Community dynamics: a palaeo perspective on regime shifts and critical transitions

Gavin L. Simpson

University of Regina · Department of Biology Seminar · September 15th 2017

Acknowledgements

Slides: **bit.ly/biolseminar**

Community response to environmental change

By Mauricio Antón [CC BY 2.5 (http://creativecommons.org/licenses/by/2.5)], via Wikimedia Commons

Community response to environmental change

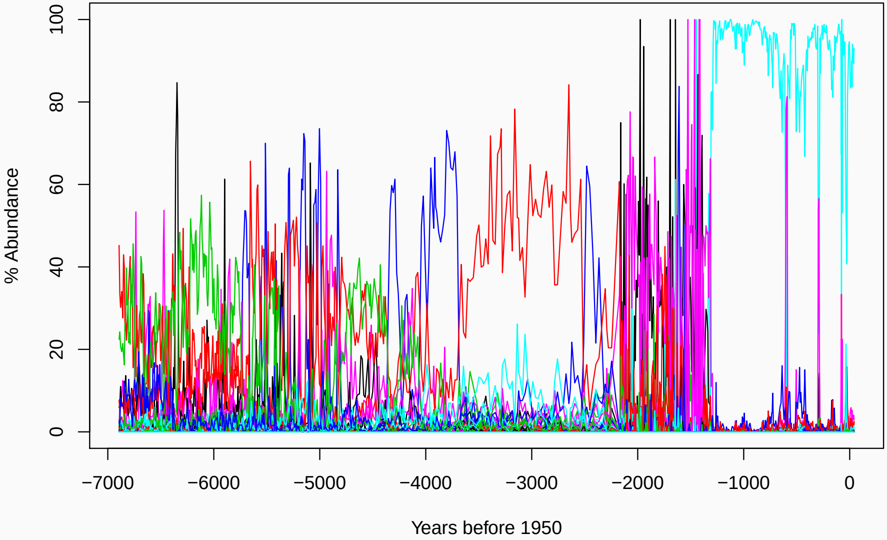
Credit: NOAA George E. Marsh Album [Public domain], via Wikimedia Commons

Community response to environmental change

Credit: Michael Knall [Public domain], via Wikimedia Commons

How have aquatic communities responded to rapid change?

Complex multivariate species data



Data provided by Jeffery Stone (Indiana State University)

Complexity

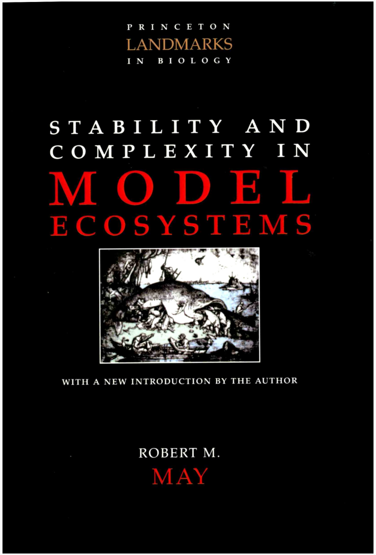
Logistic map

$$x_{n+1} = rx_n(1 - x_n)$$

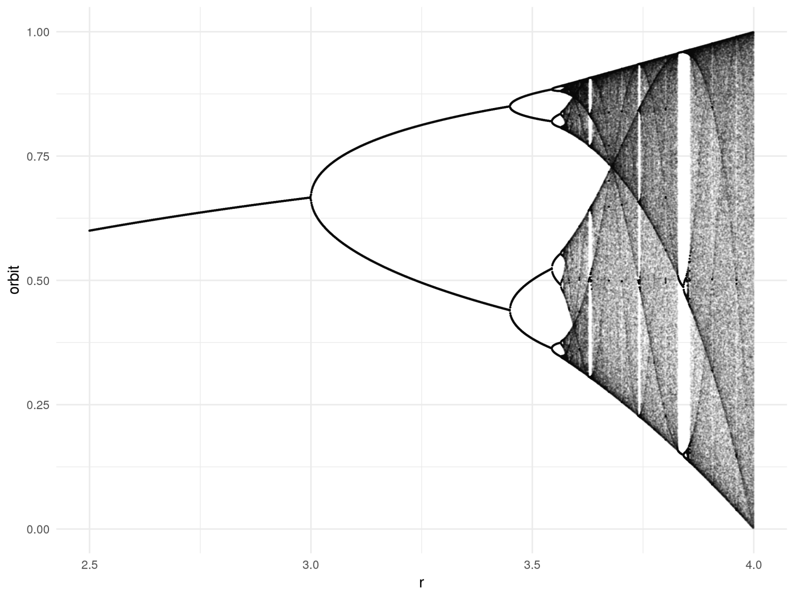
$0 < x_n < 1$ is the *fraction* of the carrying capacity of the ecosystem

r combines the birth and death rates

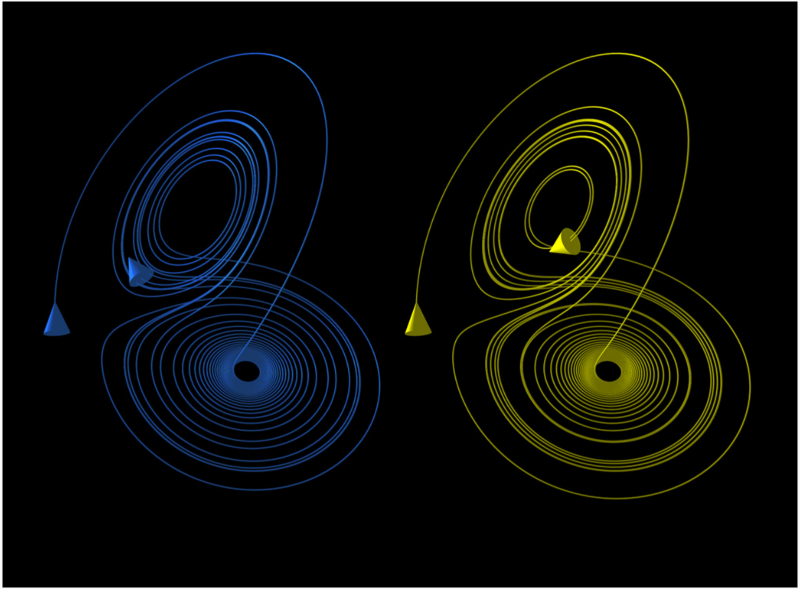
Interest is for values of r in the interval $[0, 4]$



