Changing Climate, Changing Data
A journey of statisticians and climate scientists

Claudie Beaulieu, UC Santa Cruz & Princeton University
(beaulieu@ucsc.edu)
Rebecca Killick, Lancaster University & UC Santa Cruz
(r.killick@lancaster.ac.uk)

Who we are

Claudie

- PhD in Water Sciences
- Main interest: Address climate and **environmental change** related questions using **statistics** & data science, mostly **time series** and particularly **changepoints**

Rebecca

- PhD in Statistics
- Main interest: developing **statistical** models and methods for **nonstationary time series** analysis and **changepoints**, multiple domains of applications including the **environment**

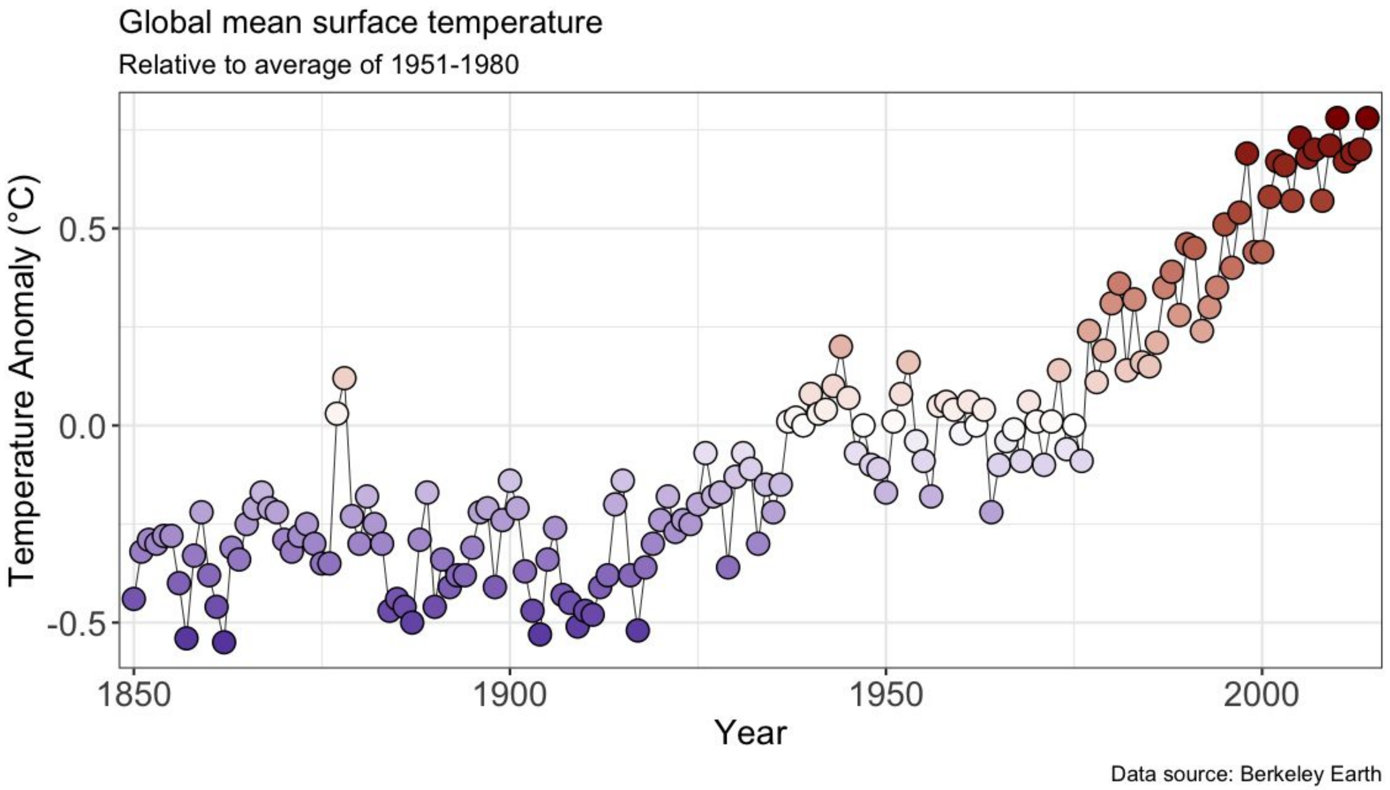
Where we met

Isaac Newton Institute for Mathematical Sciences

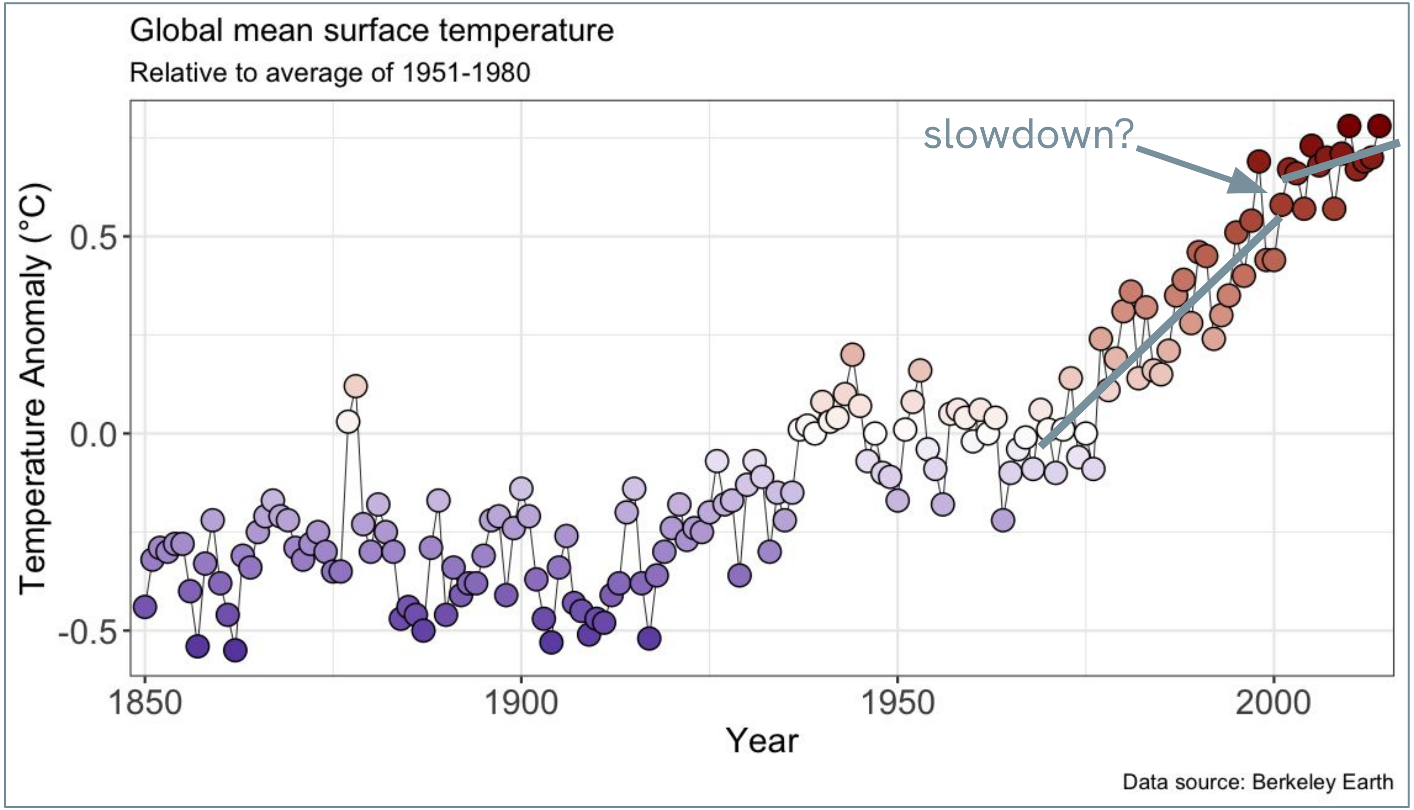
Inference for change-point and related processes

