Mathematics & Statistics | Lancaster University

Automatic Detection of Seasonality Using Wavelets

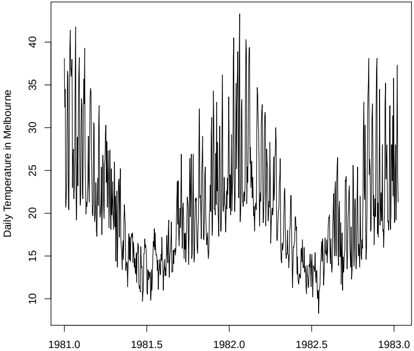
Rebecca Killick
and Ben Norwood
ONS Workshop 2018/03

Contents

MOTIVATION

Motivation

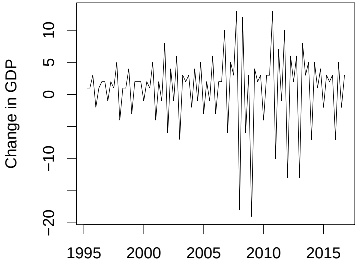
- As data collection is growing with the rise of big data, larger seasonal periods are expected.
- Data shows daily, monthly, yearly patterns.
- Automatic Detection of such periods would aid adjustment of them substantially.

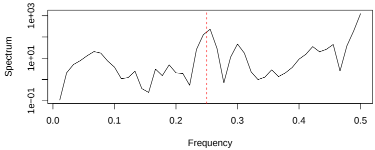


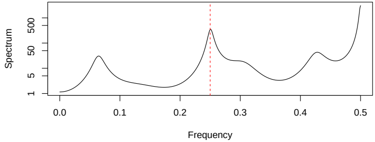
EXISTING METHODS

Fourier Methods

- Most common way to determine periodicity
- Compute the (smoothed) Fourier periodogram
- Identify spectral peaks as periodicities

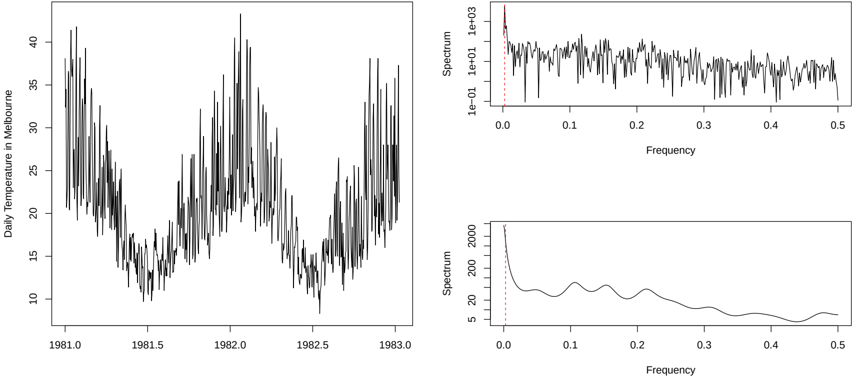






- This requires human intervention or peak-detection systems.

- For longer periodicities the cycle is often too close to the origin to detect.



Detection

- Manually, i.e. by eye or guessing
- Seasonal F test (Friedman Test)
- Automatically with visual significance (Soukup & Findley 1999, McElroy & Roy 2017)
- Automatically with findFrequency (first peak of smoothed periodogram)

There are other ways to determine seasonality using autocorrelations at various lags.