Logistic map



Butterfly effect — Ed Lorenz

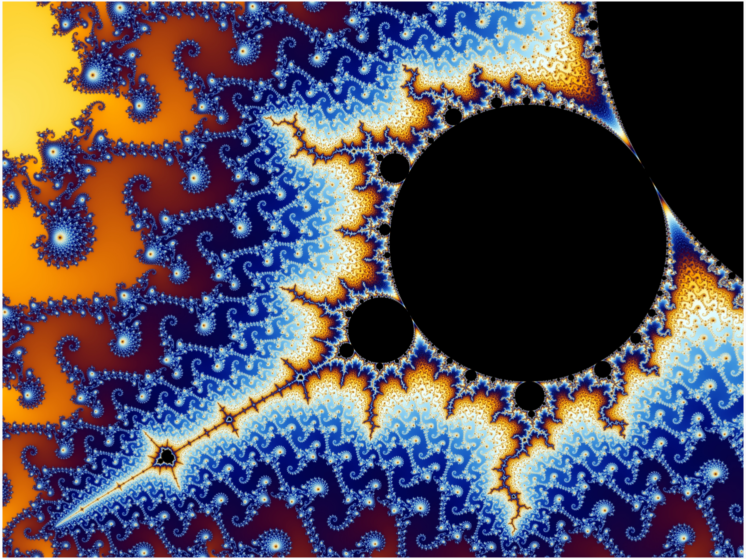


Credit: Hellisp [CC-BY], via Wikimedia Commons

Turbulent flow — Lev Landau

Credit: NASA [Public domain], via Wikimedia Commons

Fractal geometry — Benoit Mandelbrot



Credit: Wolgangbeyer [CC BY-SA], via Wikimedia Commons

Alternative stable states — Marten Scheffer

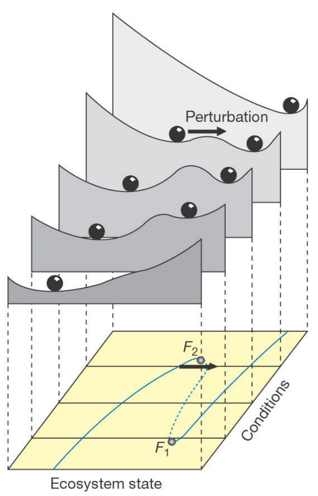
Under range of conditions multiple ecosystem states may exist

Stability landscapes depict equilibria and basins of attraction

Valleys are stable states

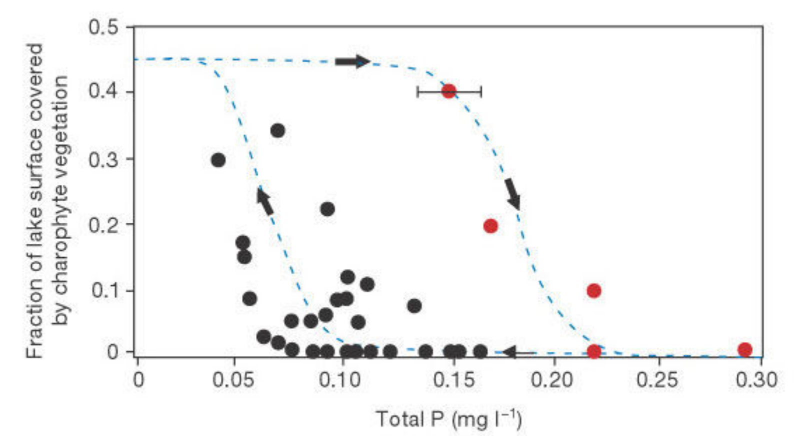
External conditions alter the stability landscape

Disturbances can kick a system between states



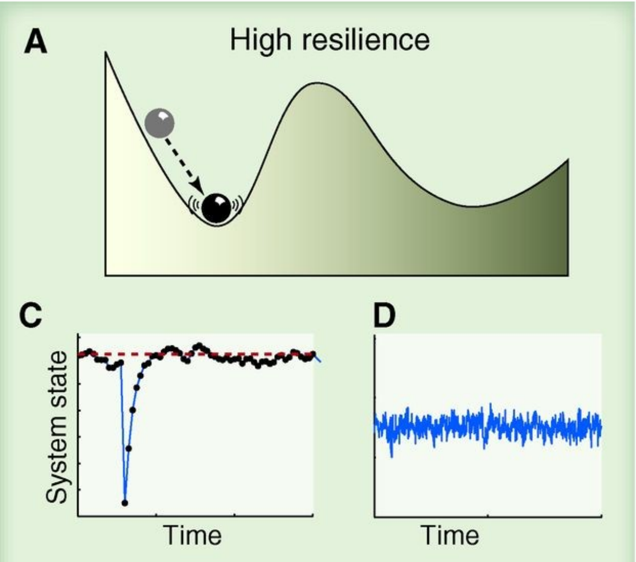
Source: Scheffer et al *Nature* 2001

Hysteresis



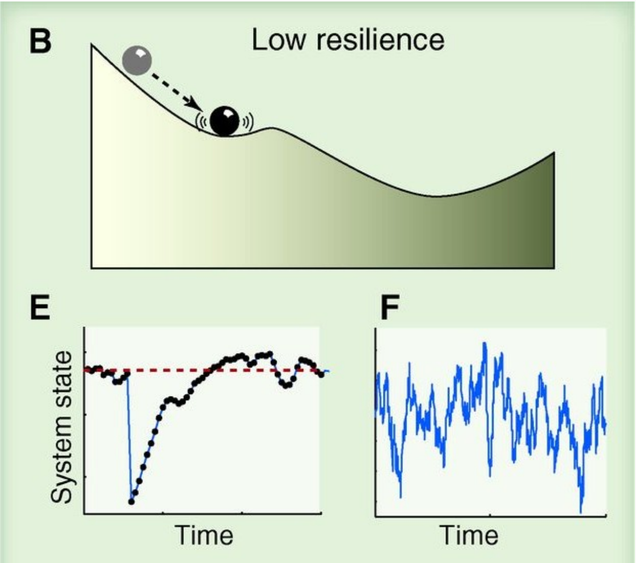
Source: Scheffer et al *Nature* 2001

Early warnings & Resilience to change



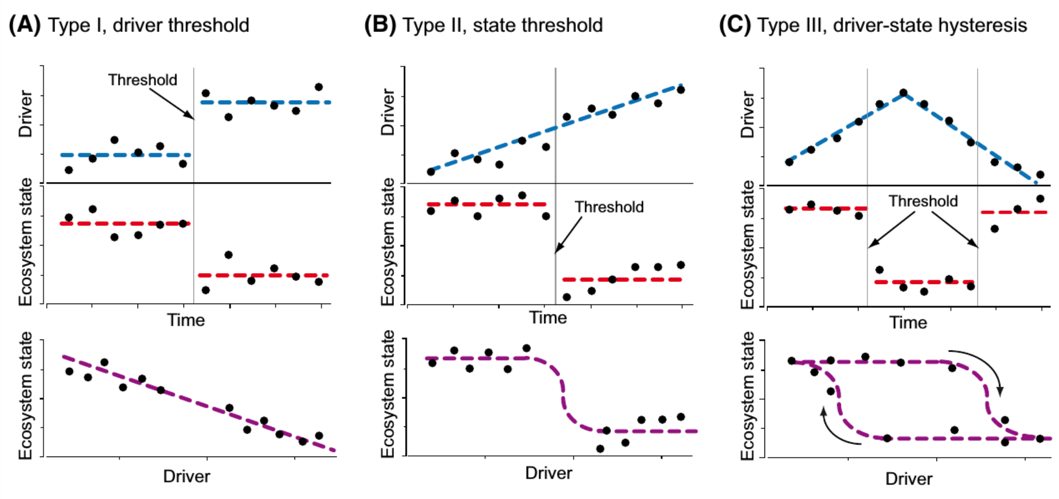
Source: Scheffer et al *Science* 2012

Early warnings & Resilience to change



Source: Scheffer et al *Science* 2012

Regime shifts



Source: Randsalu-Wendrup et al *J. Paleolimnology* 2016

Palaeoecology — Palaeolimnology

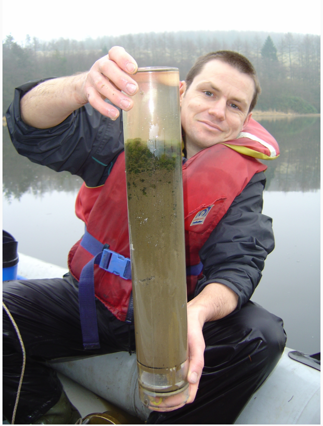
Palaeoecological data: valuable & unique source of data to study effects of environmental change

Most lakes not annually laminated

- mixing as a form of temporal averaging
- also age/dating uncertainty
- limits temporal resolution of data
- compaction of older sediments → varying time per sample

How do dynamics of species composition vary with environmental change?

