School of Mathematical Sciences | Lancaster University

Understanding tree water usage and stress via sap flux density time series

Rebecca Killick(r.killick@lancs.ac.uk)
Joint work with Andy Hirons (Myerscough), Mengyi Gong & Ian Dodd (Lancaster)
EnvEcoStats 2025

What is the problem?

- Tree planting needs to match environment
- Changing climate is challenging
- Lack of understanding of drought and waterlogging effects across species

What type of data?

Time series data from the (lab) pot water logging experiment

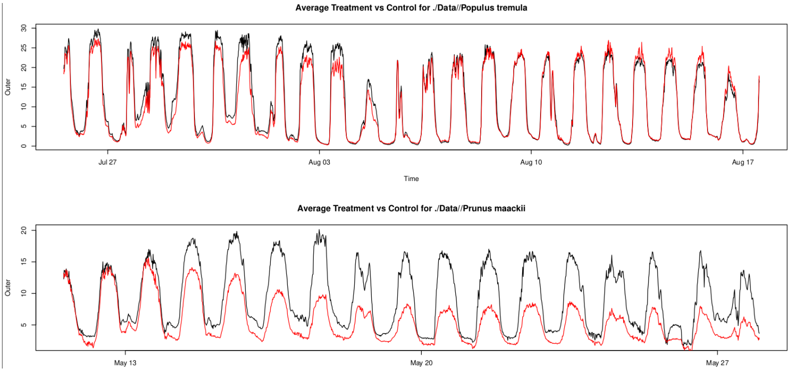
- Three phases: pre-treatment, waterlogged, recovery
- Treatment v.s. Control
- Sap flux density
- Weather data: air temperature, relative humidity, photosynthetically active radiation (PAR) and rainfall

A range of temperate tree species

- Acer rubrum, Alnus glutinosa, Carpinus betulus, Liquidambar styraciflua, Populus tremula, Prunus maackii, Salix alba, Tilia cordata

Long term we want to be able to detect tree stress in the field.

Motivation



- Multiple seasonal frequencies
- Regressors for conditions
- Explanatory series

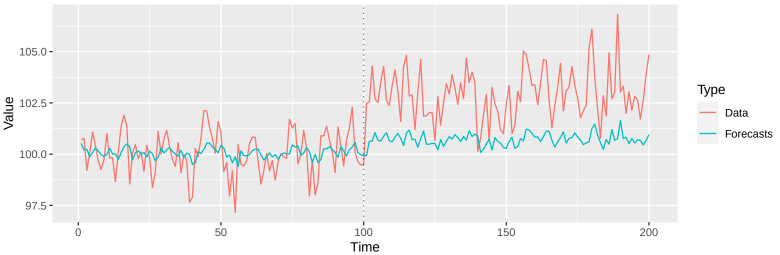
Problems

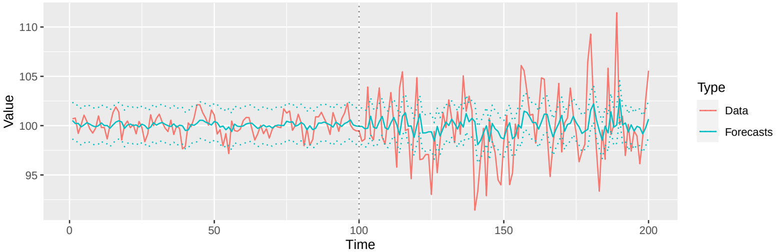
Automatically identifying changes in data requiring complex modelling is problematic:

- Most methods need to be able to fit the model after the change ...
- ... leading to delays in online detection
- ... poor model estimation in short segments
- ... identifiability issues when changepoints are close together
- ... inability to fit long seasonal frequencies (e.g. yearly)

- Those that don't are often restricted to mean/linear trend series only.

Using Forecasts





Poor Forecasts

A change in the underlying process (data) could cause our forecasting model to behave poorly!

Intuitively:

- Mean change in raw data $\rightarrow$ forecasts become biased.
- Variance change in raw data $\rightarrow$ prediction intervals become poor.
- Other changes in raw data $\rightarrow$ potential combination of biased and inaccurate prediction intervals.

Our Aim

Statistical Aim: Create a framework for detecting changes in complex models which doesn't require fitting the model to the post-change data.

Intuitively, changes in the raw data = forecasts become poor.

Solution: Use sequential forecast errors.

Bonus: This also provides a framework for identifying changes in forecast performance.