A NEW APPROACH

Wavelets

Discrete Non-Decimated Approach

- The discrete transform can be seen as a filtering operation over different bands.
- Given a time series Y_t we calculate the wavelet coefficients as

$$d_{j,k} = \frac{1}{\sqrt{2^j}} \left[\left\{ \sum_{i=0}^{2^{j-1}} Y_{k+i} \right\} - \left\{ \sum_{i=2^{j-1}+1}^{2^j-1} Y_{k+i} \right\} \right],$$

for

$$t = 1, 2, \ldots, n = 2^J,$$
$$j = 1, 2, \ldots J - 1,$$
$$k = 1, \ldots, 2^J - 2^j + 1.$$

Periodic Series

Length of period

- Throughout we now assume the series we are dealing with can be expressed as a sinusoid with noise:

$$Y_t \sim \sin\left(\frac{2\pi t}{p}\right) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \quad p \in \mathbb{R}$$

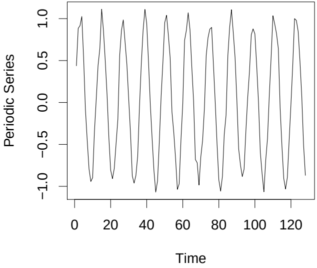


Figure: $p = 12, \sigma_\epsilon = 0.1$

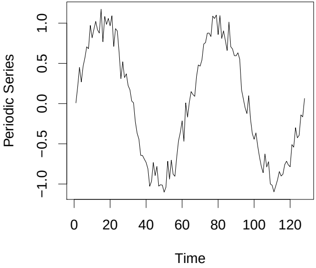
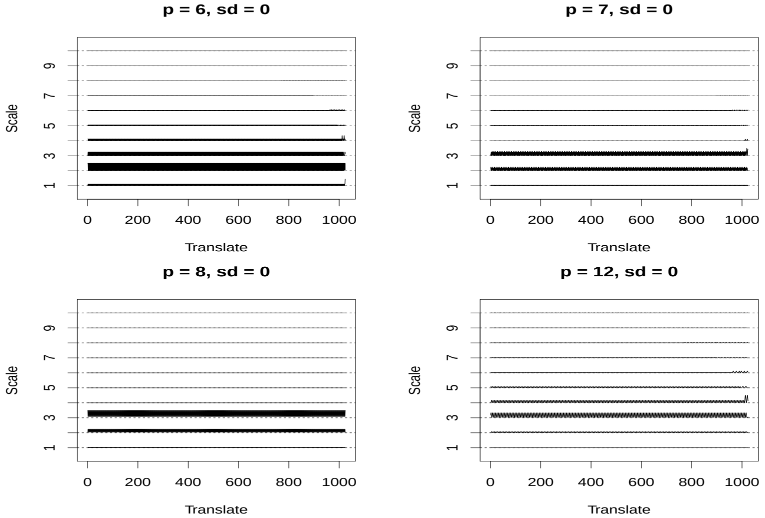


Figure: $p = 64, \sigma_\epsilon = 0.1$

Wavelet Representation

Square of the $d_{j,k}$ coefficients



Wavelet Energy

Definition

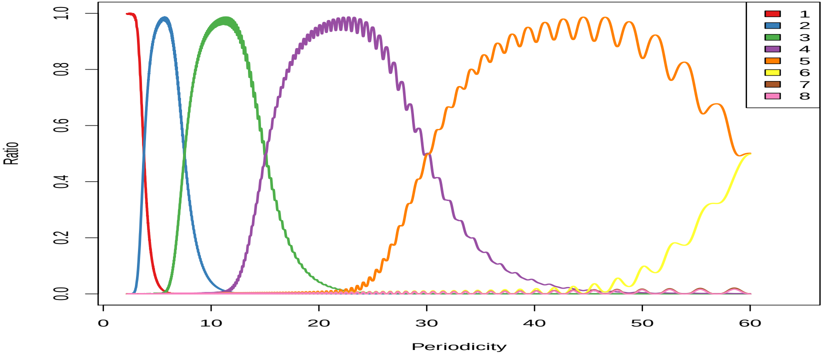
- To monitor the variance of each scale, we analyse the squares of the coefficients.
- This is such that we are monitoring the 'energy' of each scale.
- We can normalize this energy at each scale to give Relative Wavelet Energy:

$$RWE_j = \frac{\sum_{k=1}^{n} d_{j,k}^2}{\sum_{j=1}^{J} \sum_{k=1}^{n} d_{j,k}^2}.$$

Wavelet Energy

A Beautiful Relationship

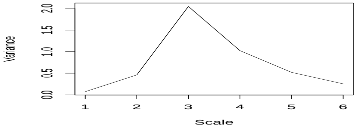
If we look at how the variance for each periodicity across the scales we get an interesting relationship.



