Abrupt changes in climate and ecosystems: automatic model selection

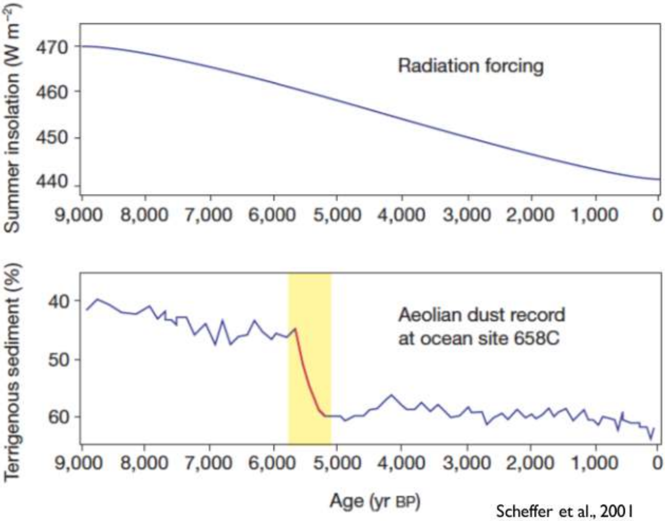
Rebecca Killick
Joint work with Claudie Beaulieu
(Southampton Oceanographic Centre)
20 Sept 2016

Summary of Talk

- Motivation
- Intro to changepoint detection
- Introduce the PELT (Pruned Exact Linear Time) method
- Automatic model selection
- Simulation Study
- North Pacific Example

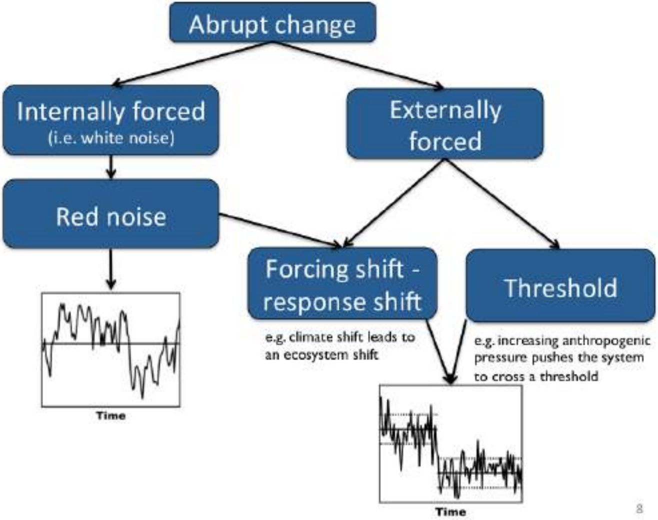
Motivation

Sahara Desert

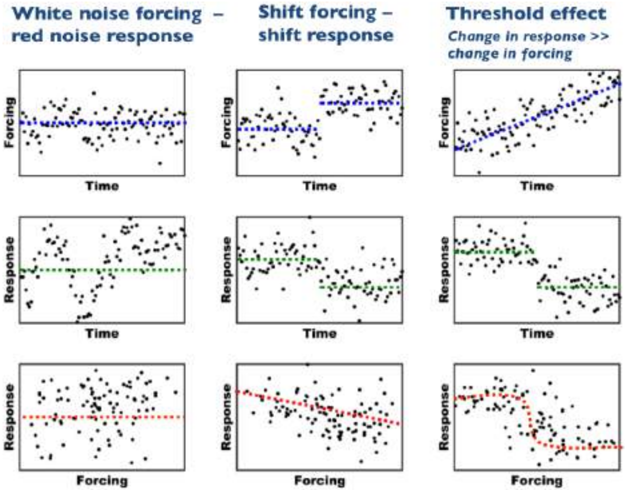


Scheffer et al., 2001

External forcing or random reorganization?



Forcing-Response relationship

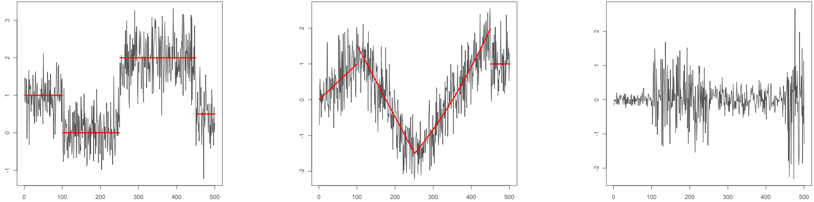


Adapted from Andersen et al., 2008; Bestelmeyer et al., 2011

Intro to changepoint detection

What are changepoints?