Detecting Changes

Detector:

- Measures the difference between mean in training period and mean in monitoring period.
- CUSUM Detector: $Q(m,k) = \sum_{t=m+1}^{k} y_t - \frac{k}{m}\sum_{t=1}^{m} y_t$.
- Page's CUSUM: $D(m,k) = \max_{0\leq t\leq k} |Q(m,k) - Q(m,t)|$

Stopping Rule:

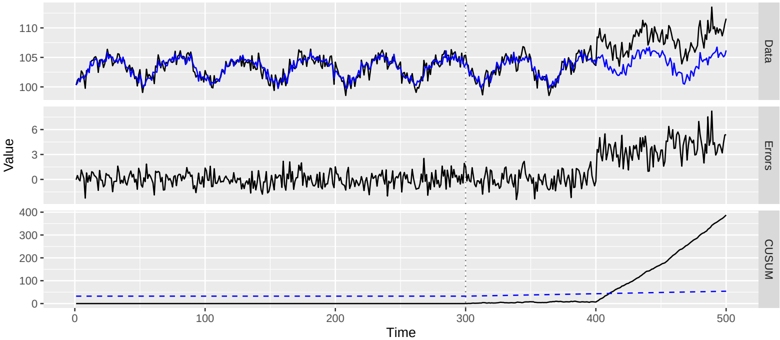
- Defines a rule for when to flag a change.
- Typically when the detector exceeds some threshold.
- Threshold controls false positive rate.
- $\tau = \min\{k \geq 1 : D(m,k) \geq c\hat{\sigma}_m g(m,k)\}$.

Solution

What about the forecast errors?

$$e_t = Y_t - \hat{y}_t(1)\ , \qquad Q(m,k) = \sum_{t=m+1}^{k} e_t - \frac{k}{m}\sum_{t=1}^{m} e_t$$

- Forecasting model accounts for data complexities.



Theory & Simulations

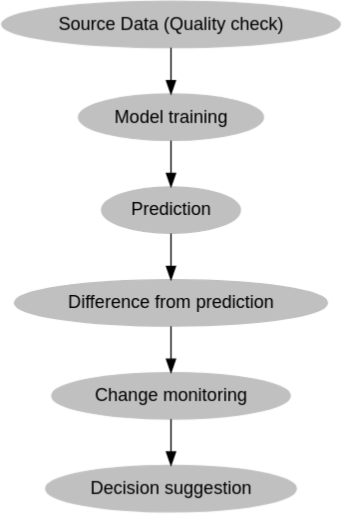
Theory: Under certain assumptions,

- Mean change in raw data $\rightarrow$ mean change in forecast errors.
- Mean and/or variance change in raw data $\rightarrow$ mean change in squared forecast errors.

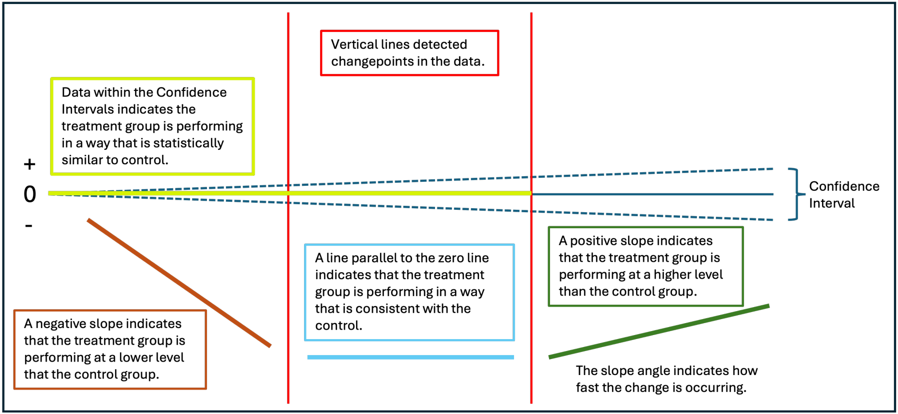
Simulations:

- Using forecast errors performs better than raw CUSUM.
- Can also be used to detect changes in,
  - Trend.
  - Dependence structure.
  - Error structure.

