Detecting Changes In Environmental Time Series

Tasmin L. Symons,
MPECDT, Imperial College London & University of Reading

EPSRC
Centre for
Doctoral
Training

Mathematics of Planet Earth

Imperial College London

University of Reading

What is a change point and why do I care?

A *change point* is a discontinuity in a time series, where one (or more) of the statistical properties – mean, variance, auto-covariance, regression coefficients, etc. – undergo a sudden change.

For our purposes, the main aim of detecting such sudden changes is *homogenisation* – the identification of changes unrelated to climate. These can then be taken into account before we begin any analysis.

The Schwarz Information Criterion

One of the most popular methods for detecting changes in meteorological time series (Beaulieu et al, 2012, Killick et al, 2010) is the *Schwarz Information Criterion* (SIC). The idea is simple: First calculate, for $k = 0$ (no change occurs) to $k = N - 1$ (where N is the number of datapoints):

$$\mathrm{SIC}(k) = -2 \log L(\hat{\Theta}(k)) + p \log N,$$

where $\hat{\Theta}(k)$ is the value of the parameters if the change occurs at time k, and p is the number of parameters of the model we are fitting.

If

$$\mathrm{SIC}(0) < \min \mathrm{SIC}(k) \qquad \forall k \in \{2, \ldots, N-1\}$$

then accept the hypothesis that no change has occurred. Otherwise declare a change at time $\hat{k} = \operatorname{argmin} \mathrm{SIC}(k)$.

A small modification

One major issue with using the SIC is its very high Type I (false positive) error rate.

	Empirical Type I error rate	
ϕ	CUSUM	SIC
0	30.05%	37.95%
0.25	32.2%	39.75%
0.5	32.25%	38.8%
0.75	30.28%	38.15%

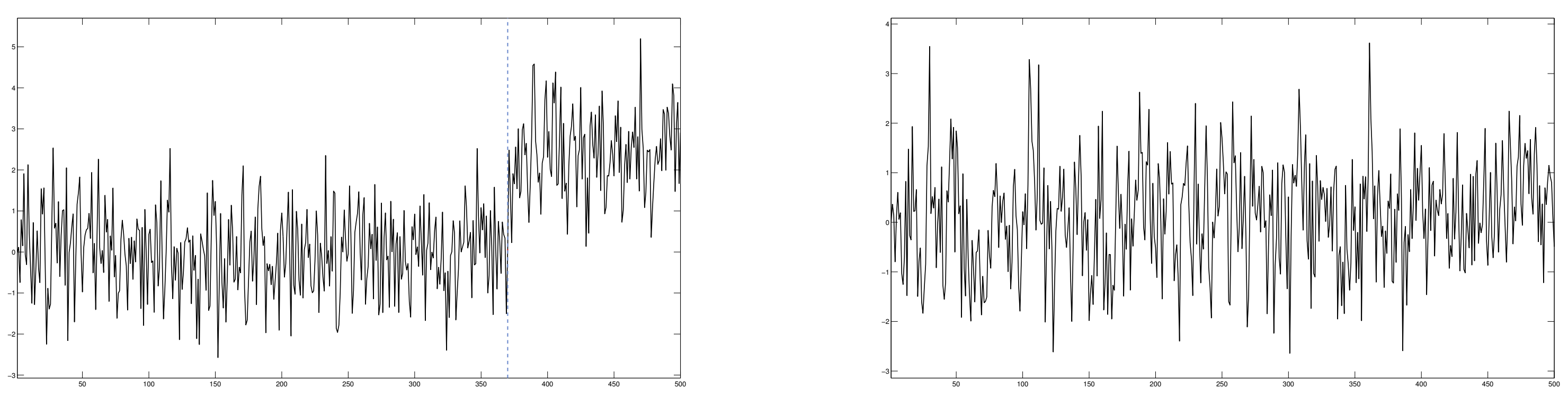
Figure : Type I error rates for the CUSUM and SIC algorithms.

