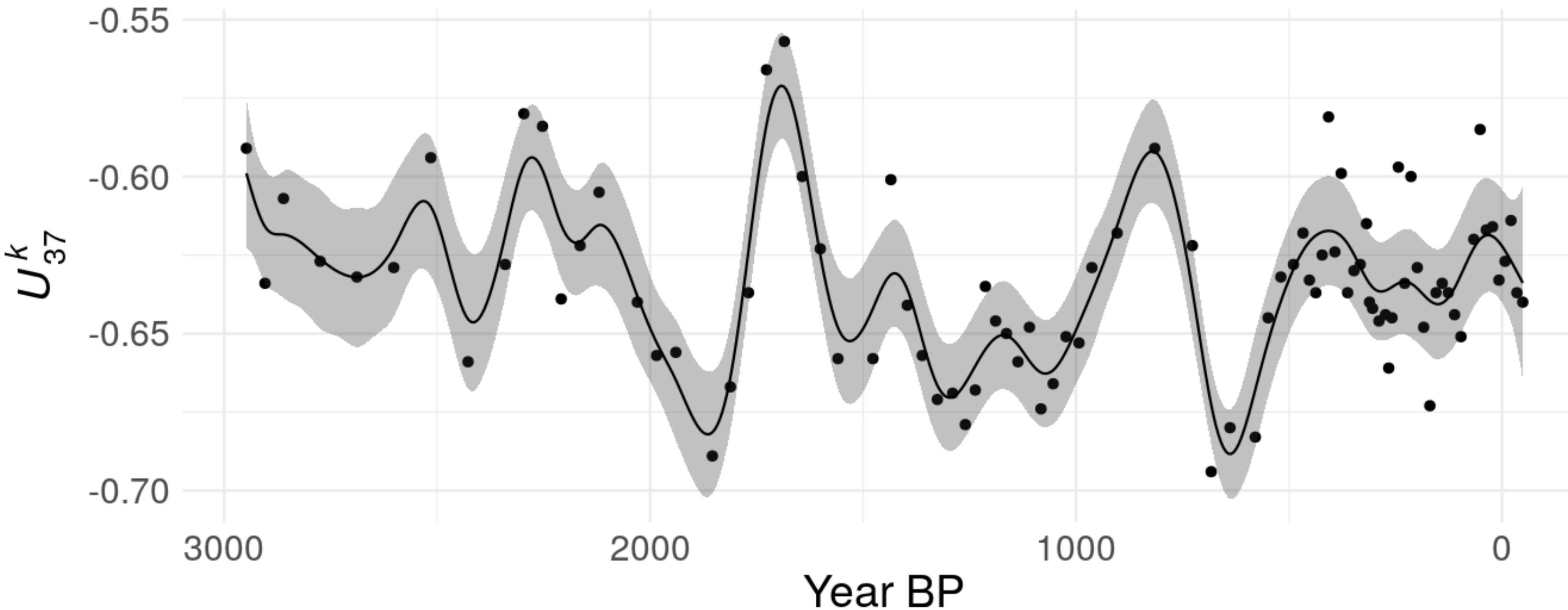
Alkenone unsaturation index — temperature interpretation