13 January 2014 to 7 February 2014

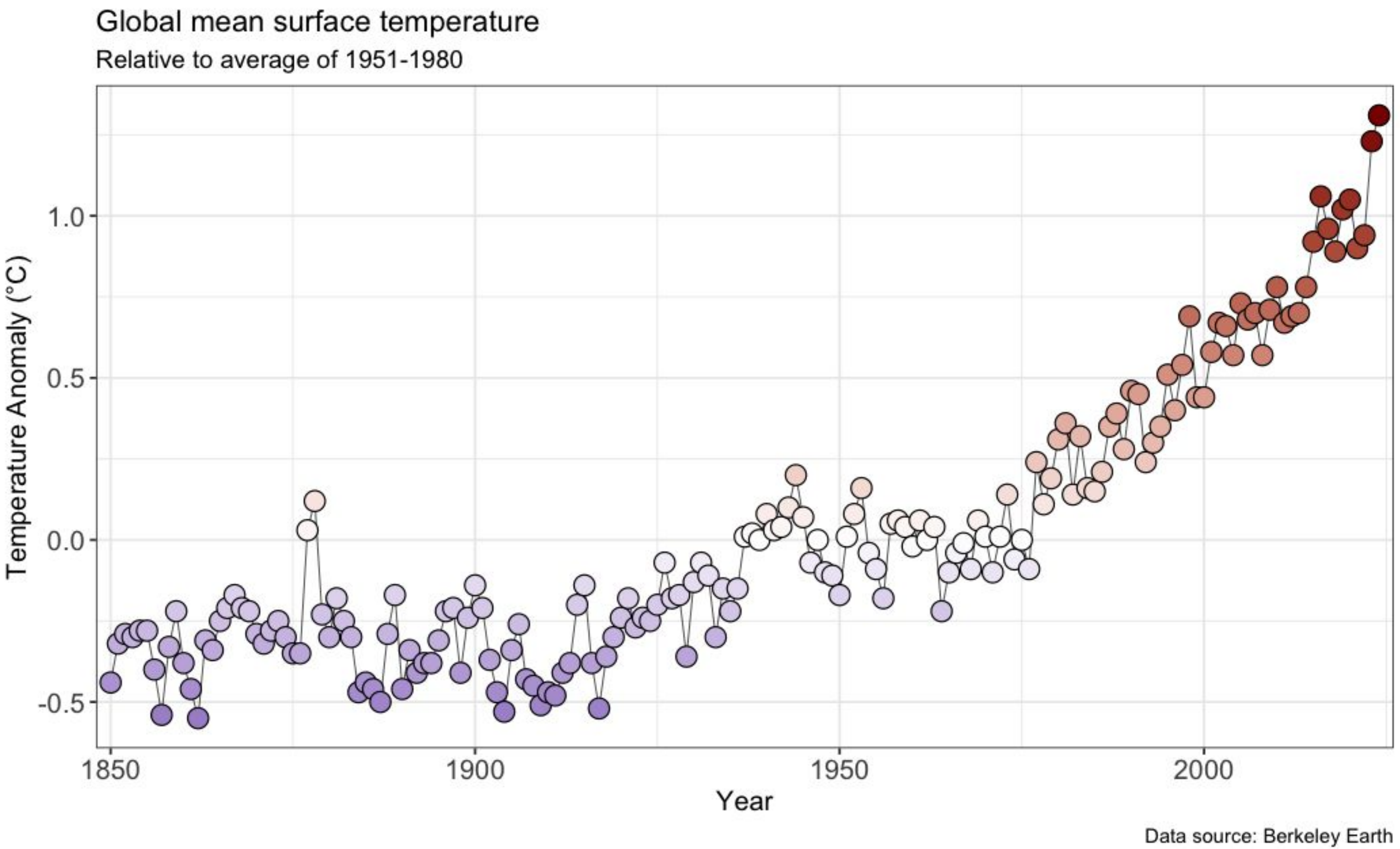
A motivation - the rate of global warming