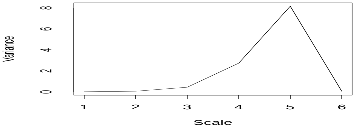
Variance Profiles

A Representation of a Series

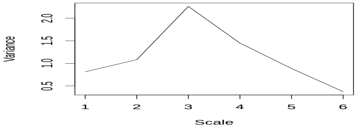
For each series we can now take a wavelet transform, and monitor the variance of the coefficients on each scale:



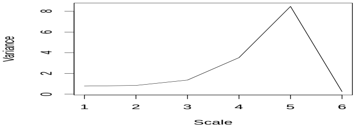
(a) p = 12, $\sigma_\epsilon = 0.1$



(b) p = 36, $\sigma_\epsilon = 0.1$



(a) p = 12, $\sigma_\epsilon = 1$



(b) p = 36, $\sigma_\epsilon = 1$

Theoretical Results

Theorem

Using the variance profile, on a sample of size T, we have proved that:

$$\mathbb{E}\left(\hat{d}^2_{j,k} - d^2_{j,k}\right)^2 \to 0 \qquad \text{as } T \to \infty$$
$$\hat{p} \to p \qquad \text{as } T \to \infty.$$

SIMULATIONS

Results

A Comparison

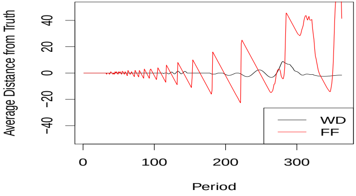
- Compare to the *findFrequency* function in *forecast* R package.
- We measure how close the predicted periodicity is to the truth.
- Also record computing time taken.

Simulation Constants:
$N = 1000$ Amount of realisations
$T = 1024$ Length of each series
$p \in [3, 365]$ periodicity simulated

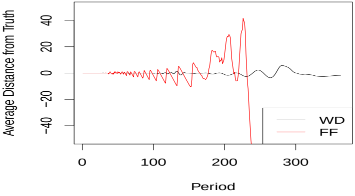
Weighted by number of coefficients in scale.

