Correlation and Coherence Analysis of Paired Time-Series

R G M Crockett

THE UNIVERSITY OF NORTHAMPTON
School of Science and Technology

Data Analysis Group, School of Science and Technology, University of Northampton, NN2 6JD, UK.

Introduction

Changes in radon and other soil-gas concentrations, and other parameters, before and after earthquakes have been widely reported. However, in the majority of such radon cases, changes in magnitude in single time-series have been reported, often large changes recorded using integrating detectors, and the majority of radon time-series analysis is reported for single time-series. With a single time-series, recorded at a single location, there is no measure of the spatial extent of any anomaly and, to a great extent, only anomalies in magnitude can be investigated. With two (or more) time-series from different locations, it is possible to investigate the spatial extent of anomalies and also investigate anomalies in time, i.e. frequency and phase components, as well as in magnitude.

The focus here is correlation and coherence analysis, i.e. windowed cross-correlation in the time and frequency domains, of paired time-series for the identification of simultaneous similar anomalies as probable responses to common stimuli.

Paired Time-Series

The time-series used to develop and illustrate this approach is a radon dataset comprising two hourly-sampled time-series spanning 5.5 months from late June to mid December 2002. This period also included the M_L=5 Dudley (UK) earthquake of 23 September (22 September GMT), which was widely felt by people in Northampton (and elsewhere in the English Midlands), and the Manchester (UK) earthquake swarm of 21-29 October, which wasn't felt in Northampton but was widely felt in southern parts of NW England and northern parts of the English Midlands. Such events are unusual for the UK and, the Dudley earthquake in particular, were the stimulus for the original investigation (Crockett *et al.*, 2006).

The paired time-series are shown in Figure 1

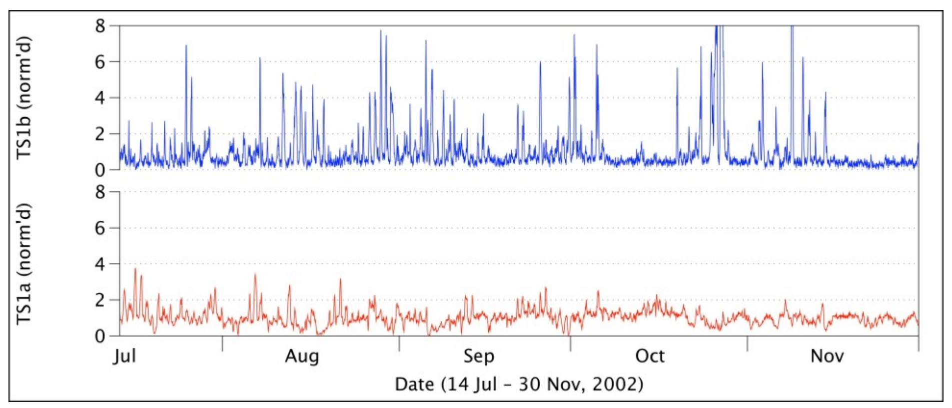

Fig 1. Normalised Paired Time-Series: central 20-week period.

Date / Time (GMT)	Lat.	Lon.	Depth (km)	Mag. (M_L)	Dist. (km)	Location
26/08/2002 23:41	50.048	-0.009	4.0	3.0	247	Eng. Channel
22/09/2002 23:53	52.520	-2.150	9.4	5.0	94	Dudley
23/09/2002 03:32	52.522	-2.136	9.3	3.2	93	Dudley
21/10/2002 07:45	53.475	-2.000	5.0	3.7	161	Manchester
21/10/2002 11:42	53.478	-2.219	5.0	4.3	169	Manchester
22/10/2002 12:28	53.473	-2.146	4.2	3.5	165	Manchester
23/10/2002 01:53	53.477	-2.157	5.0	3.3	166	Manchester
24/10/2002 08:24	53.485	-2.179	3.7	3.8	168	Manchester
29/10/2002 04:42	53.481	-2.198	5.0	3.1	168	Manchester
22/11/2002 01:40	52.921	2.430	10.0	3.4	237	North Sea

Tab 1. Earthquakes (M_L ≥3) within 250km of Northampton.