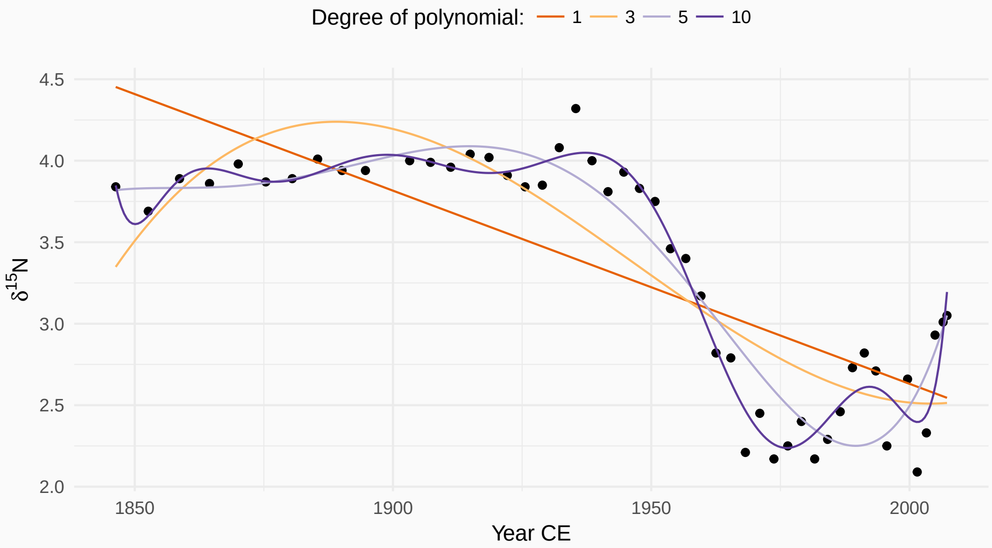
- Trend (mean), Variance, Higher moments…?



Source: © Ewan Shilland

Modelling variance

Generalized additive models



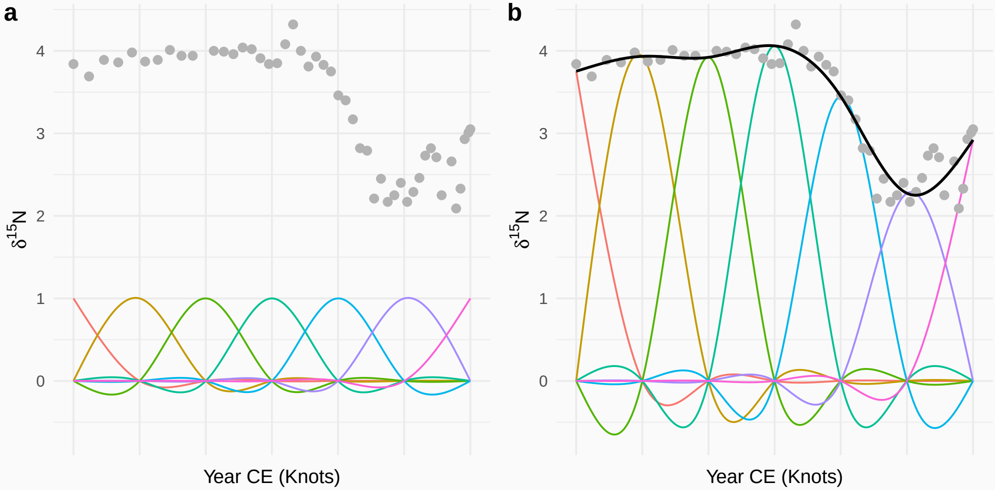
Generalized additive models (GAMs)

$$y = \beta_0 + f(\text{time}) + \varepsilon$$

f is a smooth function

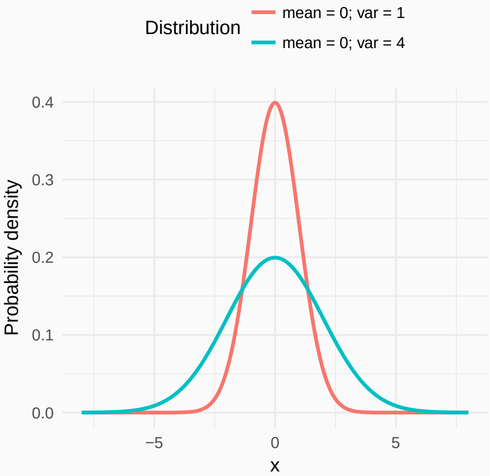
Learn about the shape of f from the data

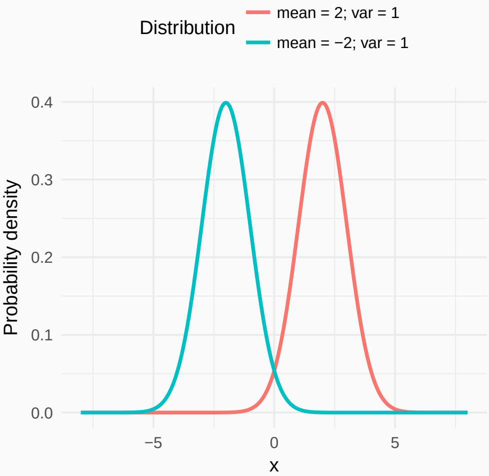
Generalized additive models



Gaussian distribution — defined by 2 parameters

- mean — μ
- variance — σ^2





Generalized additive model for location, scale, and shape

GAMLSS allows, for certain suitable distributions, the ability to model one or more of the mean, variance, skewness, and kutosis.

- GAMLSS models of Rigby and Stasinopoulos (2005)
- Vector GAM (VGAM) models of Yee (e.g. Yee and Mackenzie (2002))
- General smooth models of Wood/mgcv
- E.g.: y_i are Gaussian with mean μ_i and variance σ_i^2, each of which are modelled via a linear predictor η of smooth functions of covariates x_j and z_j

$$y_i \sim \mathcal{N}(\mu = \eta_{1,i}, \sigma^2 = \eta_{2,i})$$

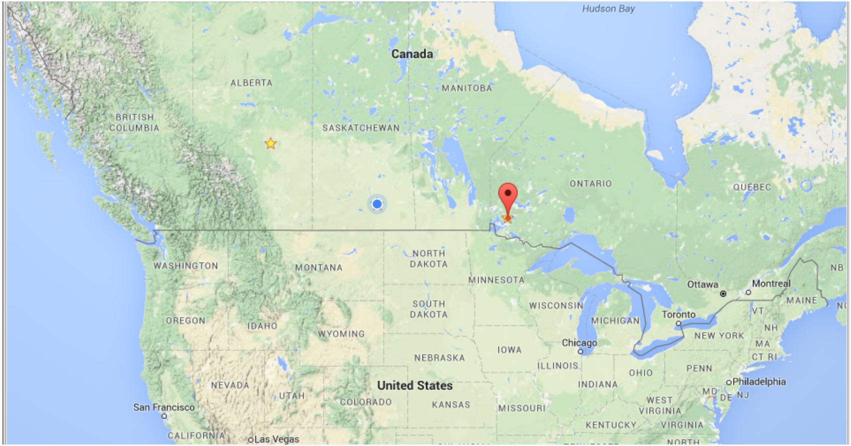
$$\eta_{1,i} = g\left(\sum_{j=1}^{k} f_j(x_{ij})\right)^{-1}$$

$$\eta_{2,i} = g\left(\sum_{j=1}^{k} f_j(z_{ij})\right)^{-1}$$

Lake 227

Lake 227

Experimental Lakes Area, NW Ontario, Canada — Dilute headwater lake (5 ha, 10 m deep)