For data $y_1, \ldots, y_n$, a changepoint is a location τ where the statistical properties of $y_1, \ldots, y_\tau$ are different from $y_{\tau+1}, \ldots, y_n$.

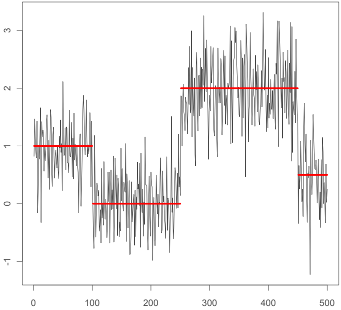


Example: Change in mean

Assume we have time-series data where

$$Y_t | \theta_t \sim \mathrm{N}(\theta_t, 1),$$

but where the means, θ_t, are piecewise constant through time.



Example: Inferring Changepoints

We want to infer the number and position of the points at which the mean changes. One approach:

Likelihood Ratio Test
To detect a single changepoint we can use the likelihood ratio test statistic:

$$LR = \max_{\tau}\{\ell(y_{1:\tau}) + \ell(y_{\tau+1:n}) - \ell(y_{1:n})\}.$$

We infer a changepoint if $LR > \beta$ for some (suitably chosen) β. If we infer a changepoint its position is estimated as

$$\tau = \arg\max\{\ell(y_{1:\tau}) + \ell(y_{\tau+1:n}) - \ell(y_{1:n})\}.$$

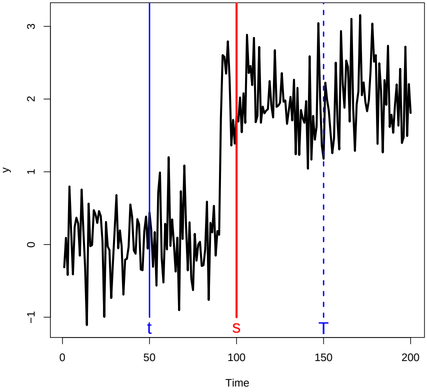
This can test can be repeatedly applied to new segments to find multiple changepoints.

The PELT Method to identify multiple changes

(Pruned Exact Linear Time)

PELT in a nutshell

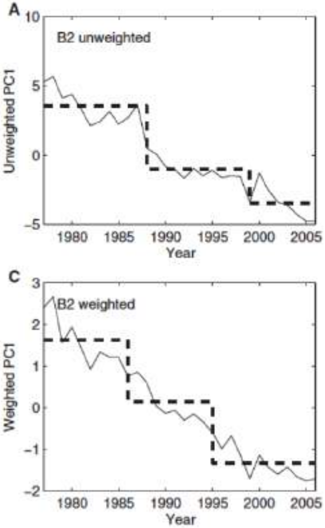
- Dynamic programming allows us to only worry about the location of the *last* change.
- Pruning means that as we go through the data we are smart about which locations are potential last change locations.



Model Selection

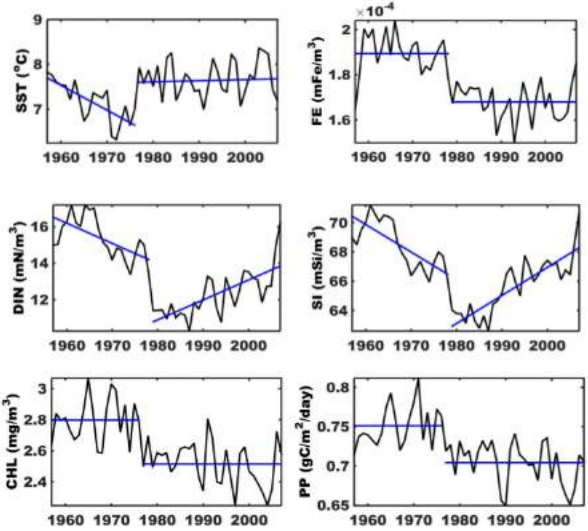
Motivation - from bad practice

- From a publication in Marine Ecology (not the only one)
- Used the Rodionov (2004) method very popular.
- Cannot deal with trend or autocorrelation.