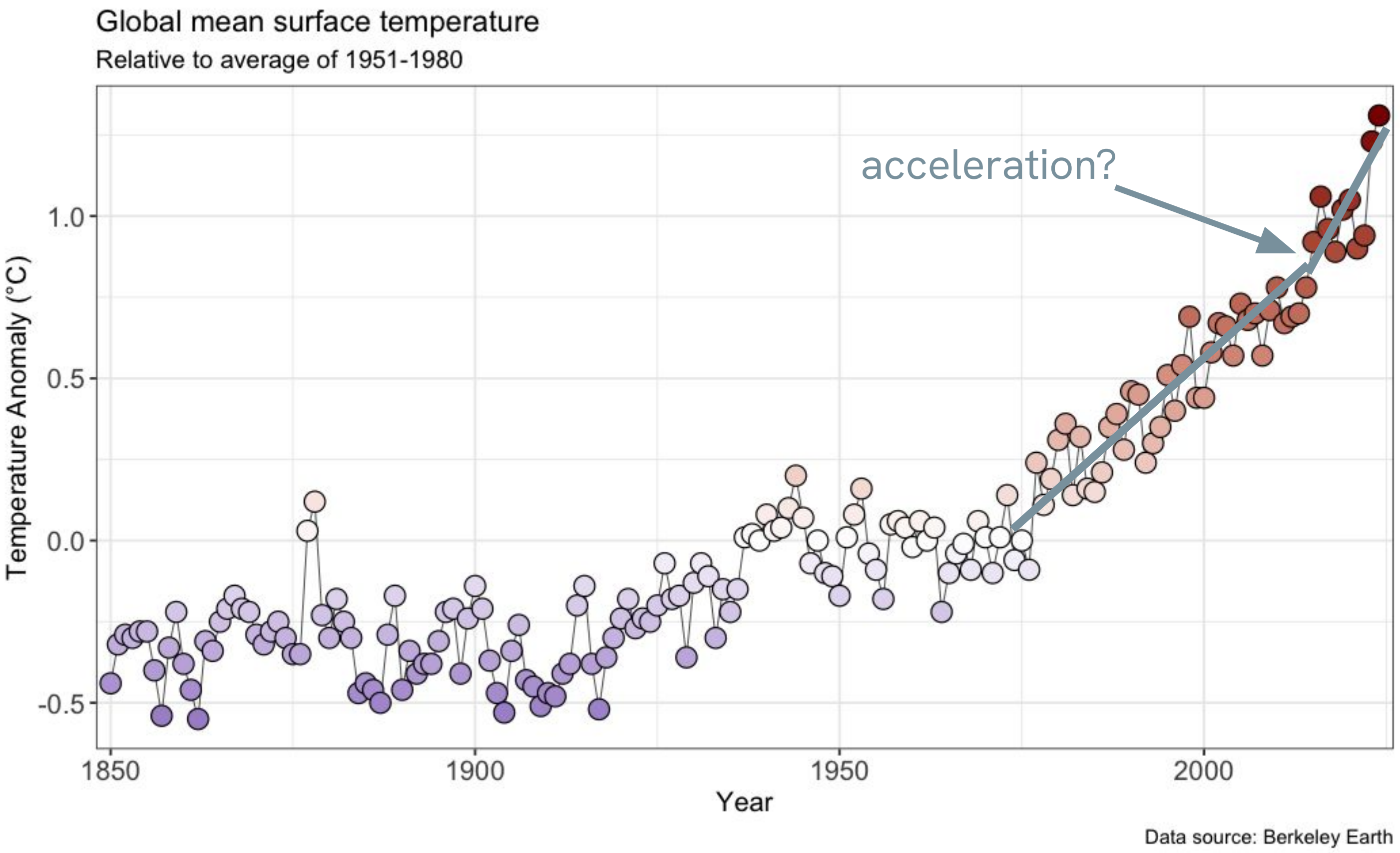
A motivation - the GMST record then



A motivation - the GMST record now



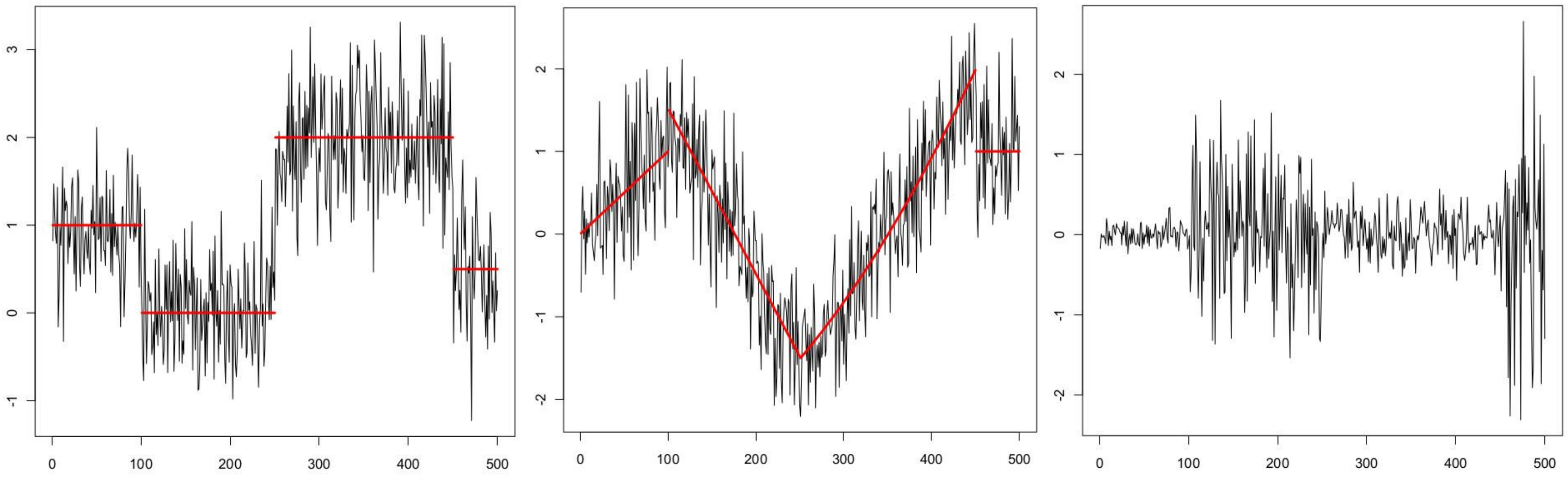
A motivation - the GMST record now



Part 1: Has there been a slowdown in warming?

What do we (statistically) mean by changepoint?

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 $y_{\tau+1}, \ldots, y_n$.



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 > β for some (suitably chosen) β. If we infer a changepoint its position is estimated as

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

Multiple changepoints

Define m to be the number of changepoints, with positions

$$\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_{m+1}) \text{ where } \tau_0 = 0 \text{ and } \tau_{m+1} = n.$$

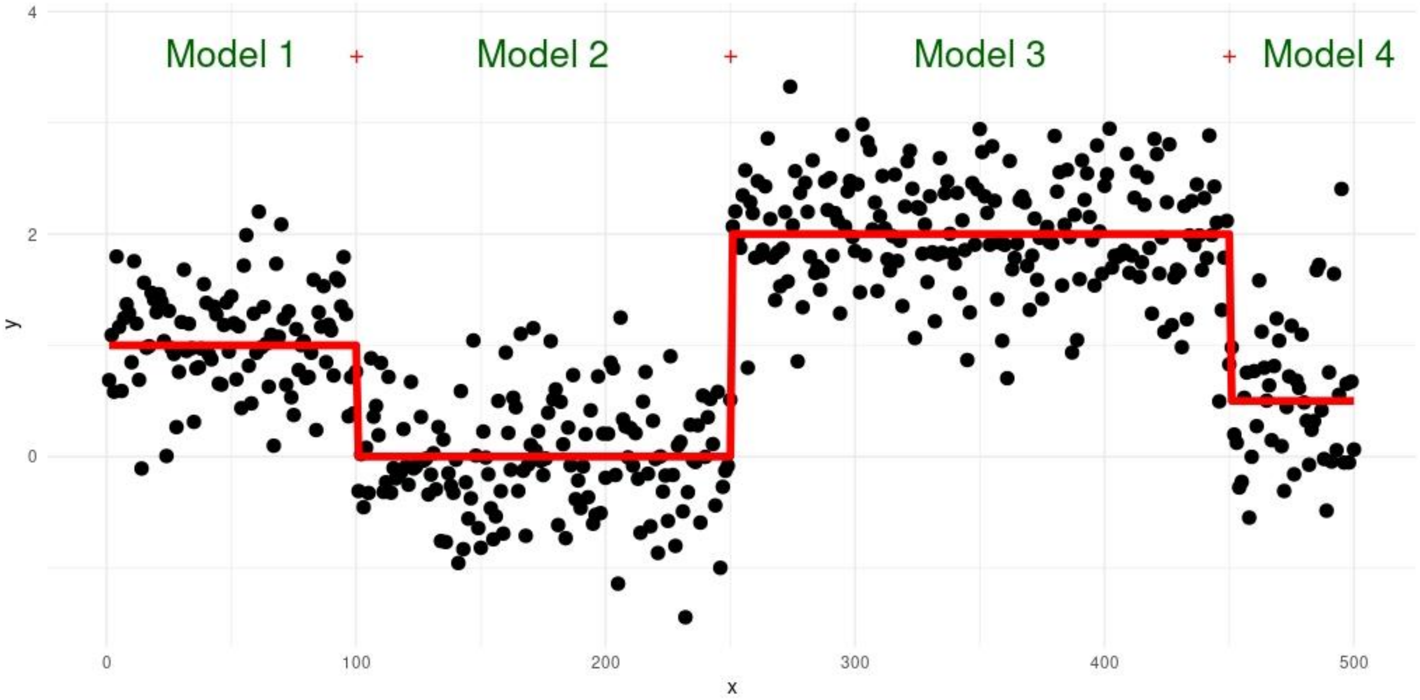
Then one application of the Likelihood ratio test can be viewed as

$$\min_{m \in \{0,1\}, \boldsymbol{\tau}} \left\{ \sum_{i=1}^{m+1} \left[-\ell(y_{\tau_{i-1}:\tau_i}) \right] + \beta m \right\}$$