D'Andrea, *et al* (2011). PNAS *108*: 9765–9769. doi: 10/brnm2n

# Braya-Sø — estimated trend

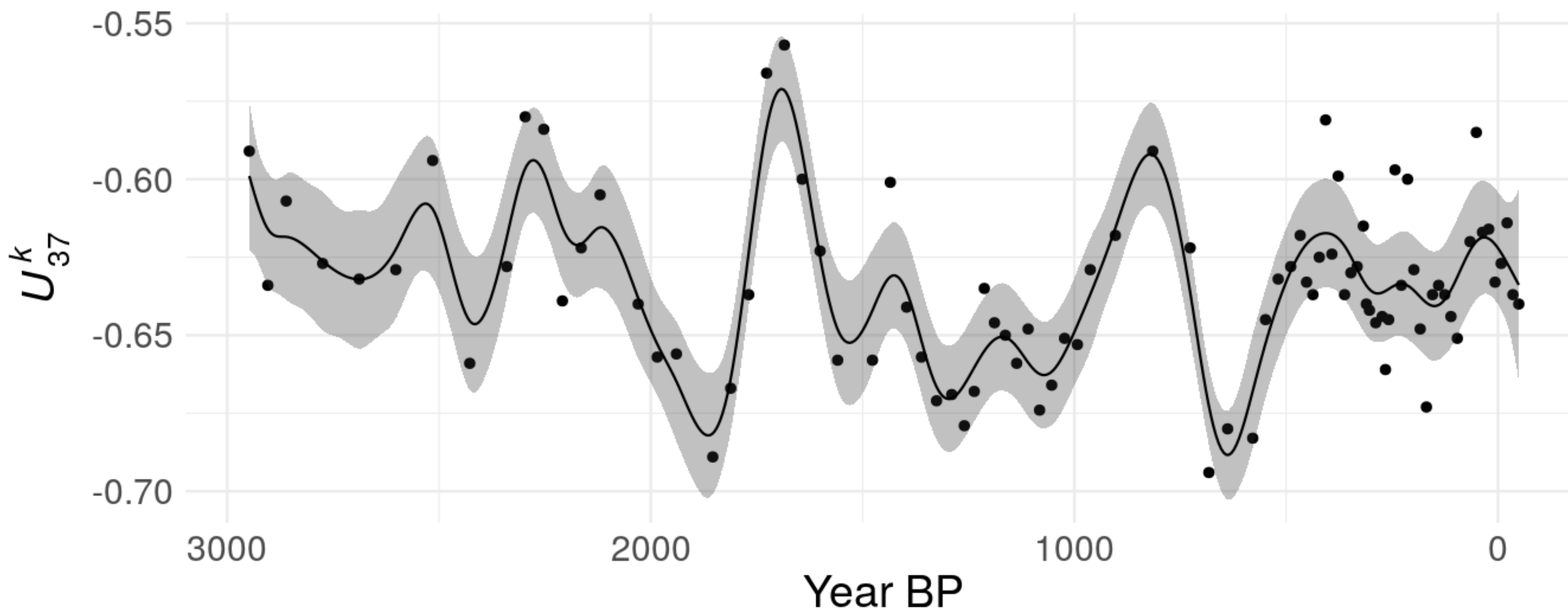
We get some wiggles — but which are real?



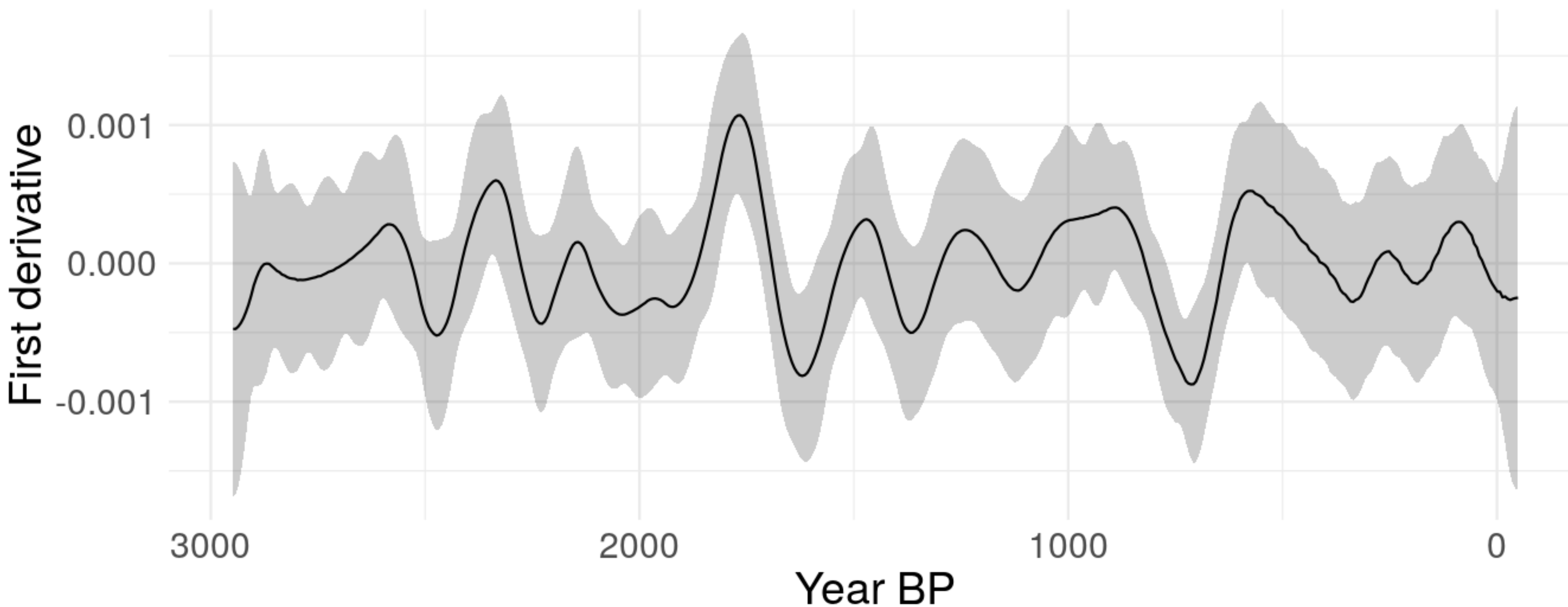
# Calculus FTW!