Cross-Correlation

In outline: starting at the beginning, the paired time-series are windowed and cross-correlated across the window. The window is repeatedly rolled/slid forwards a specifed shift/lag until the end of the time-series is reached, cross-correlating at each shift/lag. This yields a time-series of correlation coefficients.

This can be repeated for different window-durations and, for example, the results presented as a contour plot to reveal the time-duration relationships of any periods of significant cross-correlation between the time-series, analogous to the more familiar spectrogram representation of time-frequency relationships in single time-series, and also to cross-coherence (Crockett *et al.* 2006).

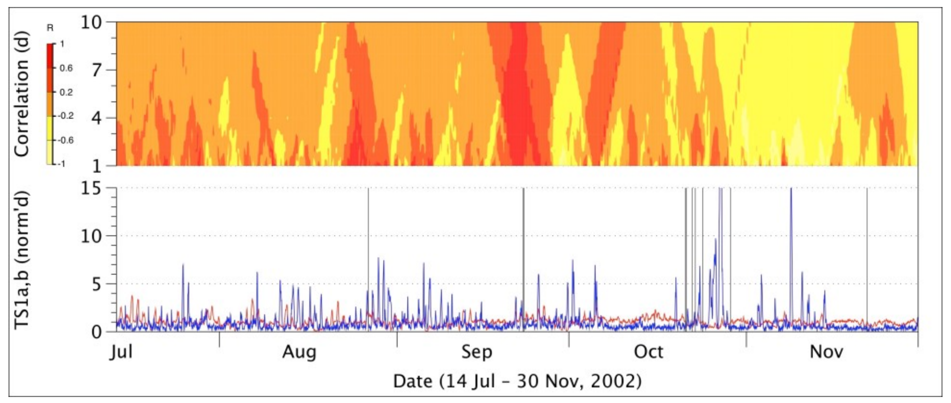

Fig 2. Cross-Correlation (upper plot): contour-plotted with time-series and earthquake incidence (lower plot, TS1a and TS1b in red and blue respectively, earthquake timings as vertical black lines).

Cross-Coherence

Coherence (cross-coherence, magnitude-squared coherence) can be useful in that it measures the similarity of two signals, i.e. time-series in this context, in terms of their frequency composition. It is a normalised measure of power cross-spectral density and is a frequency domain measure of correlation of the two signals (time-series) (Crockett, 2012).

The power spectral density is obtained via the Discrete Fourier Transform and is the proportion of the total power content, i.e. square-of-magnitude, carried at given frequencies. As defined by the Wiener-Khintchine Theorem, the power spectral density is the Fourier transform of the autocovariance and the power cross-spectral density of two time-series is the Fourier transform of their cross-covariance.

For paired time-series $\{x_n\}$ and $\{y_n\}$ the basic expression for unlagged covariance over window length N is:

$$\sigma_{xy} = \sum_{n=1}^{N} (x_n - \mu_x)(y_n - \mu_y)$$

Correlation, R , is normalised covariance, i.e.:

$$R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad -1 \le R_{xy} \le 1$$

Thus, power spectral density, G_{xx}, and power cross-spectral density, G_{xy} , as Fourier transforms of lagged auto- and cross-covariance respectively are:

$$G_{xx}(k) = \sum_{n=0}^{N-1} \gamma_{xx}(n) e^{-i\frac{2\pi}{N}nk} \qquad G_{xy}(k) = \sum_{n=0}^{N-1} \gamma_{xy}(n) e^{-i\frac{2\pi}{N}nk}$$