Repeated application is thus aiming to minimise

$$\min_{m, \boldsymbol{\tau}} \left\{ \sum_{i=1}^{m+1} \left[-\ell(y_{\tau_{i-1}:\tau_i}) \right] + \beta m \right\}$$

Problem

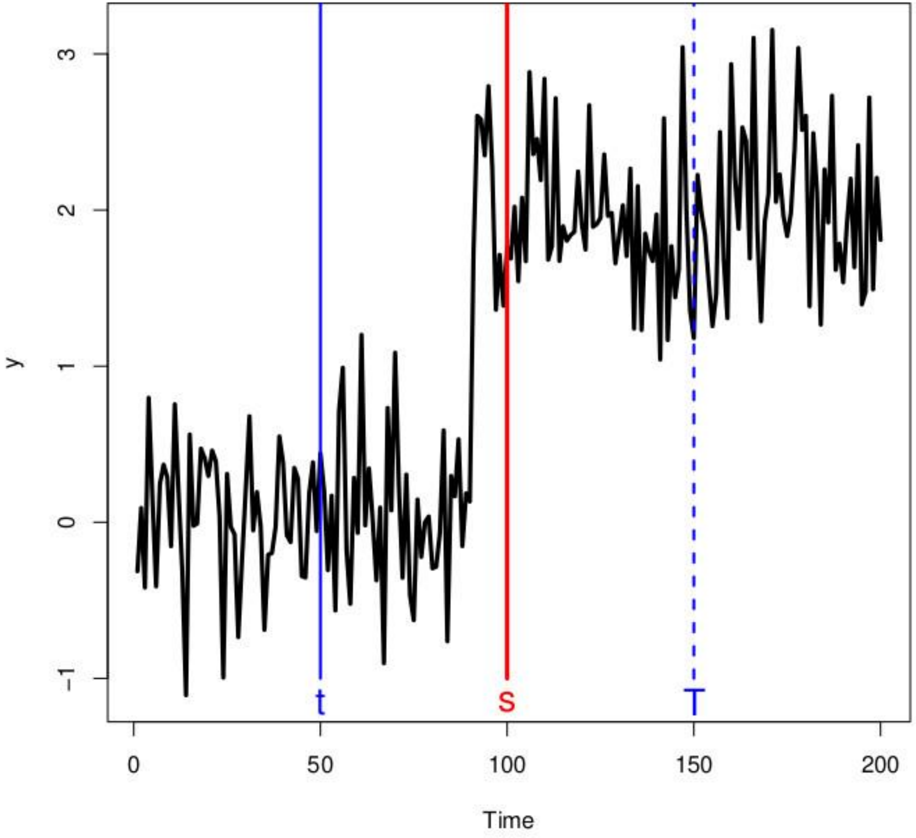


- How many changes?
- Where are the changes?

2^{n-1} possible solutions!

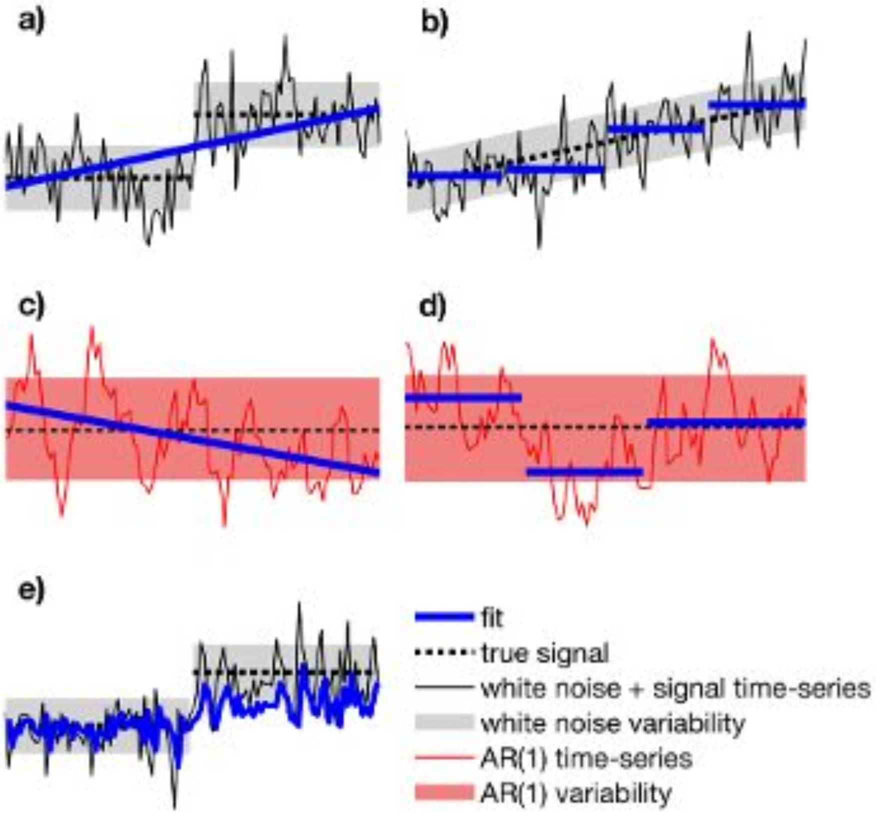
PELT algorithm (Killick et al. 2012)

- 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.



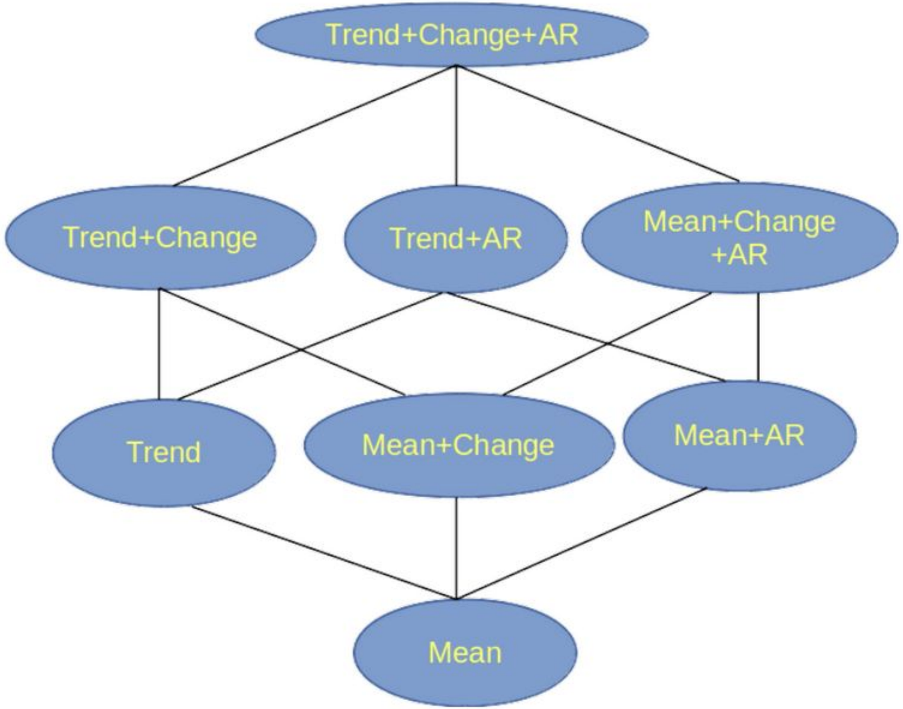
Separating signal and noise

- Tools are not enough! (Lund et al. 2023)
- Potentially hundreds of time series to analyze
- Different series may have different properties including: trend, autocorrelation, changes