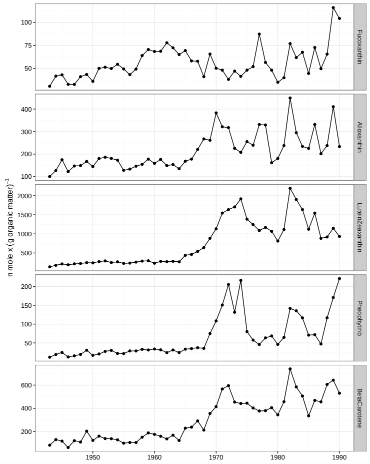


Lake 227

Experimental Lakes Area, NW Ontario, Canada — Dilute headwater lake (5 ha, 10 m deep)

Lake 227 sedimentary pigments

- Experimentally manipulated
    - 1969– :~10x increase in P load
    - 1969–74: N added in 14:1 ratio
    - 1975–82: N load reduced to 5:1
- Annual sediment samples 1943–1990
- Analysed for fossil pigments
    - *Fucoxanthin*; diatoms, chrysophytes
    - *Alloxanthin*; cryptophytes
    - *Lutein-zeaxanthin*; cyanobacteria, chlorophytes
    - *Pheophytin* b; chlorophytes
    - *β carotene*; total algae
- Cottingham, Rusak, and Leavitt (2000)



Lake 227 sedimentary pigments

Fitted Gaussian location, scale additive model using **mgcv**

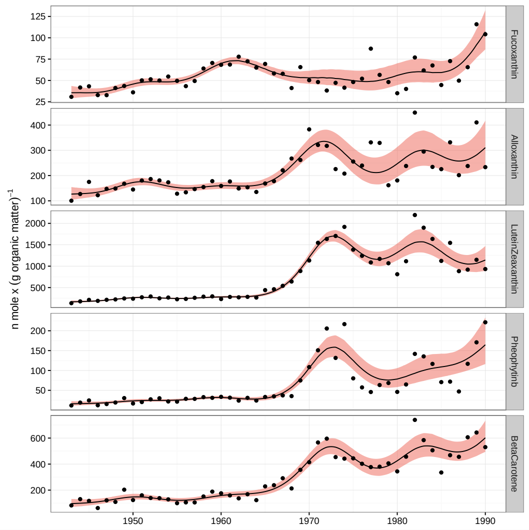
$$
\begin{aligned}
y_t &\sim \mathcal{N}(\mu = \eta_{1,t}, \sigma^2 = \eta_{2,t}) \\
\eta_{1,t} &= \exp(f_1(\text{Year}_t) + f_{1,\text{pigment}}(\text{Year}_t)) \\
\eta_{2,t} &= g(f_2(\text{Year}_t) + f_{2,\text{pigment}}(\text{Year}_t))^{-1}
\end{aligned}
$$

Mean & variance modelled with:

1. Global function of Year, plus
2. "Random effect" spline (spline-factor interaction `bs = "fs"`) with penalty on first derivative — penalizes departures from global function
3. $g()$ is a log link function modified to avoid $\hat{\sigma}^2 = 0$

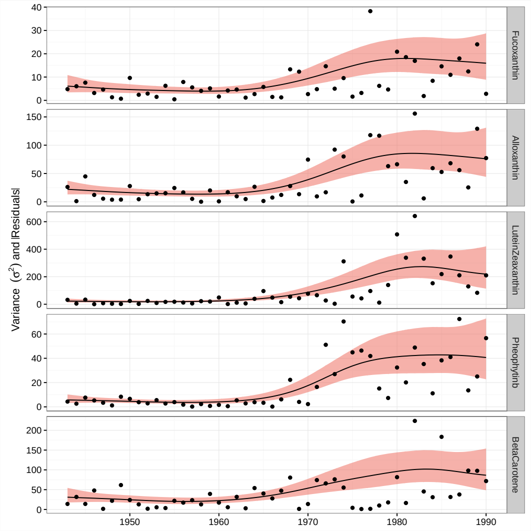
Lake 227 sedimentary pigments — fitted mean functions

- Random effect spline provides better fit of trend (mean)
- Undersmoothing of several pigments if separate models fit
- Increase in most algal groups following fertilization



Lake 227 sedimentary pigments — fitted variance functions

- Increasing variance in pigment concentration following manipulation
- Stabilisation post manipulation
- Consistent with median-log Levene's test (Cottingham, Rusak, and Leavitt 2000)



Baldeggersee

Baldeggersee

- Nutrient rich, hardwater lake, located on the central Swiss Plateau
- Developed anoxic bottom waters in 1885
- Intensive farming & agriculture in catchment
- Artificial oxygenation & circulation as remediation from 1982
- Cored in 1993 · 91 samples · 75 diatom taxa
- Source: Lotter (1998) *The Holocene* 8(4), 395–405

Source: © Andy Lotter

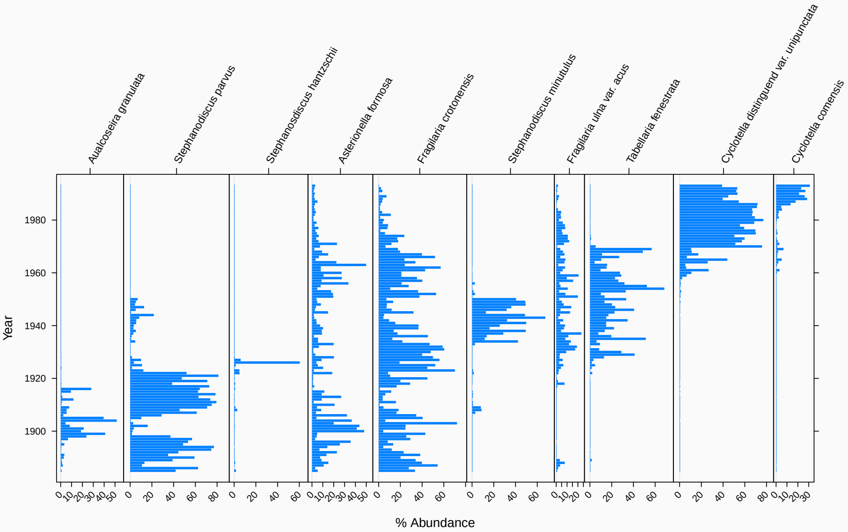
Baldeggersee

- Iconic example of a critical transition
- Deep lakes can exhibit bistable conditions
    - Nutrients increase phytoplankton
    - Decomposition can increase anoxia
    - Stimulates P release from sediments
- Marl lake
- Flickering (?)

