Reconciling time-series of environmental support data

Robert J. Clement

Exeter University

Swinburne University

Environmental data

- Collected to support agricultural management study in Mississippi
- 76 measurements of ~ 20 different variables
- Redundant measures of most variables
- Captured as time-series (either 2 sec or 5 sec sample rate)

Objective

Reduce the 76 measured time-series to a single set of continuous time-series of required variables

$$\left.\begin{matrix} T_1 \\ T_2 \\ T_3 \end{matrix}\right\} \quad T$$

Problem(s) to address

Measurement errors and noise

Lags

Gaps in data

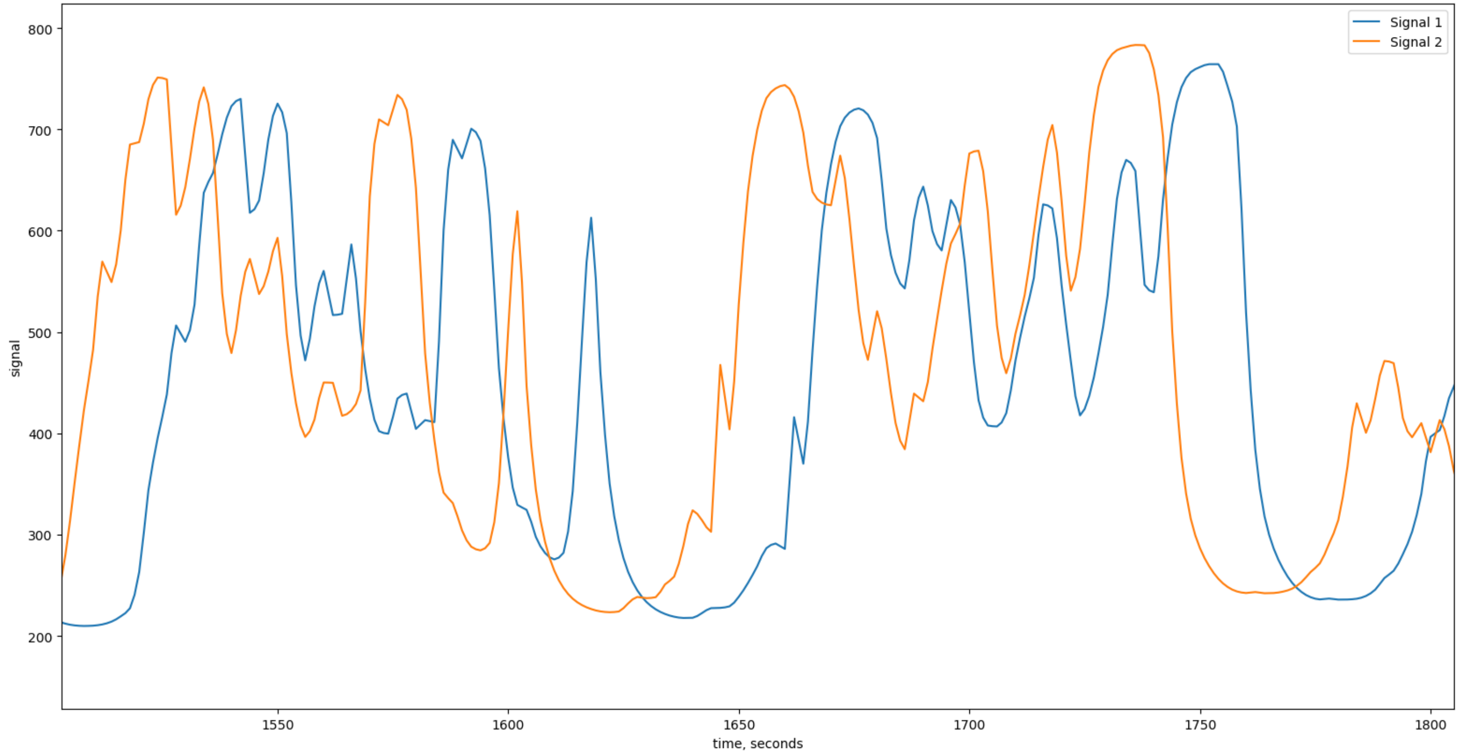
Lags

Variable time lags

Logger timestamp, Spatial separation

Noise & Gaps

Example time series: Shortwave Solar radiation

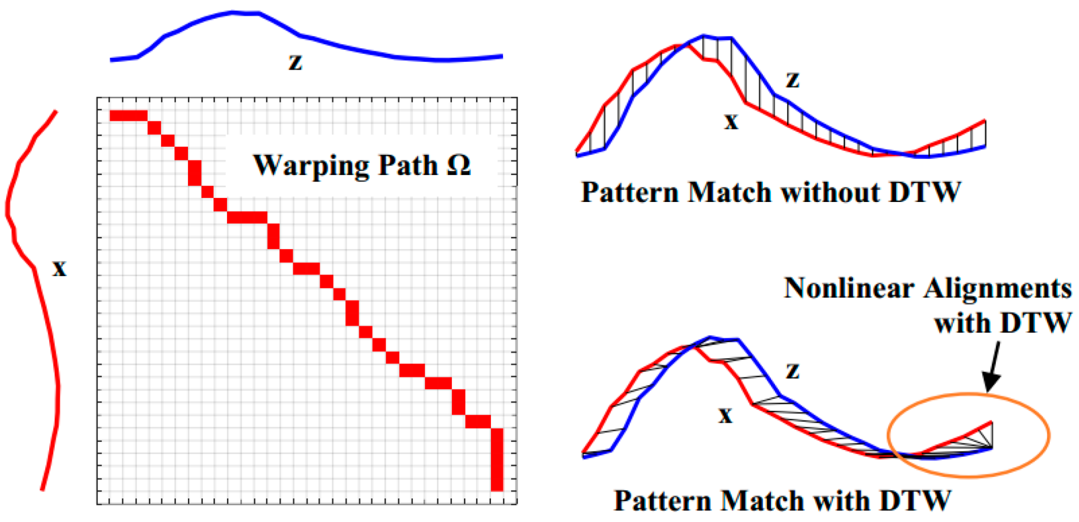


Clean signals
(noise is small)

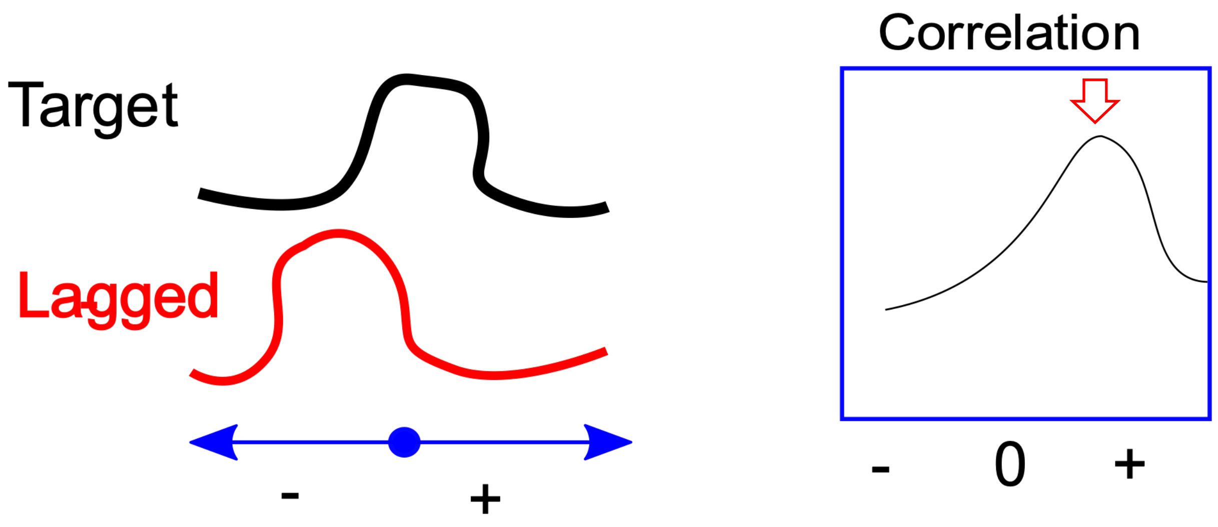
Lag caused by spatial separation is small (i.e. primarily logging system lags)