Beaulieu and Killick, 2018, JCLIM

Model selection



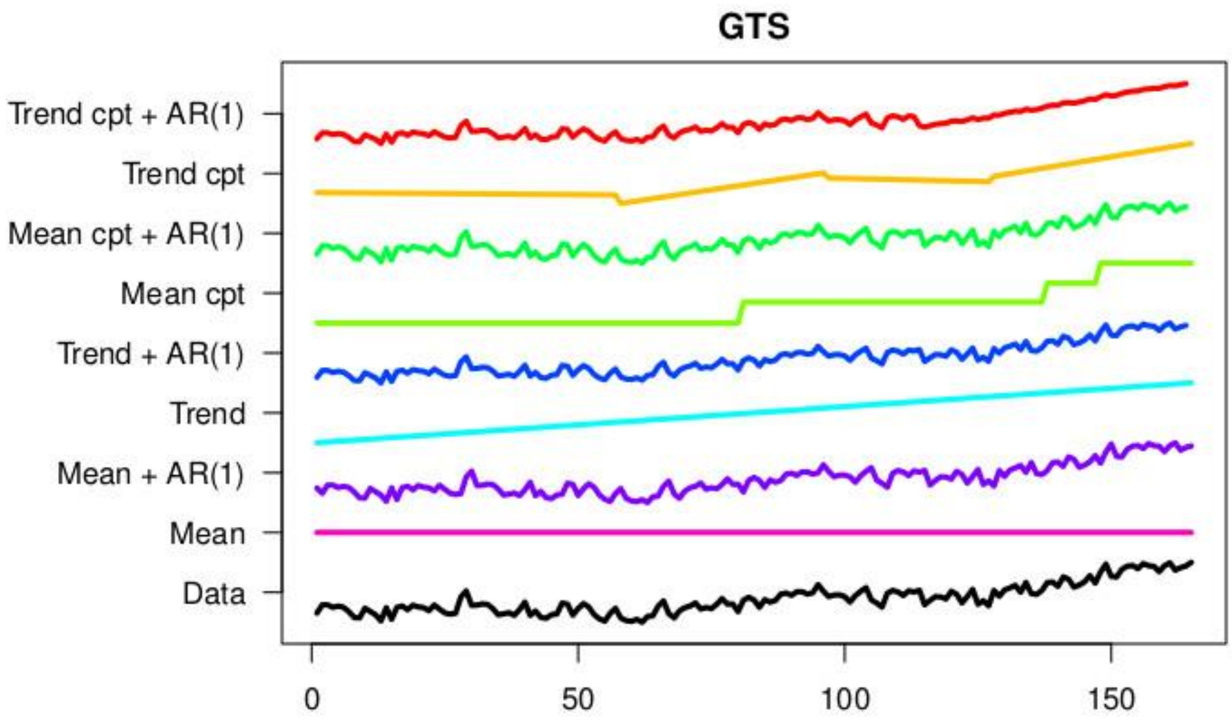
`EnvCpt`: select the most parsimonious but accurate model for the data.

Simple to extend with other types of models.

Choose the “best” model according to an information criterion e.g., AIC.

Killick et al., 2021, CRAN

EnvCpt for GMST (up to 2014)

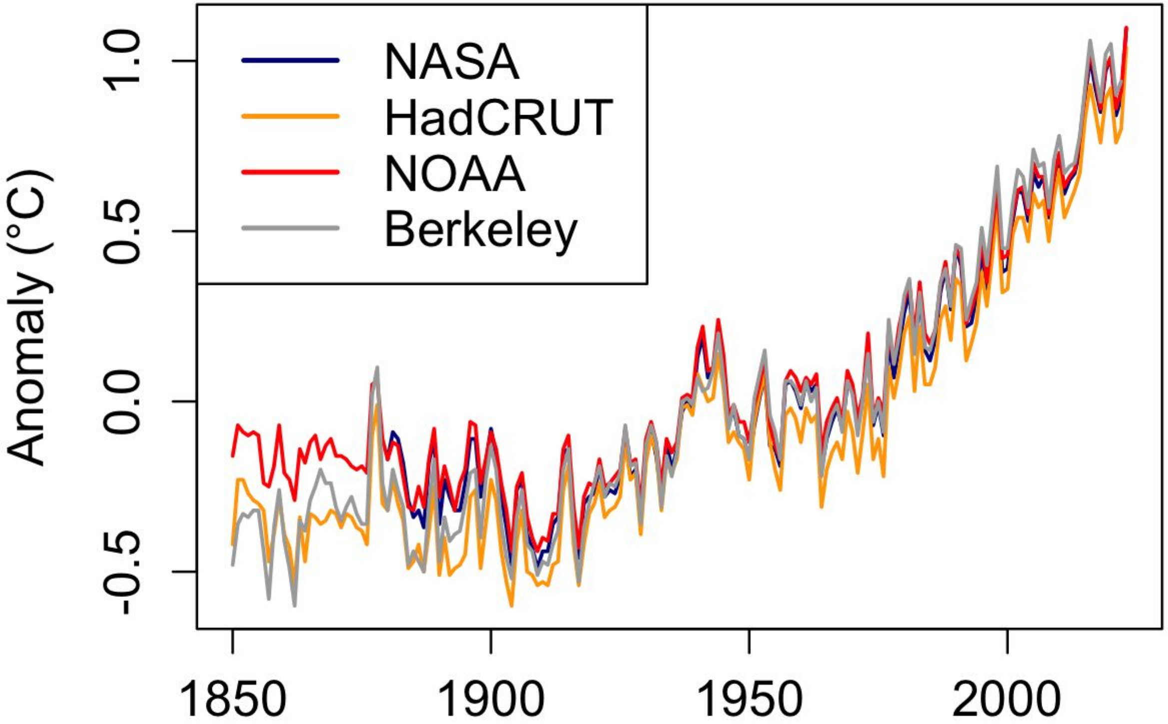


No change detected beyond the 1970s!