Results

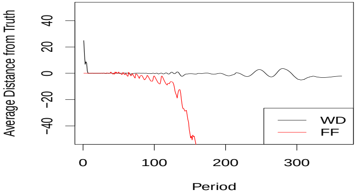
Average Distance



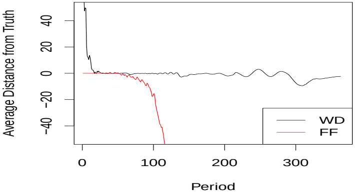
(a) $\sigma_\epsilon = 0.1$



(b) $\sigma_\epsilon = 0.4$



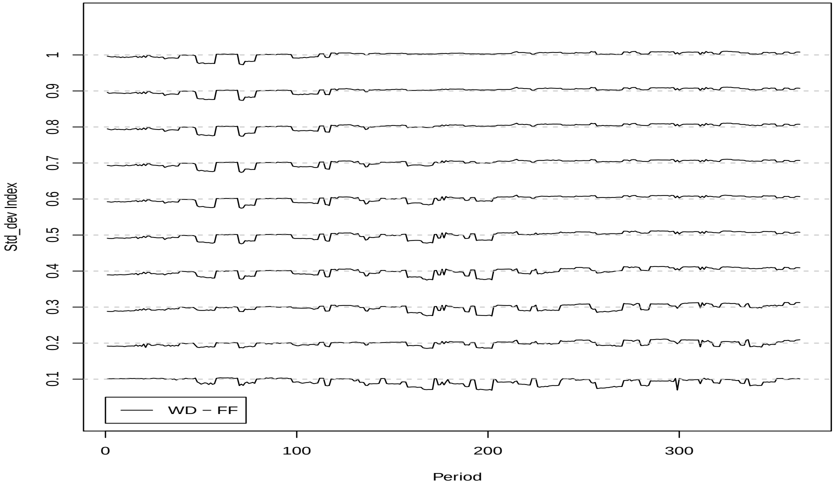
(a) $\sigma_\epsilon = 0.7$



(b) $\sigma_\epsilon = 1$

Results

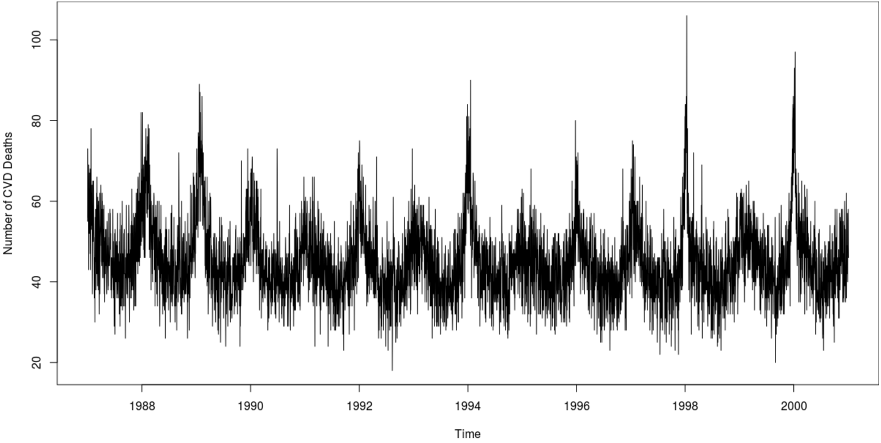
Computing Time



APPLICATION

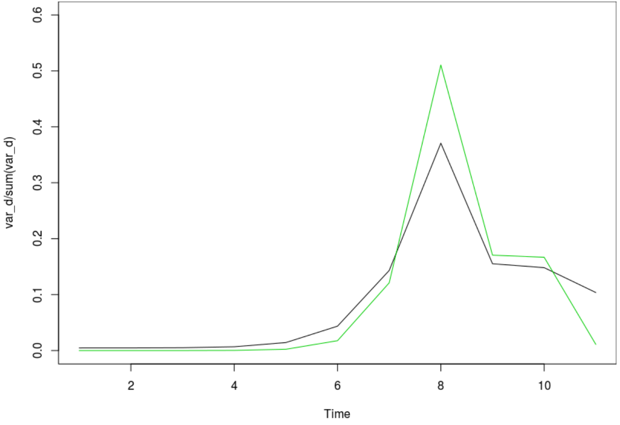
Cardiovascular Deaths

- Daily cardiovascular deaths from 1987-2000 (seasonal R)
- findFrequency function returns 1 = no seasonality
- wavelet method returns 373