Thus, the (magnitude-squared) coherence, C_{xy} , and phase-lag, $\tan(\Phi_{xy})$, between two time-series are:

$$C_{xy}(k) = \frac{\left|G_{xy}(k)\right|^2}{G_{xx}(k) G_{yy}(k)} \qquad \tan(\Phi_{xy}(k)) = \frac{\mathrm{Im}(G_{xy}(k))}{\mathrm{Re}(G_{xy}(k))}$$

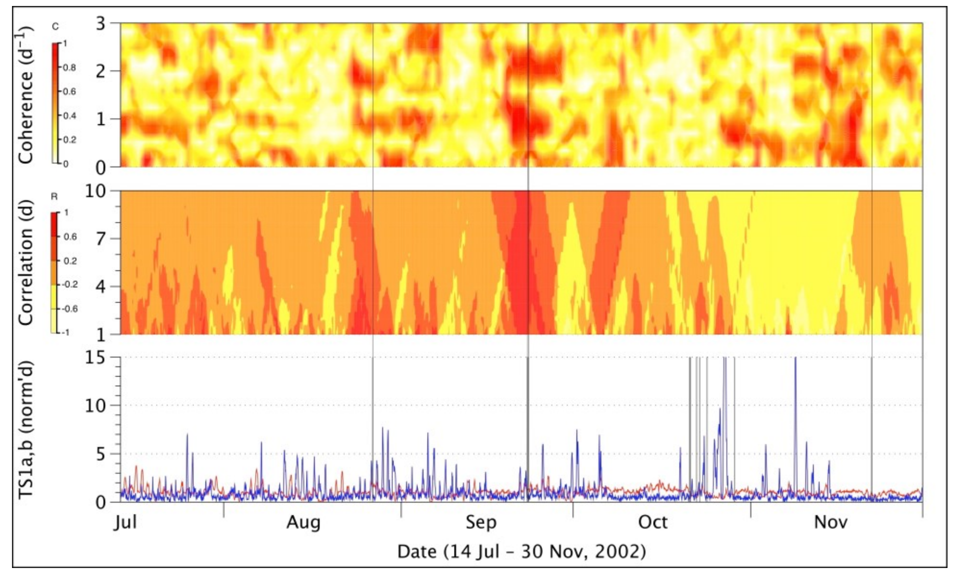

Fig 3. Cross-Coherence (upper plot) and Cross-Correlation (middle plot): contour-plotted with time-series and earthquake incidence (lower plot, TS1a, TS1b and, earthquake timings as in Figure2).

Discussion

Correlation can be misleading, particularly if used in isolation as the sole means of comparison. For example, consider a pair of equal-frequency sinusoids in the simplest case:

i) C = 1 & R = 1 ⇒ in-phase
ii) C = 1 & R = –1 ⇒ half-cycle out-of-phase (anti-phase)
iii) C = 1 & R = 0 ⇒ quarter-cycle out-of-phase

It is intuitively obvious that a pair of equal-frequency sinuosoids (or other waveform) must "correlate" in some sense. However, these three cases show that while a zero, or small, correlation coefficient can indicate a real lack of similarity between two time-series, which is (too) often the default interpretation, it can be easily misinterpreted if there is no information with regard to common frequency content and phase relationship.

The particular time-series considered here are characterised by weak, intermittent, out-of-phase 24-hour cycles and no common meteorological influence. The correlation analysis reveals two anomalous short periods where the time-series correlate: these periods temporally correspond to the Dudley and English Channel earthquakes. The coherence analysis reveals three anomalous short periods where the time-series cohere at 24-hour and 12-hour cycles: two of these periods confirm the periods revealed by the correlation analysis but the third period temporally corresponds to a North Sea earthquake.

The reasons for the absence of correlation/coherence around the Manchester earthquakes are not currently understood.

The correlation/coherence periods also correspond temporally to full-moon earth-tidal maxima, but not exclusively. There is no such correlation/coherence corresponding to one full-moon tidal maximum and none for any of the new-moon tidal maxima, which suggest that earth tides are not the main stimulus. This is shown in Figure 4, with phase-coherence also shown.