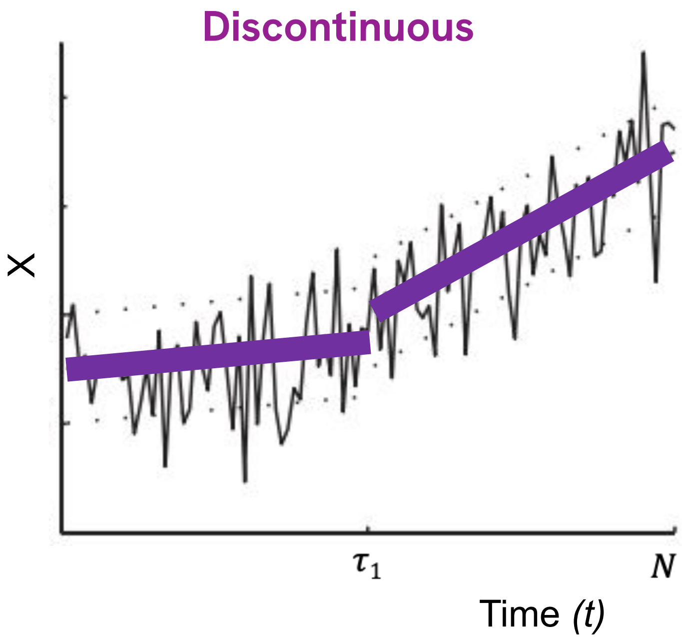
Beaulieu and Killick, 2018, JCLIM

Part 2: Has there been a recent acceleration in the rate of warming?

Is global warming speeding up?



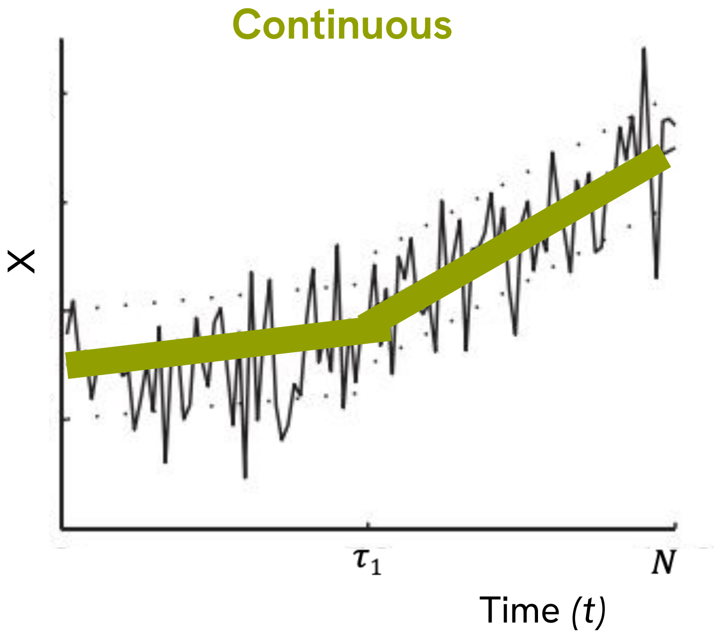
Changepoint models



$$X_t = E[X_t] + \epsilon_t$$

$$E[X_t] = \begin{cases} \alpha_1 + \beta_1 t, & 0 = \tau_0 < t \leq \tau_1, \\ \alpha_2 + \beta_2 t, & \tau_1 < t \leq \tau_2, \\ \vdots & \vdots \\ \alpha_{m+1} + \beta_{m+1} t, & \tau_m < t \leq \tau_{m+1} = N \end{cases}$$

Changepoint models



$$X_t = E[X_t] + \epsilon_t$$

$$E[X_t] = \begin{cases} \alpha_1 + \beta_1 t, & 0 = \tau_0 < t \leq \tau_1, \\ \alpha_2 + \beta_2 t, & \tau_1 < t \leq \tau_2, \\ \vdots & \vdots \\ \alpha_{m+1} + \beta_{m+1} t, & \tau_m < t \leq \tau_{m+1} = N \end{cases}$$

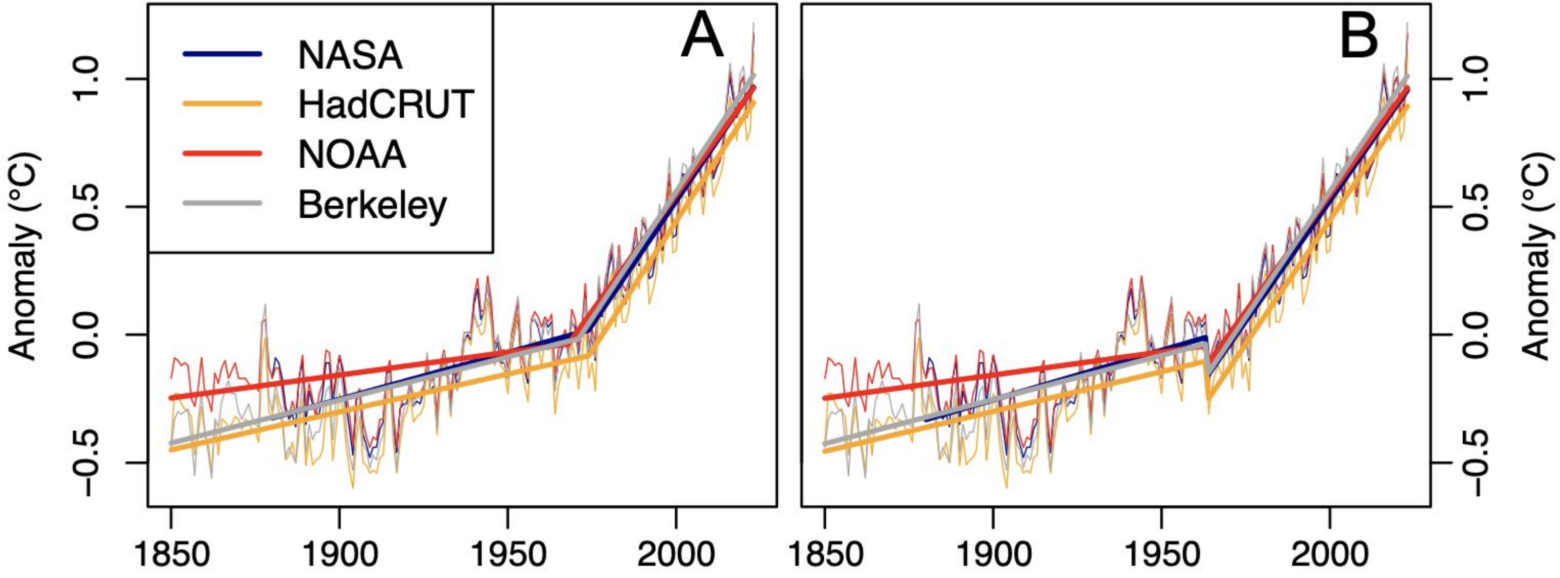
Continuity constraint:

$$\alpha_i + \beta_i \tau_i = \alpha_{i+1} + \beta_{i+1} \tau_i, \qquad 1 \leq i \leq m$$