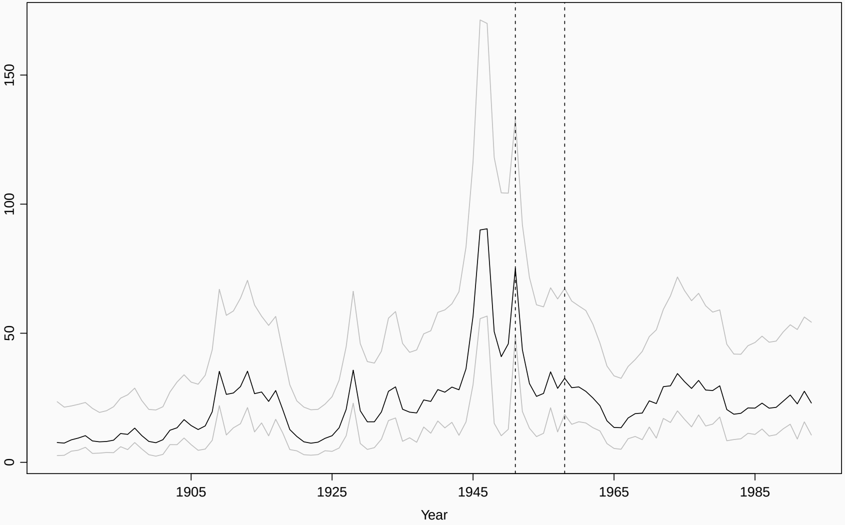
Are early warning signals (increased variance) present?

Source: © Andy Lotter

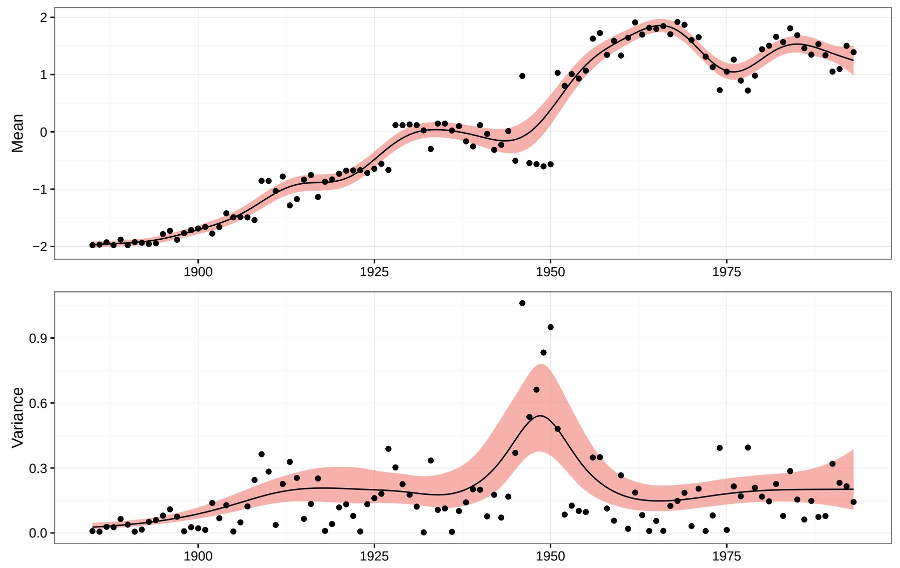
Baldeggersee — Stratigraphy



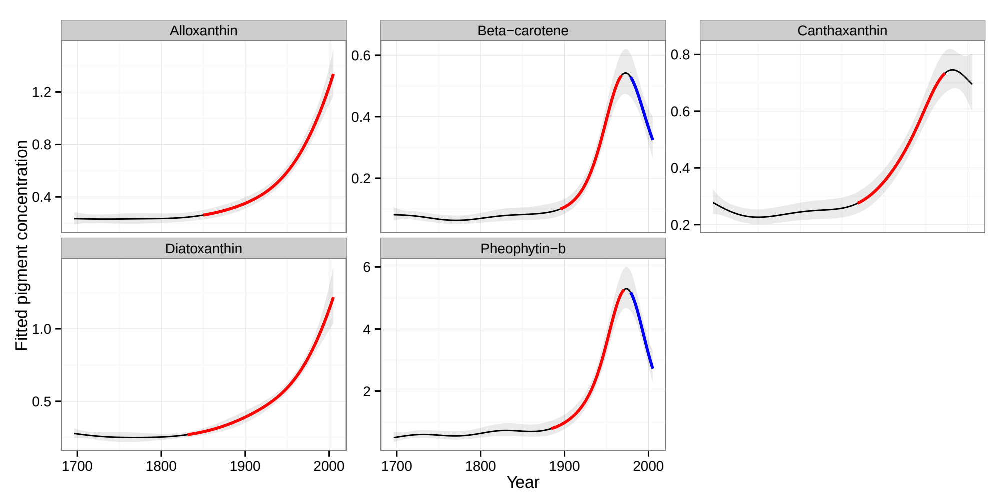
Baldeggersee — Stochastic Volatility Model



Baldeggersee — GAM Location and Scale fits

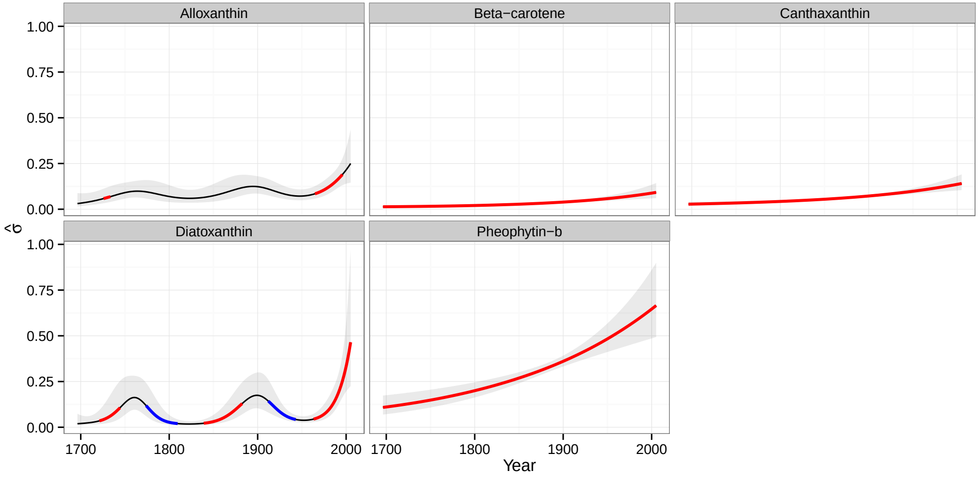


Has Lake Winnipeg suffered a catastrophic regime shift?