To solve this, we introduce the significance level α, together with a *critical value* c_α such that if

$$\mathrm{SIC}(0) < \min_{1<k<N-1} \mathrm{SIC}(k) + c_\alpha,$$

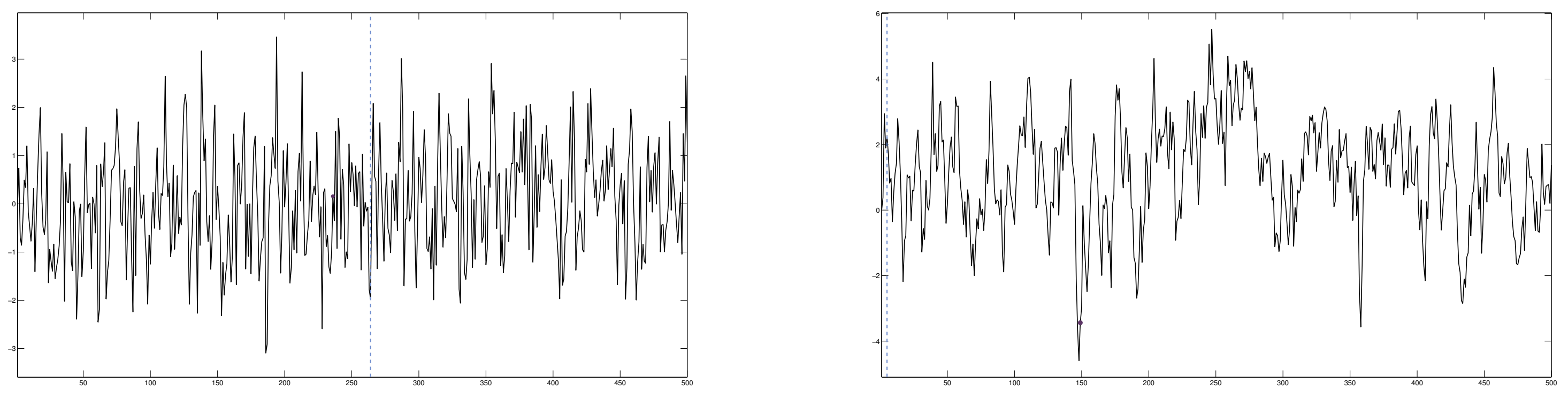
then H_0 is accepted, otherwise $\hat{k} = \arg\min \mathrm{SIC}(k)$. The c_α are determined using classical results about the asymptotic distribution of the series under H_0.

Are you smarter than the SIC?



(a) $\phi = 0.2$. Change $\mu = 0 \mapsto 2$ at time 370. Detected at time 370.

(b) $\phi = 0.5, \mu = 0$ No change. None detected.



(c) $\phi = 0.3$. Change $\mu = 0 \mapsto 0.15$ at time 236. Detected at time 380.

(d) $\phi = 0.9$. Change $\mu = 0 \mapsto 0.25$ at time 149. 'Detected' at time 5.