Methodology

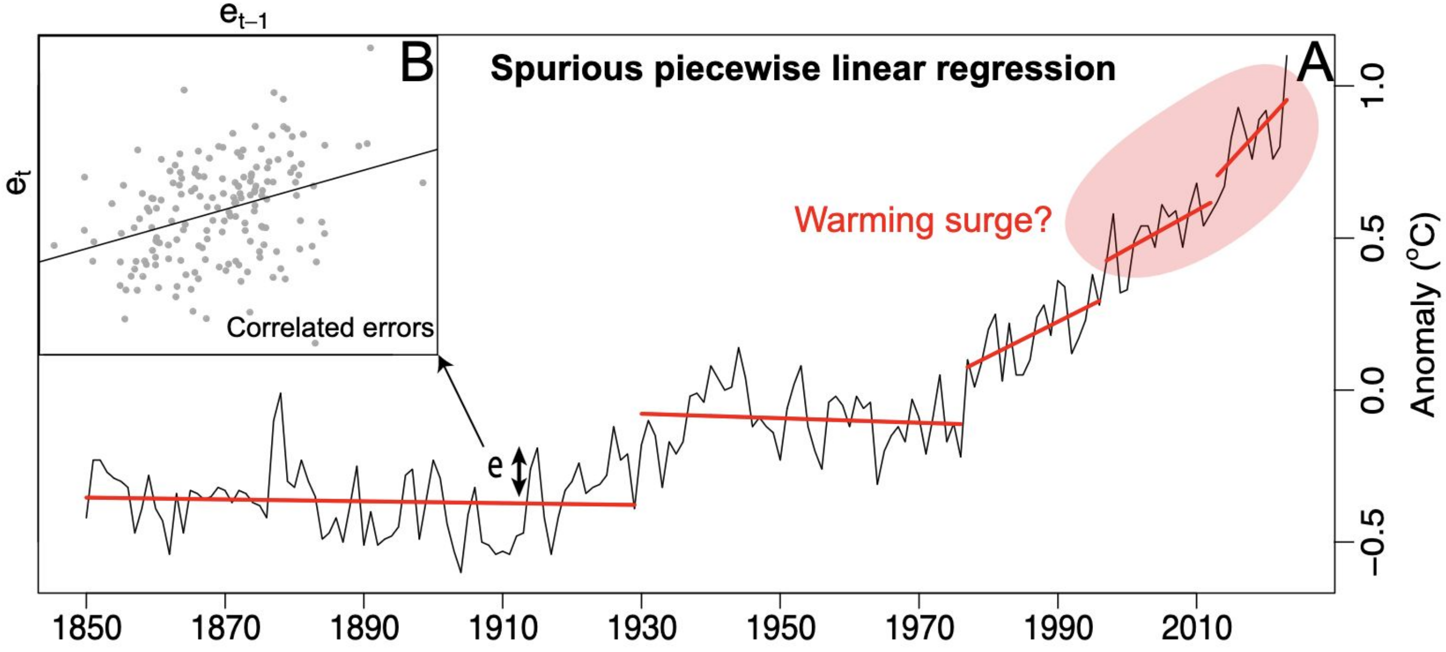
- We fit a series of changepoint models to accommodate different assumptions
    - Regression function:
        - Continuous vs discontinuous changepoints
    - Errors:
        - AR(1) fixed
        - AR(1) varies at changepoints
        - AR(4) fixed based on inspection of the residuals
        - Independent, just to see what happens
- The goal here is to assess if any configurations above yield an acceleration in warming after the 1970s

Post 1970s acceleration not detectable - with multiple types of changepoint models



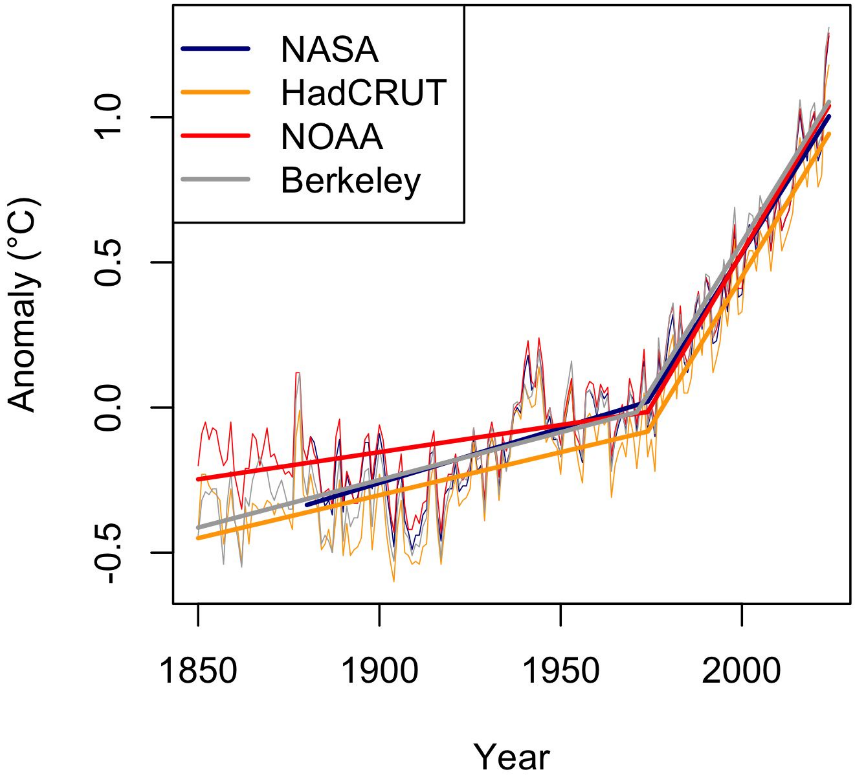
Beaulieu et al, Communications Earth & Environment, 2024