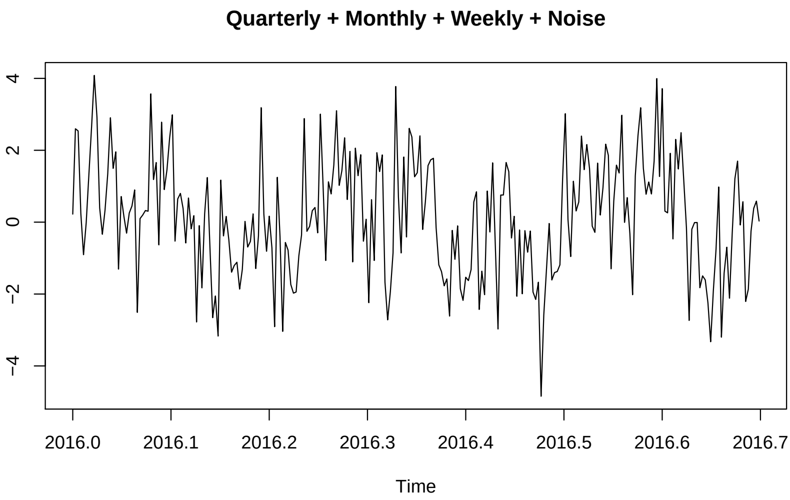


Cardiovascular Deaths

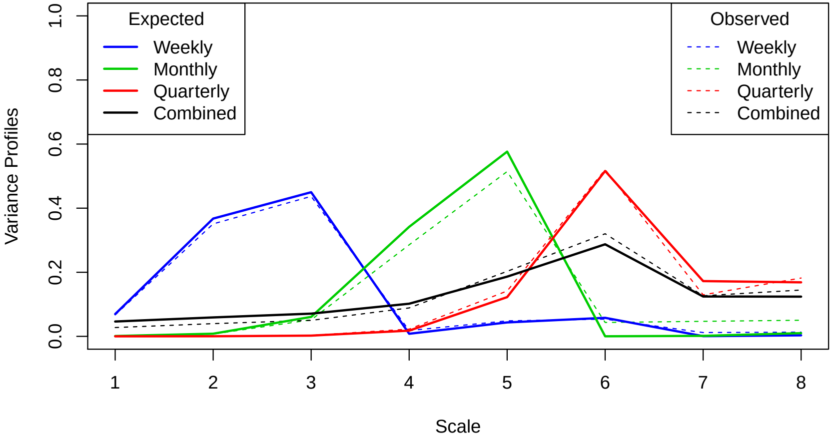
- findFrequency function returns 1 = no seasonality
- wavelet method returns 373



Mixed Frequencies



Mixed Frequencies



Conclusion

- Provided an alternative approach to detecting low frequency cycles.
- Resistant to noise.
- Computationally efficient.
- Seek to apply periodic boundary conditions to improve.
- Need to add confidence interval to estimate.
- Encouraging for further development.

Detecting Periodicity

Approaching an Algorithm

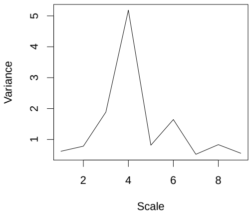
In detecting periodic behaviour we have to find the variance profile that fits correctly. This leads to a number of considerations:

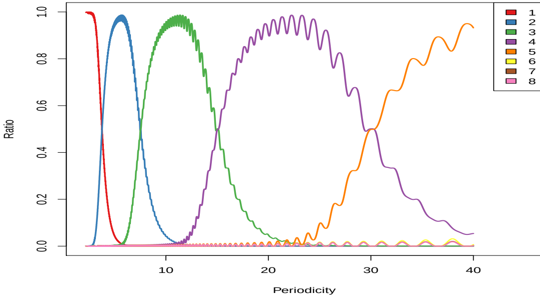
- What region should we search?
- How should we weight the differences?
- How certain are we?

Detecting Periodicity
Search Regions

Mathematics & Statistics | Lancaster University

Suppose we are given the following variance profile for a series, how could we narrow down our search?

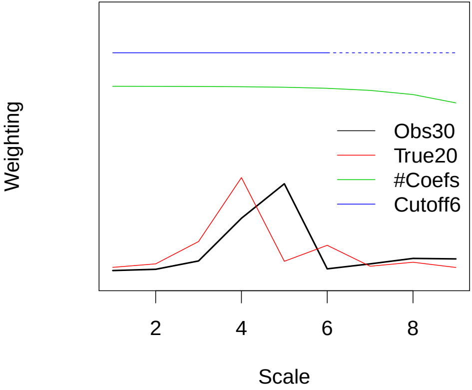




Detecting Periodicity
Weighting Functions

When considering any distance it is useful to consider a variety of weighting functions:

- According to the true variance profile.
- According to the observed profile.
- According to the amount of coefficients considered in each scale.
- Choose a cut-off scale.



Different Wavelets