Lag determination methods

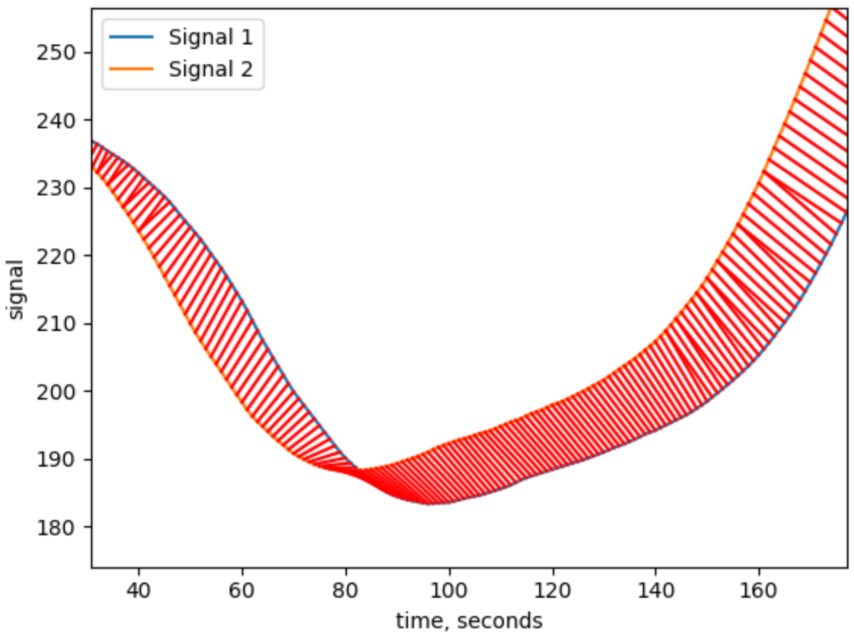
- Dynamic time warping

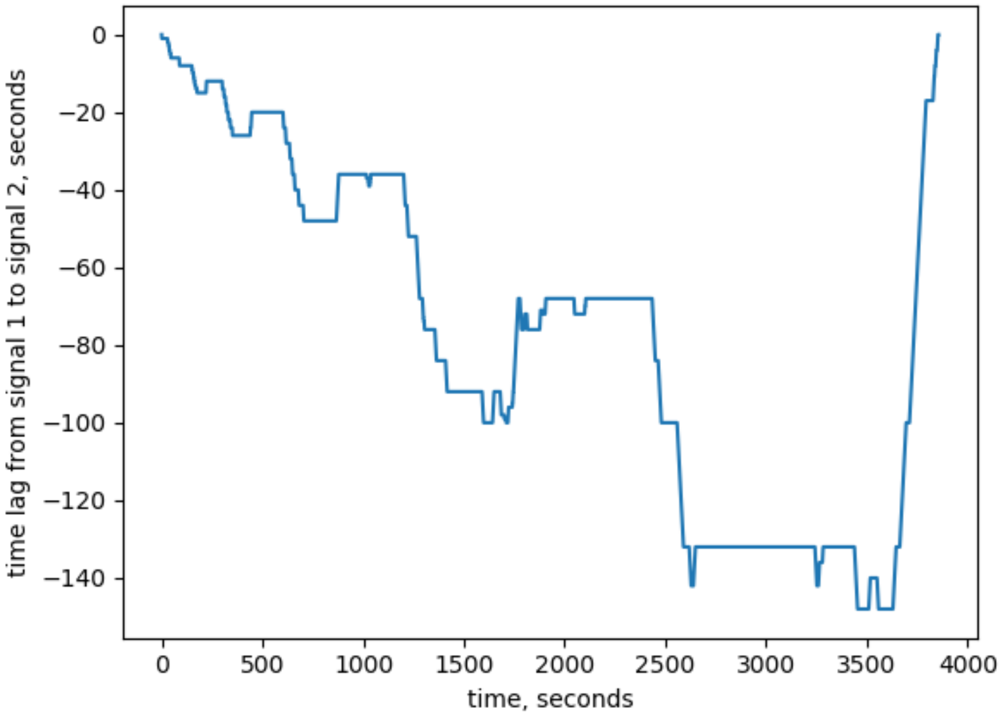


- Lagged cross-correlation



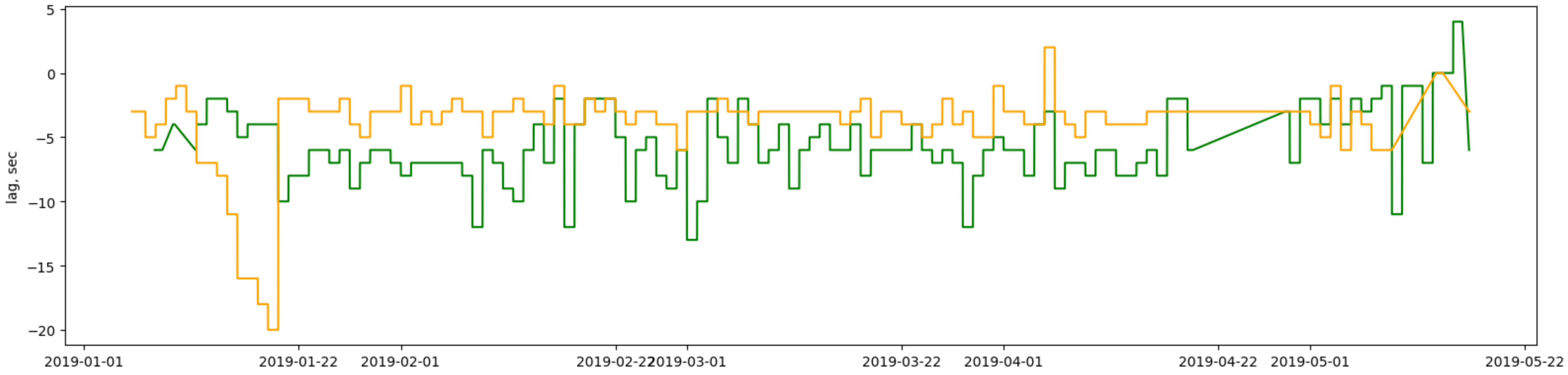
Dynamic time warping: fastdtw





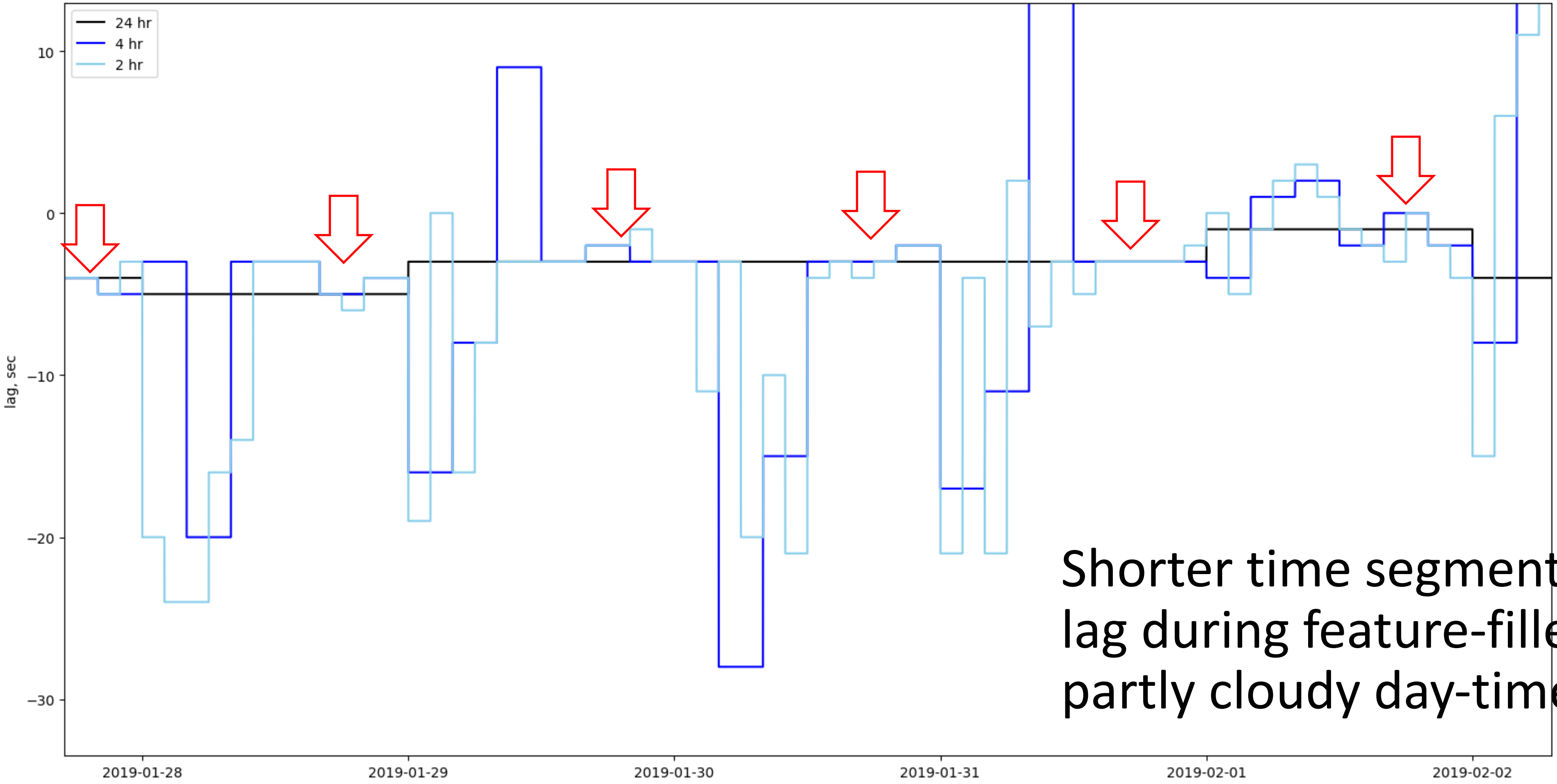
- Computationally very slow
- Unrealistic 1-to-many relationships
- Progressive worsening - incorrect event matching

Lagged cross-correlation



- Computationally slow
- Identifies lags if large sequential time-segments used (24 hours)

Lagged cross-correlation



Shorter time segments only identified lag during feature-filled periods, e.g. partly cloudy day-time conditions

Take aways

- Neither method produces satisfactory lag-time estimates for continuously variable lags.
- Feature-less time-segments are not handled well by either method, but cross-correlation ‘recovers’ better than dynamic-time-warping.
- Cross-correlation is computationally faster than dynamic-time-warping.
- Looking for a better method, ☺