Compute the first derivative of the spline and it's uncertainty

- Where confidence band *includes* **0** we fail reject the $H_0$ of *no change*
- Where *doesn't* include **0** we reject $H_0$ → conclude there's change



# Braya-Sø rates of change



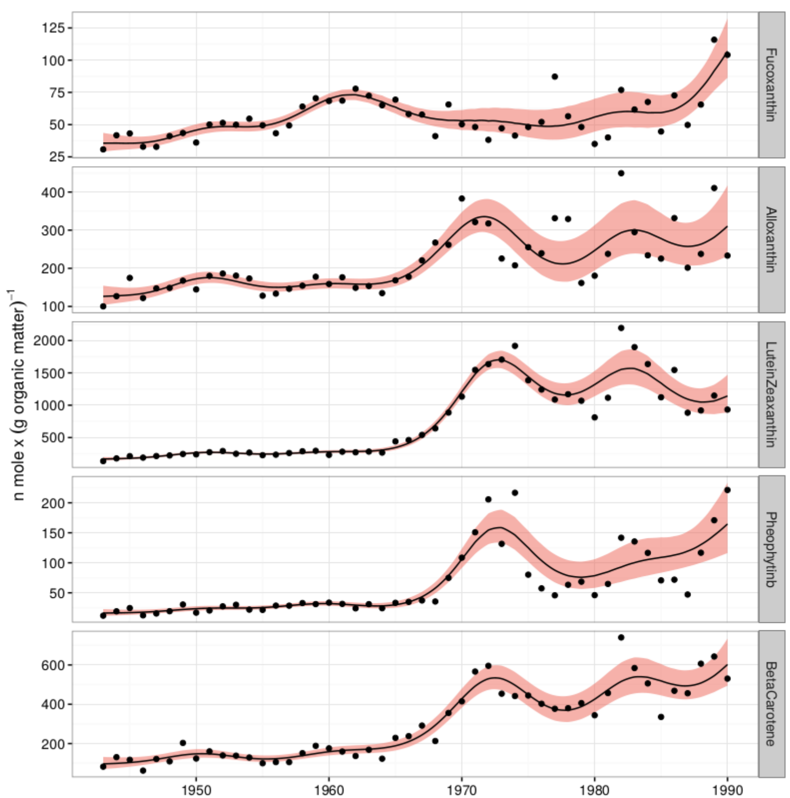
# What else...?