One model detects an acceleration, but is invalid

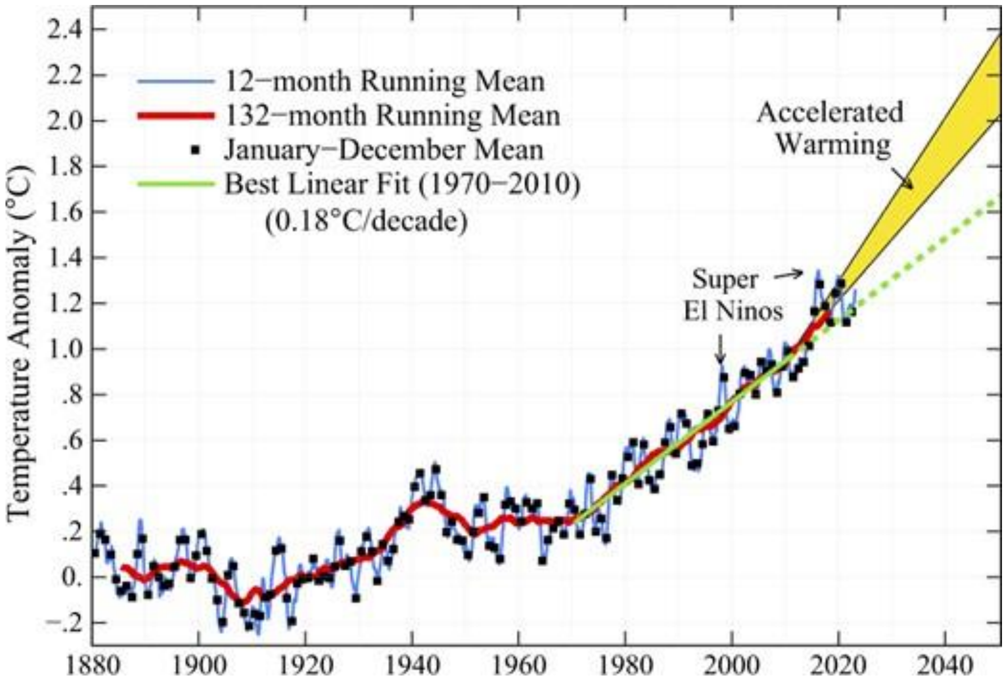


Beaulieu et al, Communications Earth & Environment, 2024

Results hold with 2024 observations



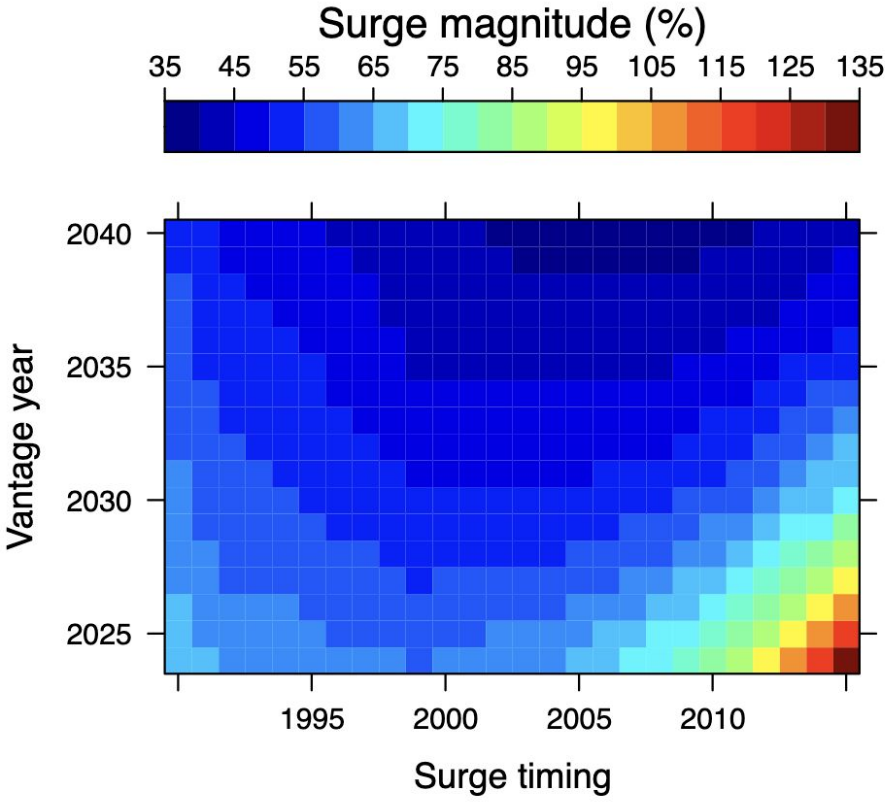
Accelerated warming prediction



Hansen et al. 2023

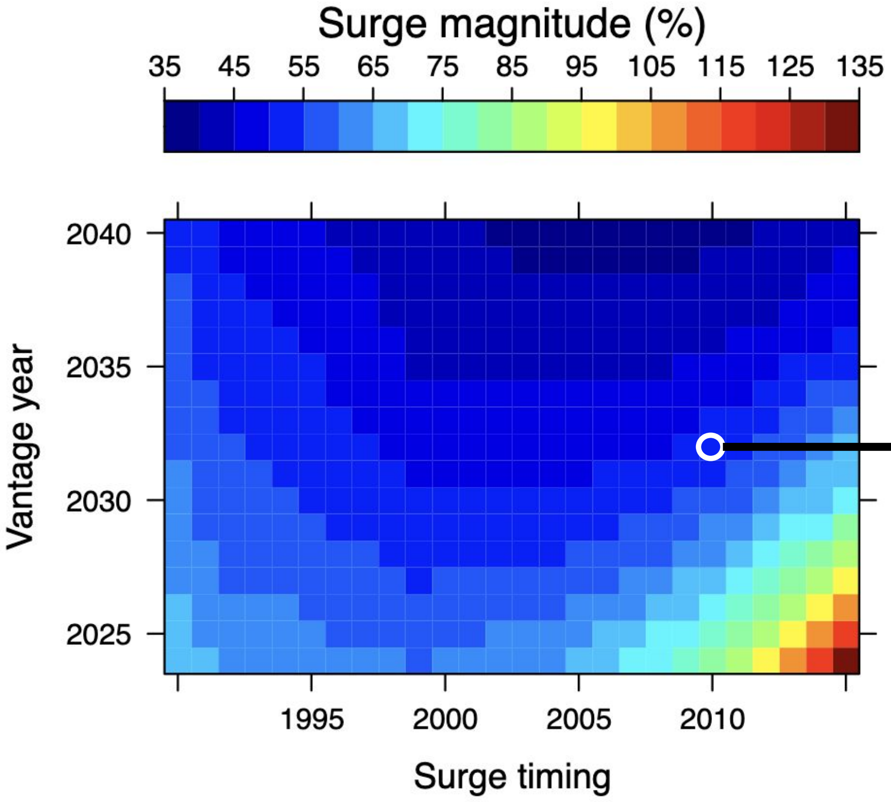
An acceleration in the rate of warming is predicted after 2010 due to reduction in cooling aerosols.

When could we detect an acceleration?



Beaulieu et al, Communications Earth & Environment, 2024

When could we detect an acceleration?



A change in 2010 is too recent and/or too small to be detectable at this time.

For example, an increase of ~50% in the rate of warming in 2010 would be detectable in the early 2030s.

Beaulieu et al, Communications Earth & Environment, 2024

Summary of our global detection study

- An acceleration in global warming is not yet statistically detectable, based on an analysis of surface temperature observations.

- These results do not challenge that anthropogenic activities are causing long-term global warming nor that 2023 & 2024 temperatures were record-breaking.

- On social media our paper was widely used to support unrelated and counterfactual claims to our study

Thoughts on collaboration

- Shared commitment to understanding each other and working together
- Mutual respect for each others research needs and interests

Thoughts on collaboration

- Fruitful and fun collaboration!
- There is intrinsically a large statistical component to climatology … and a need for collaboration with statisticians
- How the two communities can better work together more generally
    - Accessibility to climate scientists
    - Acknowledge statisticians contributions
- Find your people!

Thank you!

Beaulieu, C., Gallagher, C., Killick, R., Lund, R., Shi, X., (2024) A recent surge in global warming is not detectable yet. *Communications Earth & Environment, 5, 576.*

Lund, R.B., Beaulieu, C., Killick, R., Lu, Q., Shi, X. (2023) Good Practices and Common Pitfalls in Climate Time Series Changepoint Techniques: A review. *Journal of Climate, 36, 8041-8057.*

Shi, X, Beaulieu, C., Killick, R., Lund, R. (2022) Changepoint Detection: An Analysis of the Central England Temperature Series. *Journal of Climate*, 35, 2729-2742.

Killick, R., Beaulieu, C., Taylor, S., Hullait, H. (2021) EnvCpt: Detection of structural changes in climate and environment time-series. R package version 1.1.3.

Beaulieu, C., Killick, R., Ireland, D., Norwood, B. (2020) Considering long-memory when testing for changepoints in surface temperature: A classification approach based on the time-varying spectrum. *Environmetrics*

Beaulieu, C., Killick, R. (2018) Distinguishing trends and shifts from memory in climate data. Journal of Climate, 31, 9519-9543.

SECURE

EPSRC

NSF CAREER Award