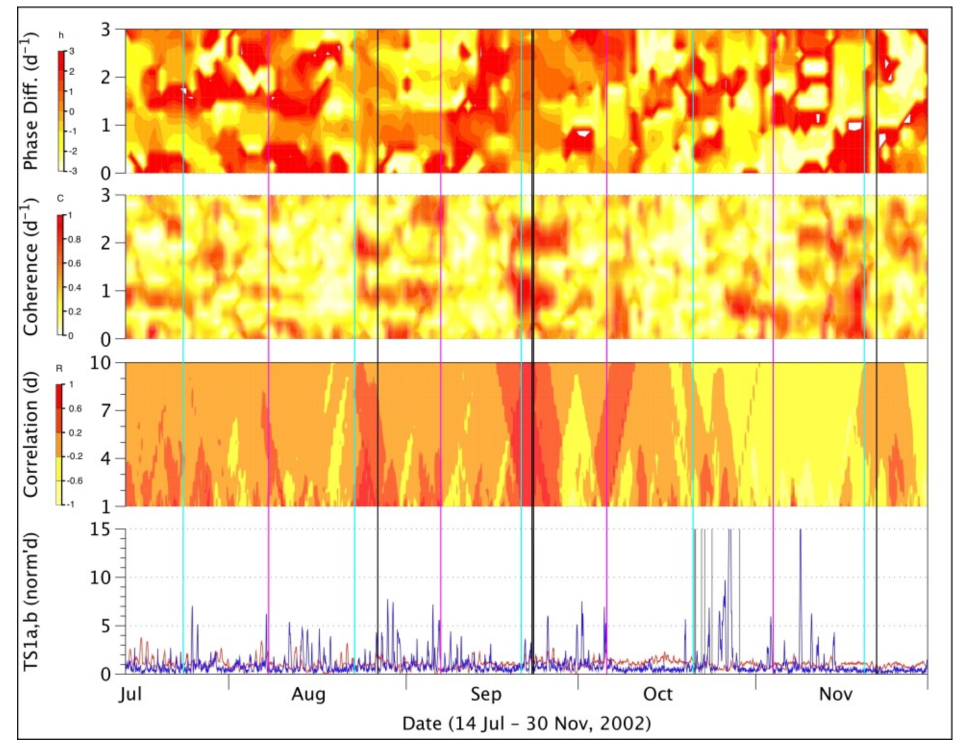

Fig 4. Phase-Coherence (upper plot), Cross-Coherence (second plot) and Cross-Correlation (third plot): contour-plotted with time-series and earthquake incidence (lower plot, TS1a, TS1b and earthquake timings as in Figure2, plus full and new moon timings in cyan and magenta respectively).

Conclusions

Correlation does not imply causality, is not proof of causality: at most, correlation might be evidence to support causality.

Both correlation and coherence show when two, or more, time-series behave similarly in the time domain, according to shape (correlation) or frequency composition (coherence). Thus, these techniques allow the identification of time-domain anomalies, i.e. periods in time when the common behaviour of two, or more, time-series changes from the typical to the anomalous.

References

1. Crockett R G M. 2012. Identification of Simultaneous Similar Anomalies in Paired Time-Series. In Earthquake Research and Analysis / Book 5. Ed. Sebastiano D'Amico, InTech. ISBN 979-953-307-681-1.
2. Crockett R G M, Gillmore G K, Phillips P S, Denman A R, Groves-Kirkby C J. 2006. Radon Anomalies Preceding Earthquakes Which Occurred in the UK, in Summer and Autumn 2002. Science of The Total Environment, 364, 138-148..
3. Crockett R G M, Gillmore G K, Phillips P S, Denman A R, Groves-Kirkby C J. 2006. Tidal Synchronicity of Built-Environment Radon Levels in the UK, Geophys. Res. Lett., doi:10.1029/2005GL024950.