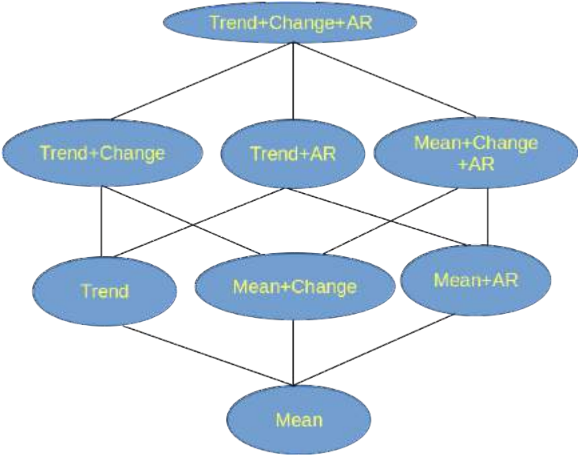
Motivation - from Beaulieu et al. 2015

- potentially hundreds or thousands of series
- no time to consider the format of change for each
- need to include both the potential for trends and also red noise (auto-correlation).



Model Selection

AIM: select the most parsimonious but accurate model for the data.



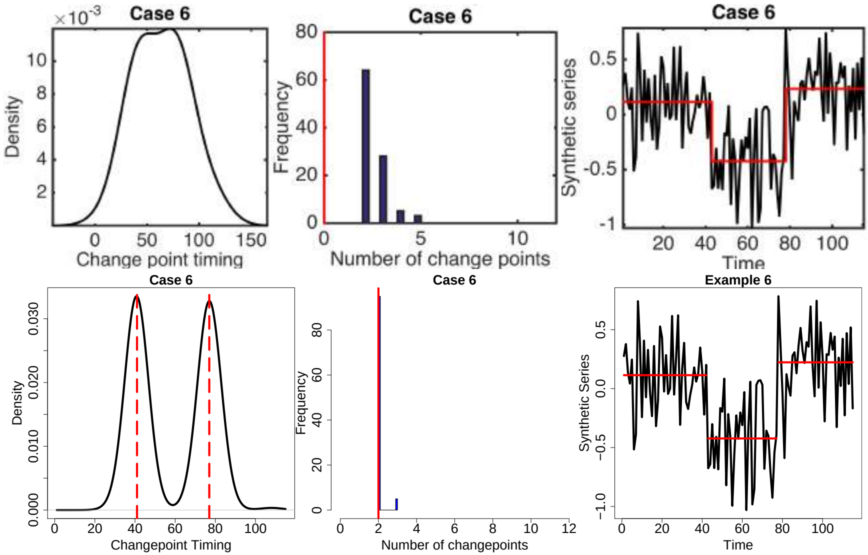
Simple to extend with other types of models.

Model Selection

- Fast changepoint detection techniques gives us the ability to fit all models
- Choose the best model according to your favourite criterion (we use AIC here).
- If you are worried about computation time, you can fit stepwise.
- All routines are available in R and Matlab packages - one function does it all.