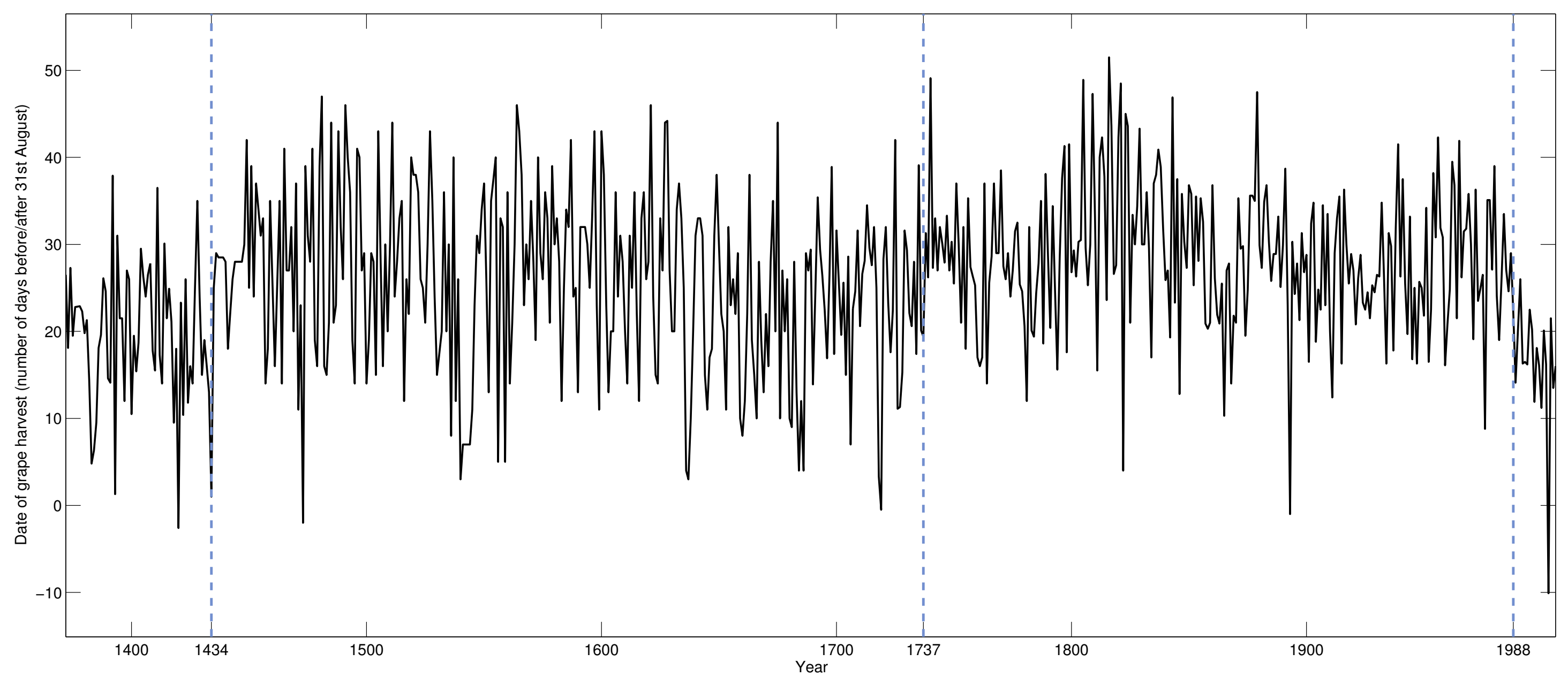
Can you successfully identify the changes in the above time series? How did your answers compare to the algorithms performance?

The French Grape Harvest

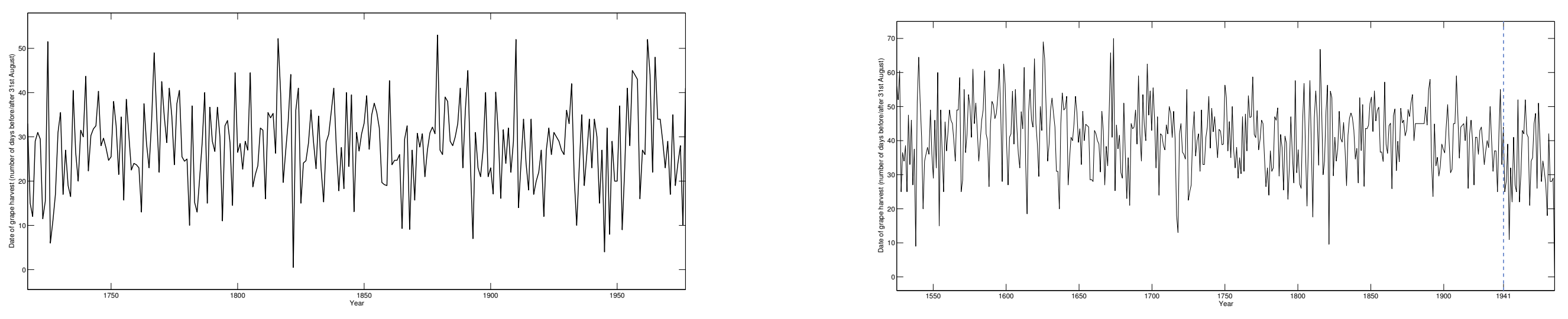
Until the end of the 19th century, French vine growers were not at liberty to harvest their grape crop at a time of their choosing. The *ban des vedanges* ('raise the ban of the harvest') was a municipally-defined date after which the harvest could begin (harvest dates are still agreed in advance today, however vine growers have rather more say).