Source: Bunting et al *Limnology & Oceanography* 2016

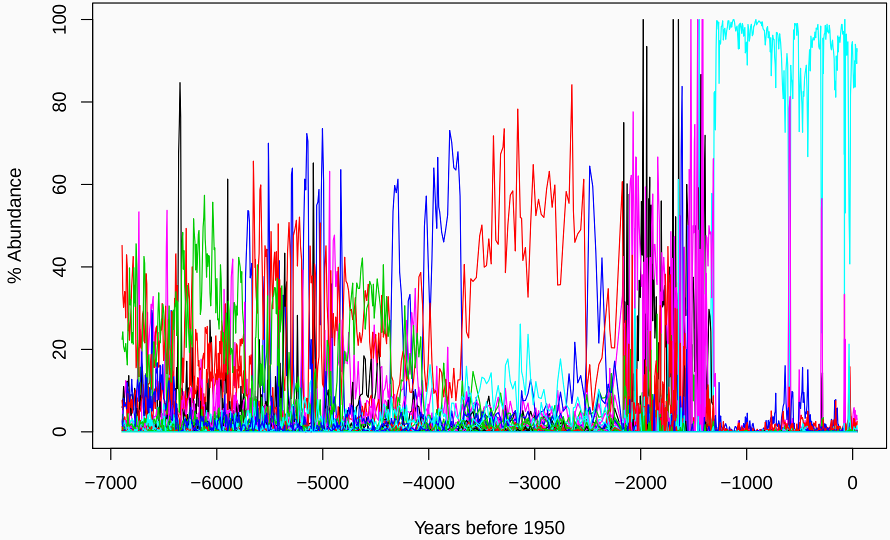
Has Lake Winnipeg suffered a catastrophic regime shift?



Source: Bunting et al *Limnology & Oceanography* 2016

Topic Models

Complex multivariate species data



Data provided by Jeffery Stone (Indiana State University)

Dimension reduction

Typically we can't model all 100+ taxa in data sets like this

- (M)ARSS-like models don't like the large n
- Most methods assume regular spacing in time

Seek a reduced dimensionality of the data that preserves the signal

Existing dimension reduction methods aren't appropriate for questions we want to ask

- Interpretation of latent factors is complex (PCA, CA, Principal Curves)
- Impose hard thresholding on the data (hierarchical clustering)
- No grouping of species or samples, or group only one

Can we group species into J associations and soft cluster sample as compositions of these associations?

Foy Lake — Montana

Foy Lake

- Deep, freshwater lake
- Drought-sensitive Flathead River Basin
- Diatom assemblages sensitive to lake depth variation
- Related to variability in effective moisture

Regime shift ~1.3ka BP

Spanbauer *et al* PLOS ONE **9**(10) e108936

OPEN ACCESS Freely available online

PLOS | ONE

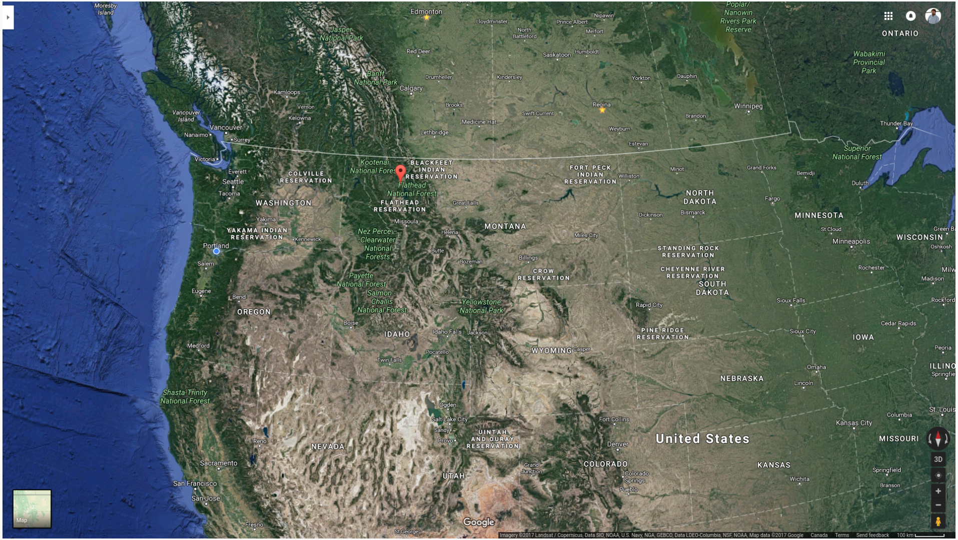
Prolonged Instability Prior to a Regime Shift

Trisha L. Spanbauer[1*], Craig R. Allen[2], David G. Angeler[3], Tarsha Eason[4], Sherilyn C. Fritz[1], Ahjond S. Garmestani[4], Kirsty L. Nash[5], Jeffery R. Stone[6]

Abstract

Regime shifts are generally defined as the point of 'abrupt' change in the state of a system. However, a seemingly abrupt transition can be the product of a system reorganization that has been ongoing much longer than is evident in statistical analysis of a single component of the system. Using both univariate and multivariate statistical methods, we tested a long-term high-resolution paleoecological dataset with a known change in species assemblage for a regime shift. Analysis of this dataset with Fisher Information and multivariate time series modeling showed that there was a ~2000 year period of instability prior to the regime shift. This period of instability and the subsequent regime shift coincide with regional climate change, indicating that the system is undergoing extrinsic forcing. Paleoecological records offer a unique opportunity to test tools for the detection of thresholds and stable-states, and thus to examine the long-term stability of ecosystems over periods of multiple millennia.

Foy Lake — Montana



Foy Lake — Montana

Topic models

Machine learning approach for organizing text documents

Latent Dirichlet Allocation (LDA) — (Blei, Ng, & Jordan, *J. Mach. Learn. Res.* 2003)

Generative model

Valle, D., Baiser, B., Woodall, C. W. & Chazdon, R. (2014) *Ecology Letters* **17**

Community of Skittles

Individual skittles from one of four flavoured packs

What are the proportion of flavours in each pack?

How many of each pack comprise the skittle community?

Latent Dirichlet Allocation

Aim is to infer the

- distribution of skittle **flavours** within each pack β_j, and
- distribution of skittle **packs** within each community (*sample*)

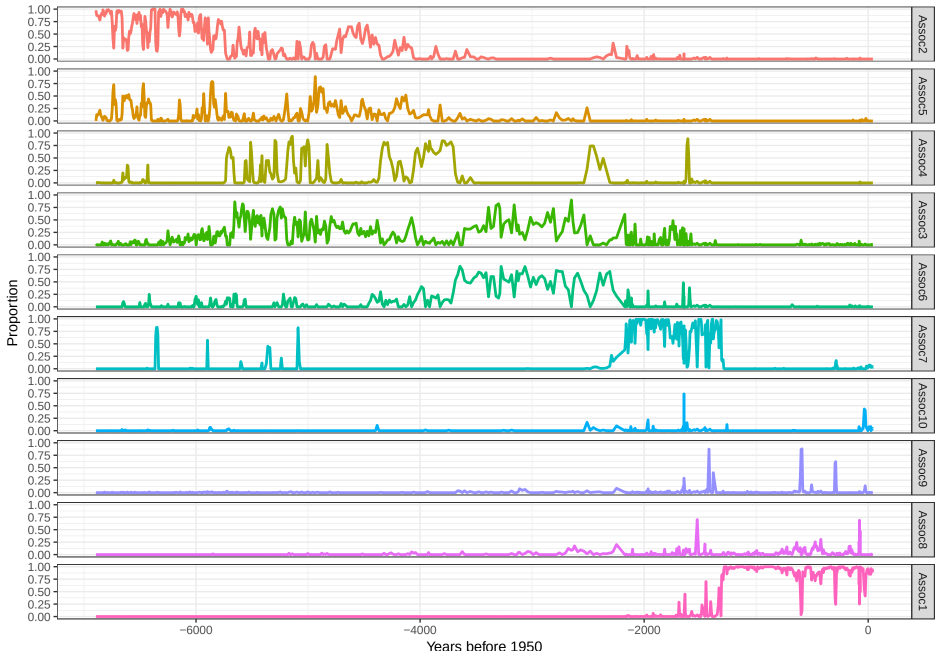
Achieve a soft clustering of samples — mixed membership model