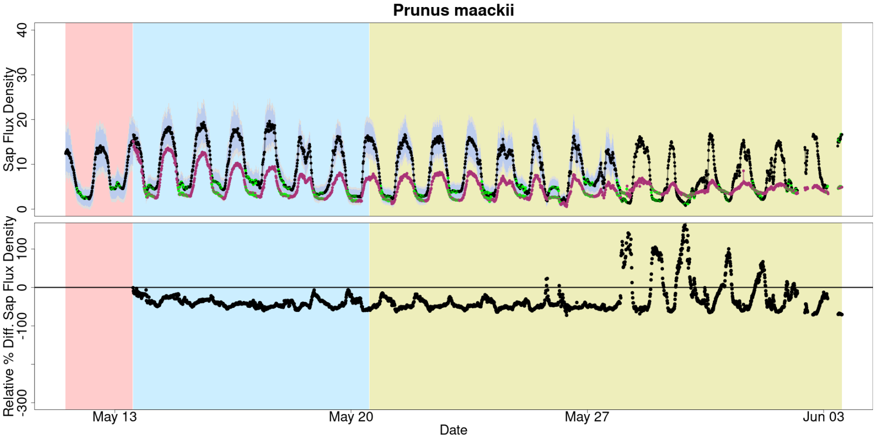
Analysis Flow



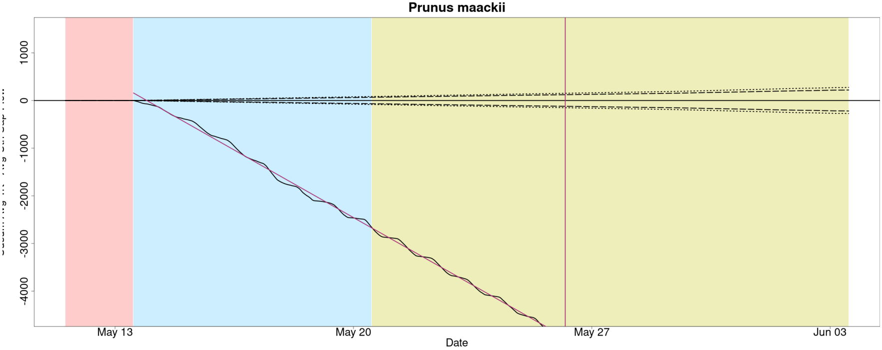
Interpretation



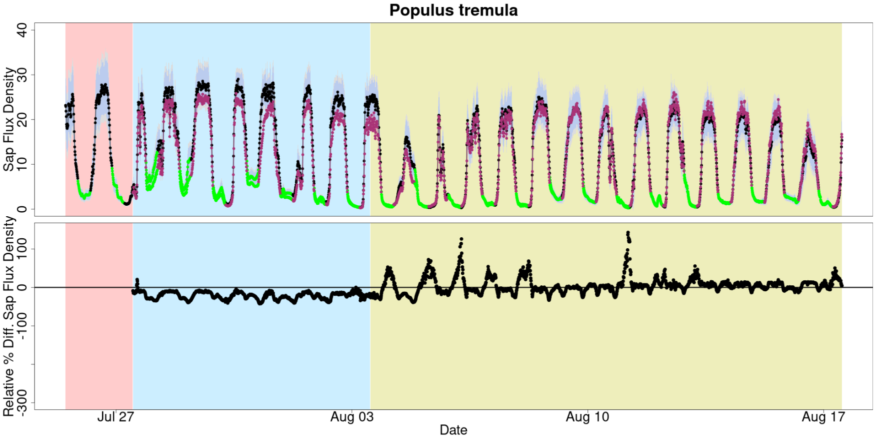
Waterlogged



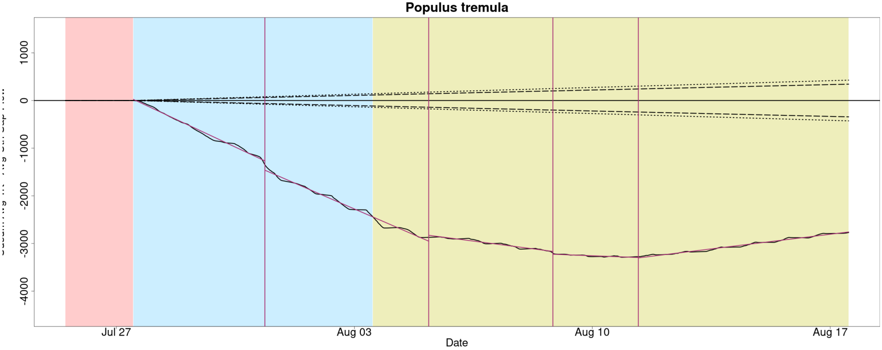
Waterlogged



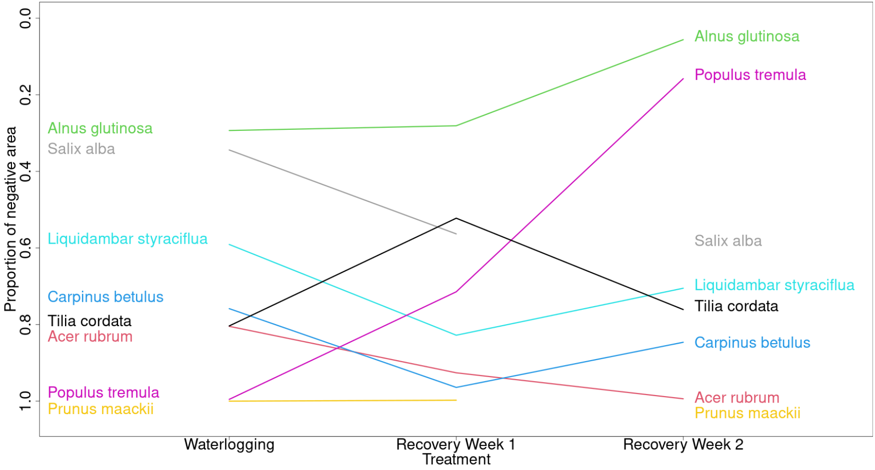
Waterlogged



Waterlogged



Waterlogged Species



Conclusions

- Changepoint analysis on the difference between the average sap flow for treatment and model predictions within a species.
- Detect mean changes in the difference (forecast errors) time series
  - CUSUM test for a significant deviation from zero.
- Detect trend changes in the trajectories of the CUSUM statistics
  - To ascertain the tolerence to treatment of the species
  - Allows species ranking