One side effect of this bureaucracy is a robust and long record of the harvest dates at a variety of Northern Europe's wine-growing regions. The ripening of grapes on the vine is strongly correlated to Northern European summer temperature – the general rule is that the warmer the summer, the earlier the harvest, and vice versa – making the record a valuable proxy dataset.

Burgundy

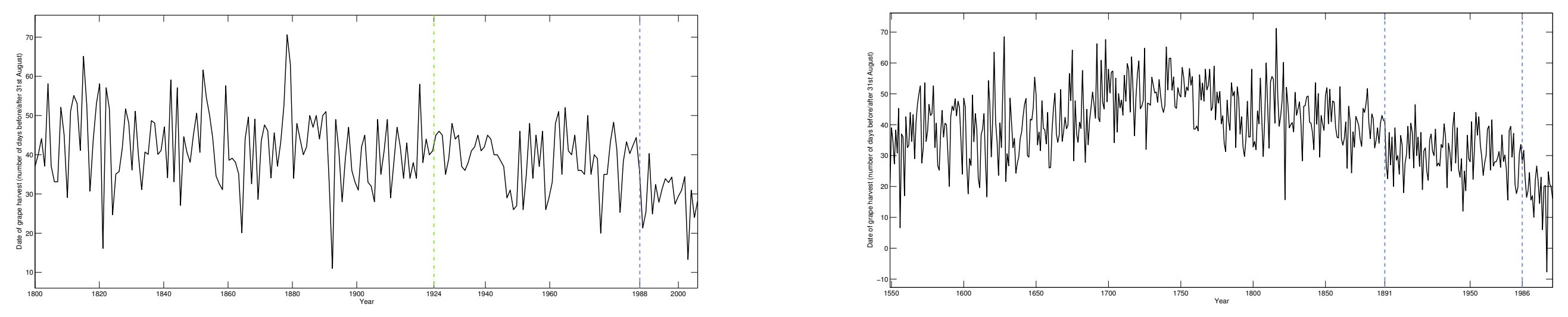


Other regions



(a) Ile de France. No changes detected.

(b) Jura. Change in mean in 1941.



(c) Low Loire Valley. Change in variance in 1924, mean in 1988.

(d) Switzerland. Two changes in mean, in 1891 and 1986.

Figure : Results of testing for abrupt changes in mean (red dashed lines) and variance (magenta dot-dashed lines) in four wine-growing regions of western Europe.

The changes

In all three series which extend to the end of the 20th century a change in mean was detected in the late 1980s. The fact that this is followed by a sharp downward trend in the remaining data suggests this is an effect of climate change – a higher CO2 content in the atmosphere makes photosynthesis more efficient (indeed, both yields and alcohol content have increased in the past quarter century).

The other detected changes are harder to explain, but seem to reflect social upheavals rather than changes in the climate (the 1434 change in Burgundy is a possible exception to this).

References

C. Beaulieu et. al. Change-point analysis as a tool to detect abrupt climate variations. Phils. Trans. Roy. Soc. A., 370 : 1288–1249, 2012.
J. Chen and A. K. Gupta. Parametric statistical change point analysis: with applications to genetics, medicine, and finance. Springer, 2011.
M. Csörgő and L. Horváth. Limit theorems in change-point analysis. Wiley Ser. Probab. Stat. Wiley, 1997. With a foreword by David Kendall.
V. Daux et al. An open-access database of grape harvest dates for climate research: data description and quality assessment. Clim. Past, 8(5):1403–1418, 2012.
D. Jaruskova. Some problems with application of change-point detection methods to environmental data. Environmetrics, 8(5):469–484, 1997.
R. Killick et. al. Detection of changes in variance of oceano- graphic time-series using changepoint analysis. Ocean Engineering, 37(13):1120–1126, 2010.
J. E. Wilson. Terroir: The role of geology, climate and culture in the making of French wines. Univ. California Press, 1998.

`http://www.imperial.ac.uk/people/tasmin.symons11`

`tls111@ic.ac.uk`