Achieve a soft clustering of **species** (flavours) into **associations** of taxa (packs)

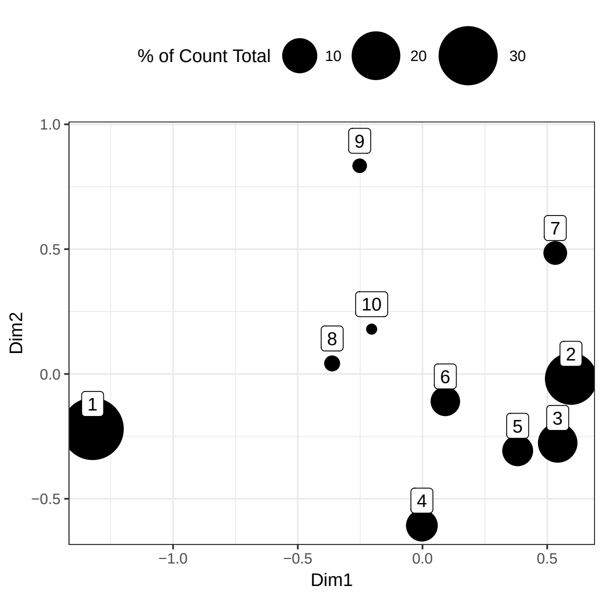
User supplies J — the number of associations *a priori* – $j = \{1, 2, \ldots, J\}$

J chosen using AIC, *perplexity*, CV, ...

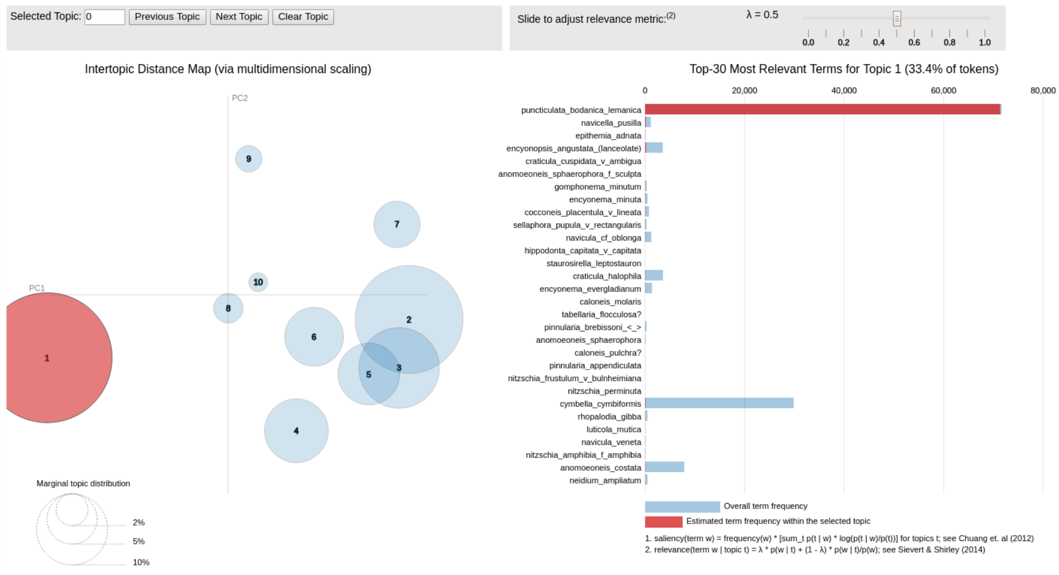
Latent Dirichlet Allocation



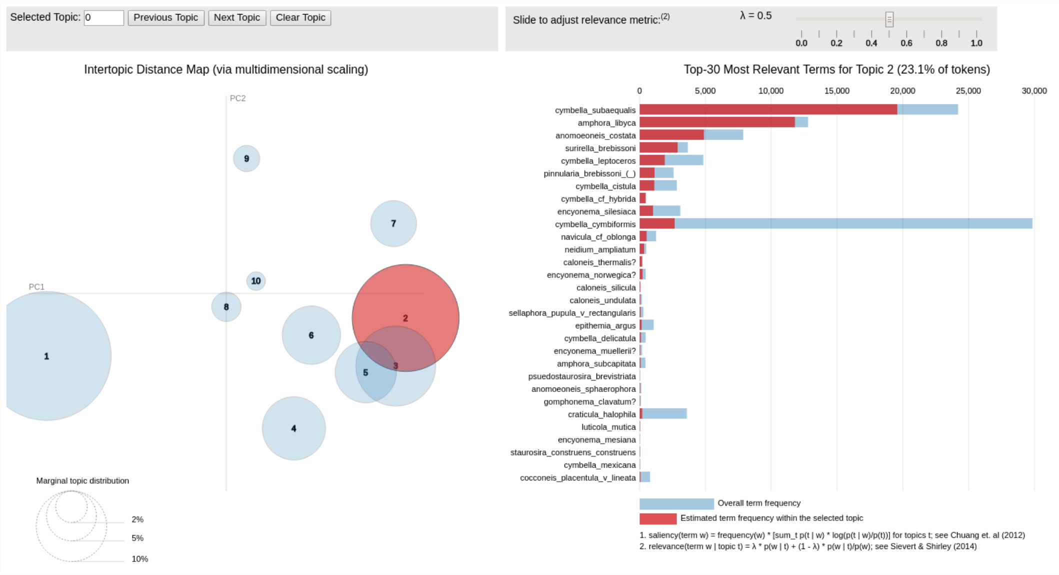
Latent Dirichlet Allocation — nMDS



Latent Dirichlet Allocation — Association 1



Latent Dirichlet Allocation — Association 2



Latent Dirichlet Allocation — Trend estimation

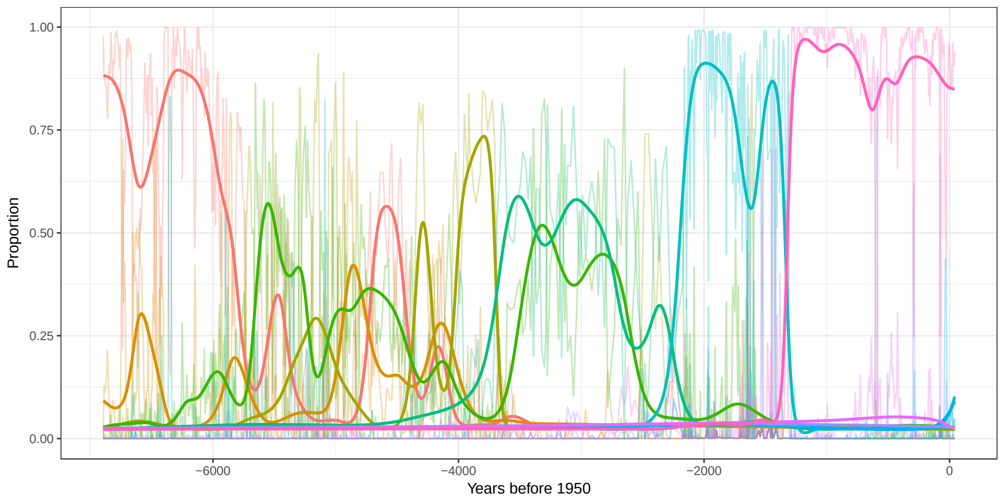
LDA knows nothing about the temporal ordering of the samples