Simulation Study

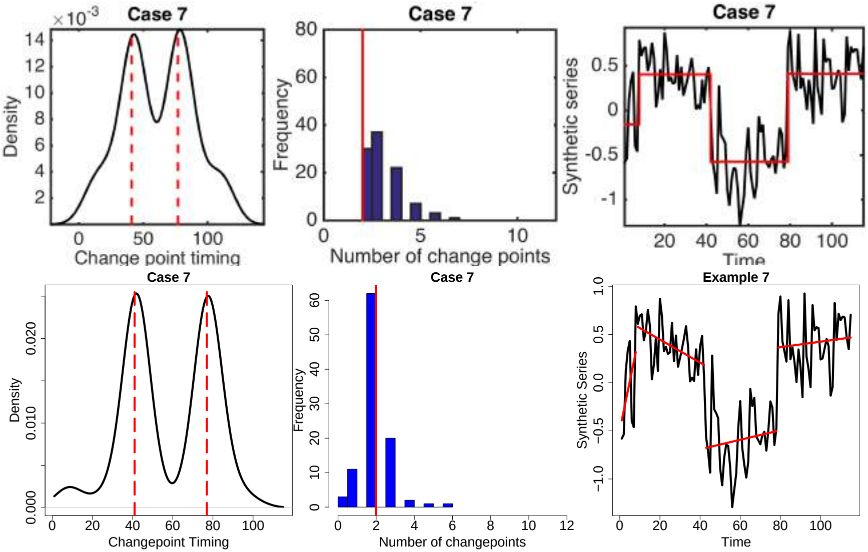
Mean+Change



Case 6
$\times 10^{-3}$
Density
Change point timing
Case 6
Frequency
Number of change points
Case 6
Synthetic series
Time
Case 6
Density
Changepoint Timing
Case 6
Frequency
Number of changepoints
Example 6
Synthetic Series
Time

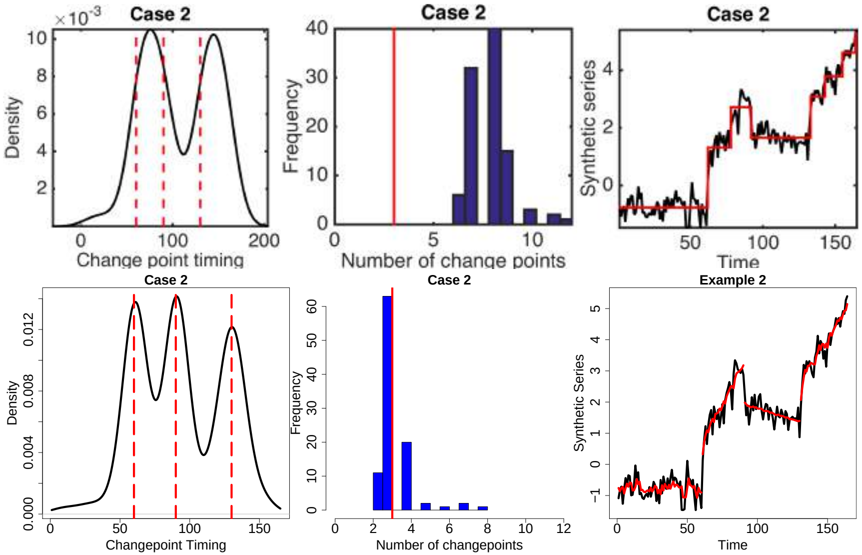
Mean+AR+Change

Data Science | Lancaster University

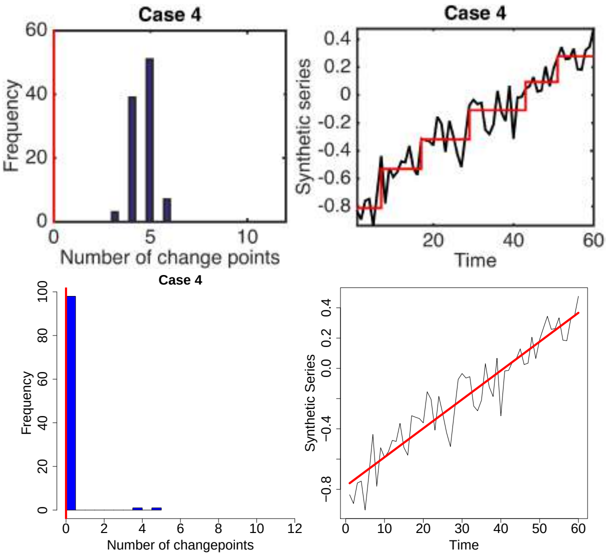


Trend+AR+change(trend)