Once you start to think in terms of models you can address problems out of reach previously:

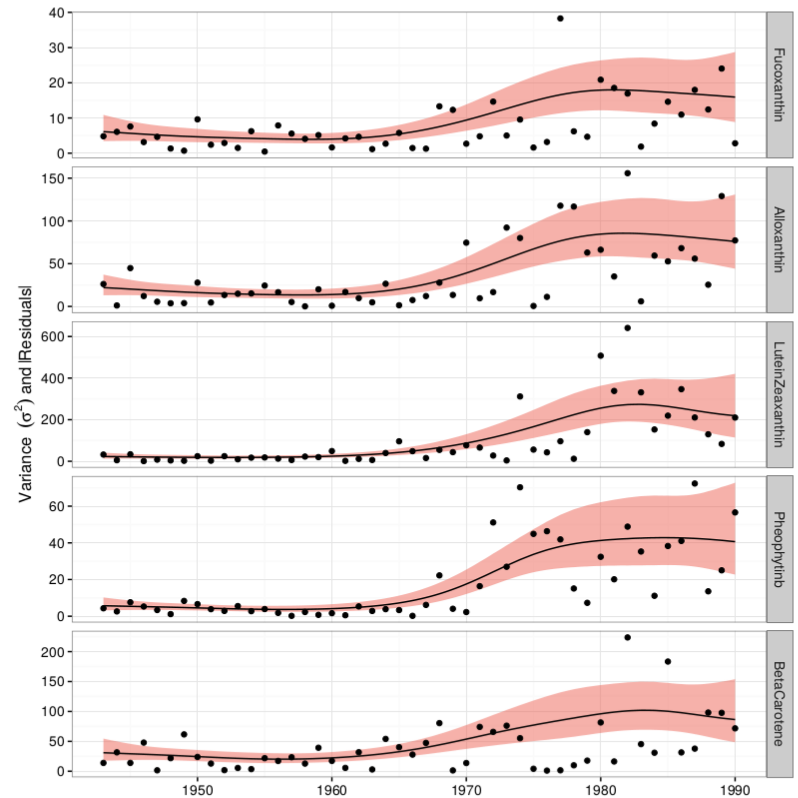
- *models for mean and variance* — resilience & tipping points (?)
- *account for age-model uncertainty*
- *compare trends across multiple sites*
- ...

# Lake 227: eutrophication → increased variance

## Trend in mean




## Trend in variance



# Conclusions

Most palaeolimnological data are time series & we're interested in estimating trends in those data

But, we rarely try to estimate those trends statistically

When we do, we often do it inappropriately

We should be modelling our data using statistical models

GAMs are a (relatively) simple model that we could use to model palaeo time series

# Want to know more...?

Preprint — doi: 10/cq5k

Blog — www.fromthebottomoftheheap.net

Slides — bit.ly/ipagams

Modelling palaeoecological time series using generalized additive models

Gavin L. Simpson
gavin.simpson@uregina.ca[1]

[1]Institute of Environmental Change and Society, University of Regina, Regina, Saskatchewan, Canada, S4S 0A2

May 11, 2018

***Keywords***— time series; generalized additive model; simultaneous interval; spline; environmental change

## Abstract

In the absence of annual laminations, time series generated from lake sediments or other similar stratigraphic sequences are irregularly spaced in time, which complicates formal analysis using classical statistical time series models. In lieu, statistical analyses of trends in palaeoenvironmental time series, if done at all, have typically used simpler linear regressions or (non-) parametric correlations with little regard for the violation of assumptions that almost surely occurs due to temporal dependencies in the data or that correlations do not provide estimates of the magnitude of change, just whether or not there is a linear or monotonic trend. Alternative approaches have used Loess-estimated trends to justify data interpretations or test hypotheses as to the causal factors without considering the inherent subjectivity of the choice of parameters used to achieve the Loess fit (e.g. span width, degree of polynomial). Generalized additive models (GAMs) are statistical models that can be used to estimate trends as smooth functions of time. Unlike Loess, GAMs use automatic smoothness selection methods to objectively determine the complexity of the fitted trend, and as formal statistical models, GAMs, allow for potentially complex, non-linear trends, a proper accounting of model uncertainty, and the identification of periods of significant temporal change. Here, I present a consistent and modern approach to the estimation of trends in palaeoenvironmental time series using GAMs, illustrating features of the methodology with two example time series of contrasting complexity; a 150-year bulk organic matter $\delta^{15}N$ time series from Small Water, UK, and a 3000-year alkenone record from Braya-Sø, Greenland. I discuss the underlying mechanics of GAMs that allow them to learn the shape of the trend from the data themselves and how simultaneous confidence intervals and the first derivatives of the trend are used to properly account for

1