Estimate trends in proportions of species associations using a GAM

- assumes smoothly varying trends
- use adaptive spline to allow for rapid adaptation to changing data
- model each association as $\sim$ Beta(μ, ϕ)

Other methods would also be appropriate: eg. Bayesian Change point model

Latent Dirichlet Allocation — Trend estimation (Adaptive GAM)



Summary

Ecosystems are under threat from a range of natural and anthropogenic factors

Research over the past 40–50 years suggests that ecosystems respond in a myriad complex ways

GAMs have proven useful for testing some of these ideas with long time series & palaeo data

GAMs can estimate trends in variance in difficult data without very complex models or *ad hoc* solutions

Summary

Topic models proved well-capable of summarizing the complex community dynamics of Foy Lake

- Reduced 113 taxa to 10 associations of species
- Species associations match closely the expert interpretation of the record
    - make autecological sense also
- The CTM was more parsimonious — removed one rare association
- Estimated trends in proportions of species associations capture
    - smooth, slowly varying trends, and
    - rapid (regime shift?) state change ~ 1.3 ky BP

Sampling process of topic models similar in spirit to those recently invoked for community assembly

Future directions

GAM-based models for other EWS?

- Autocorrelation
- Skewness & kurtosis

Distributional models & recently-developed quantile GAM approacheds might work

Future directions

Choosing J is inconvenient

Address this via **Hierarchical Dirichlet Processes** and Bayesian Nonparametrics

- assume J in infinite & put a prior distribution over J

Associations in LDA & CTM are static — distributions are fixed for all samples

- dynamic topic models allow distributions to vary smoothly with time

Many developments in this field:

- *Chinese Restaurant Process*,
- *Indian Buffet Process*, &
- ...

1. Flavour distribution for jth type of Skittle

$$\beta_j \sim \text{Dirichlet}(\delta)$$

2. Proportions of each type in the Skittle community

$$\theta \sim \text{Dirichlet}(\alpha)$$

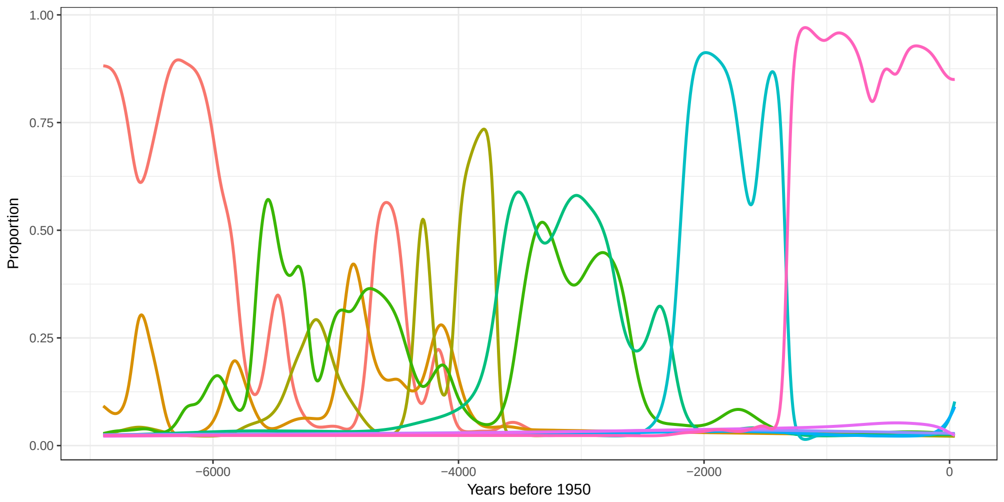
3. For each skittle s_i
    - Choose a pack in proportion

$$z_i \sim \text{Multinomial}(\theta)$$

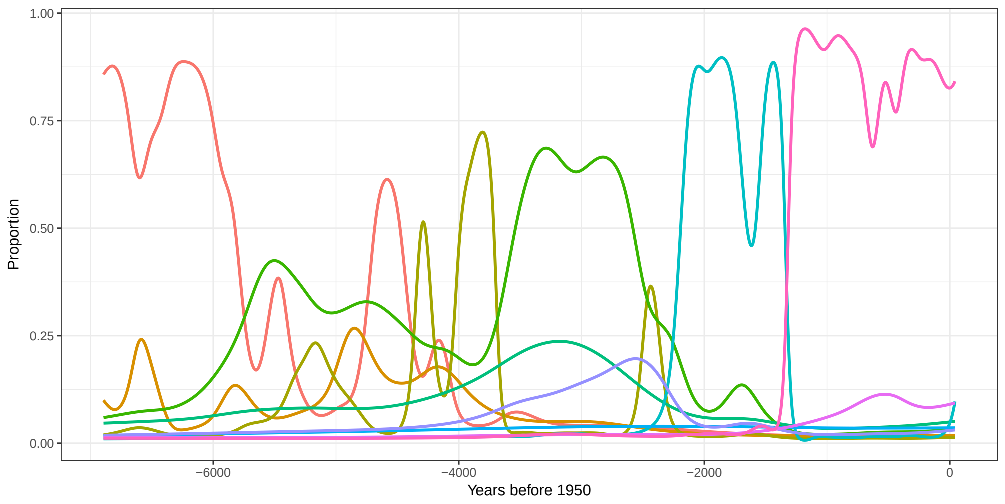
    - Choose a flavour from chosen pack with probability

$$p(s_i|z_i, \beta_j) \sim \text{Multinomial}(\delta)$$

Latent Dirichlet Allocation — Trend estimation (AdaptiveGAM)



Correlated Topic Model — Trend estimation (Adaptive GAM)



Intuition behind LDA

Latent Dirichlet allocation represents a trade-off between two goals

1. for each sample, allocate its individuals to a **few associations of species**
2. in each association, assign high probability to a **few species**

These are in opposition

- assigning a sample to a single association makes *2* **hard** — all its species must have high probability under that one topic
- putting very few species in each association makes *1* **hard** — to cover all individuals in a sample must assign sample to many associations

Trading off these two goals therefore results in LDA finding tightly co-occurring species

Cottingham, K L, J A Rusak, and P R Leavitt. 2000. "Increased Ecosystem Variability and Reduced Predictability Following Fertilisation: Evidence from Palaeolimnology." *Ecology Letters* 3 (4). Blackwell Science Ltd: 340–48. doi:10.1046/j.1461-0248.2000.00158.x.

Rigby, R A, and D M Stasinopoulos. 2005. "Generalized Additive Models for Location, Scale and Shape." *J. R. Stat. Soc. Ser. C Appl. Stat.* 54 (3). Blackwell Publishing Ltd: 507–54.

Yee, Thomas W, and Monique Mackenzie. 2002. "Vector Generalized Additive Models in Plant Ecology." *Ecol. Modell.* 157 (2-3): 141–56.