Data Science | Lancaster University



Trend

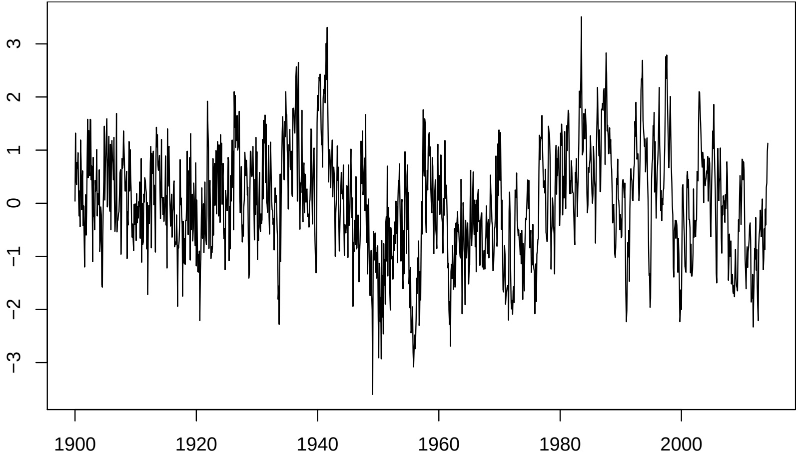


Case 4
Frequency
Number of change points
Case 4
Synthetic series
Time
Case 4
Frequency
Number of changepoints
Synthetic Series
Time

Pacific Decadal Oscillation

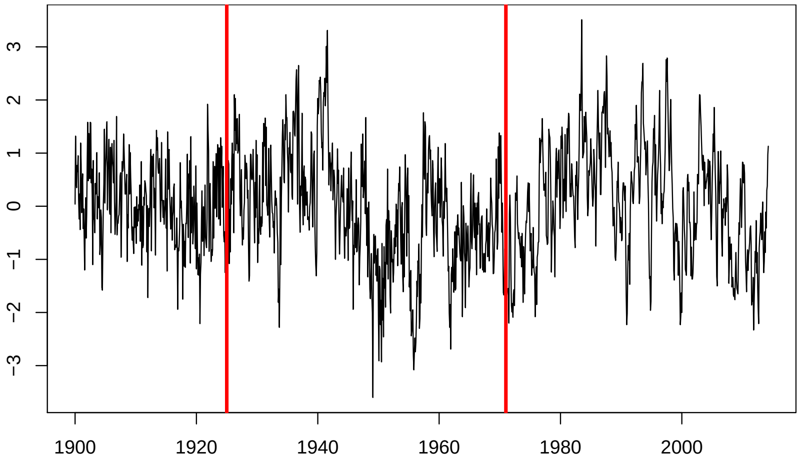
North Pacific Ocean

Monthly PDO



North Pacific Ocean - Trend(Mean)+AR+change

Monthly PDO



Summary

- Being able to find changepoints quickly is important
- Being able to fit several models is useful
- Automatic decision making saves time and bias
- Code is available within an R package (EnvCpt) on Github.

Acknowledgements

- PELT algorithm: Paul Fearnhead and Idris Eckley (Lancaster)
- Change in Trend: Rob Maidstone and Paul Fearnhead (Lancaster)
- Model selection and climate examples: Claudie Beaulieu (Southampton)

SECURE EPSRC Engineering and Physical Sciences Research Council

Assumptions of PELT

- Independence between segments
- Additivity of the cost function over segments
- Penalty that is linear in the number of changepoints

Theorem

Theorem

Define θ^* to be the value that maximises the expected log-likelihood

$$\theta^* = \arg\max \int \int f(y|\theta) f(y|\theta_0) dy \pi(\theta_0) d\theta_0.$$

Let θ_i be the true parameter associated with the segment containing y_i and $\hat{\theta}_n$ be the maximum likelihood estimate for θ given data $y_{1:n}$ and an assumption of a single segment:

$$\hat{\theta}_n = \arg\max_{\theta} \sum_{i=1}^{n} \log f(y_i|\theta).$$

Theorem cont.

Theorem

Then if

(A1) *denoting* $B_n = \sum_{i=1}^{n} \log \left[f(y_i|\hat{\theta}_n) - \log f(y_i|\theta^*) \right]$, *we have* $\mathbb{E}(B_n) = o(n)$ *and* $\mathbb{E}\left([B_n - \mathbb{E}(B_n)]^4\right) = \mathcal{O}(n^2)$;

(A2) $\mathbb{E}\left([\log f(Y_i|\theta_i) - \log f(Y_i|\theta^*)]^4\right) < \infty$;

(A3) $\mathbb{E}\left(S^3\right) < \infty$; *and*

(A4) $\mathbb{E}\left(\log f(Y_i|\theta_i) - \log f(Y_i|\theta^*)\right) > \frac{\beta}{\mathbb{E}(S)}$;

the expected CPU cost of PELT for analysing n data points is bounded above by Ln for some constant $L < \infty$.