6

Identification of Simultaneous Similar Anomalies in Paired Time-Series

R. G. M. Crockett
University of Northampton
United Kingdom

1. Introduction

Changes in radon and other soil-gas concentrations, and other parameters, before and after earthquakes have been widely reported (Asada, 1982; Chyi *et al.*, 2001; Climent *et al.*, 1999; Crockett *et al.*, 2006a; Crockett & Gillmore, 2010; Igarishi *et al.*, 1995; Kerr, 2009; Koch & Heinicke, 1994; Planinic *et al.*, 2000; Plastino *et al.*, 2002; Wakita, 1996; Walia *et al.*, 2005; Walia *et al.*, 2006; Zmazek *et al.*, 2000). 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 (e.g. Baykut *et al.*, 2010; Bella & Plastino, 1999; Finkelstein *et al.*, 1998). 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 anomalies in magnitude.

The aim of this chapter is to present techniques, developed and adapted from techniques more familiar in the field of signal analysis, for investigating paired time-series for simultaneous similar anomalous features. A paired radon time-series dataset is used to illuminate these techniques. This is not to imply that the techniques are restricted to radon time-series: it is simply that the investigation at the University of Northampton of these techniques in the context of earthquake precursory phenomena has been conducted on radon datasets. This work commenced in the autumn of 2002, following the Dudley earthquake of 23 September which was felt in Northampton and which occurred approximately three months into a radon monitoring programme being conducted as part of another project (Crockett *et al.*, 2006a; Phillips *et al.*, 2004).

1.1 UK earthquakes

The UK is not generally regarded as a seismically active region. In general, across the UK as a whole, in any given year there might be a few earthquakes of magnitude up to 3 or 4 and every 5-10 years there might be an earthquake of magnitude 5 or thereabouts (e.g. Bolt, 2004; Musson, 1996). This is simultaneously an advantage and disadvantage to this research. It is an advantage in that with so few earthquakes there is very little seismic 'noise' in any radon, or other, dataset. It is a disadvantage in that with so few earthquakes, long intervals can

elapse between events and there is an element of luck in obtaining suitably paired time-series to investigate for potential earthquake-related anomalies.
Indeed, in 2002, luck played a major role in stimulating this research. During the latter part of 2002 the University of Northampton Radon Research Group had two hourly-sampling radon detectors deployed for a period which included the Dudley earthquake and also an unusual earthquake swarm in the Manchester area. Subsequently, 5.5 years then elapsed until another UK earthquake of similar magnitude occurred at Market Rasen in February 2008. Again, the University of Northampton had two hourly-sampling radon detectors deployed and, although that paired time-series was shorter in duration, it was still possible identify simultaneous similar anomalies (Crockett and Gillmore, 2010).

1.2 Magnitude anomalies, probability of occurrence

An anomaly in magnitude is, expressed straightforwardly, a magnitude that occurs infrequently, at low probability, often determined according to user-defined probability criteria in a given context. One commonly used criterion assumes that the magnitudes are normally distributed and defines an anomalous magnitude as being one that lies more than a specified number of standard deviations from the arithmetic mean. For example, in any normally-distributed data, an interval of two standard deviations either side of the mean includes 95.45% of the data and any data lying outside this interval can be defined as anomalous, occurring only 4.55% of the time (2.28% at each tail of the distribution). This straightforward type of criterion is clearly satisfactory for normally distributed data but becomes increasingly less robust with divergence of data from normal distributions, indicating the use of more rigorous probability criteria.

1.2.1 Standardised data, Standardised Radon Index (SRI)

Where the data are not normally distributed, or not sufficiently close to normally distributed in a given context, account must be taken of the specific probability distribution. In some cases, it is possible to map onto a normal distribution, via an equiprobability mapping, and then use the mapped-onto distribution to investigate magnitude anomalies. This is essentially the approach taken with, for example, Standardised Precipitation Indices (SPIs) as described by McKee *et al.* (1993) which are a representation of (precipitation) data in terms of standard normal variables, i.e. standard deviations around a zero mean. Alternatively, where the data are lognormally distributed, as is generally the case with radon datasets, or for example, square-root or cube-root normally distributed (Fu *et al.*, 2010), it is possible to define magnitude anomalies in terms of the normal distributions of the logarithms, square-roots or cube-roots of the data respectively, again a representation of the data in terms of standard normal variables. This is the premise underpinning the Standardised Radon Indices (SRIs) proposed by Crockett & Holt (2011). SRIs are determined from the normally-distributed logarithms of lognormally distributed radon data, a representation of the data in terms of standard normal variables. Thus, a given magnitude of SRI is determined by the probability of occurrence of a given magnitude of radon concentration within a dataset.
In addition to transforming radon data, or other data by extension, such that the familiar normal-distribution definitions of magnitude anomaly can be used reliably, SRIs also allow radon time-series to be compared more fully than by considering relative magnitude as obtained by normalising the data, e.g. scaling the data to unit mean value. Such

normalisation does not account for different radon responses to identical stimuli as might arise from differences in emission properties of rocks, soils, groundwater etc. SRIs, in standardising according to probability of occurrence, allow comparison of different radon datasets having different, possibly non-linear, responses to changes in radon emission in response to identical stimuli. In the case of paired time-series, as being considered herein, if values in the time-series have the same probabilities of occurrence they will have the same SRI even if their (relative) magnitudes are different.
Whilst not the focus of this chapter, this technique is discussed briefly, with an example, in Section 3.

1.3 Quality of data, validity of comparison

In any application of any of the correlation and coherence techniques described herein, the durations and sampling intervals of the paired time-series must be equal. Ideally, the two time-series should be sampled at the same times so as to avoid a built-in time-difference between pairs of data. Where there is such a time difference, it might be possible to pre-process the time series, e.g. via a moving average, to minimise its significance, although this must be balanced against the resultant loss of high-frequency content. Under some circumstances where the sampling intervals are different but one is a multiple of the other, it can be possible to aggregate shorter-interval data to correspond to longer-interval data.
In the following sections, it is implicitly assumed that differences between the paired time-series do not arise from different monitoring equipment responses. Where both time-series are of the same parameter or process recorded using the same equipment, this is generally a safe assumption. However, where this is not the case, e.g. the same parameter or process recorded using different equipment for each time-series or different parameters or processes (recorded using different equipment), it will generally be necessary to pre-process the time-series to minimise such differences. In such cases, filtering or spectral decomposition techniques can be used. In particular, Empirical Mode Decomposition (EMD), in decomposing a time-series into separate components (Intrinsic Mode Functions) according to frequency content, has shown promise (Crockett & Gillmore, 2010; Feng, 2011; Huang *et al.*, 1998; Rilling *et al.* 2003).

2. Correlation and coherence

Correlation and coherence are techniques for comparing time-series (more generally, waveforms, signals). In brief, correlation compares shape, i.e. envelope, and is a time-domain technique; coherence compares composition, i.e. frequency (harmonic) content, and is a frequency-domain technique. Neither technique directly compares scale, i.e. neither directly detects magnitude anomalies.
Correlation is a relatively familiar and straightforward technique, widely used in various forms to compare datasets in general. However, correlation can be misleading, particularly if used in isolation as sole means of comparison. For example, consider a pair of identical time-series, such as two equal-frequency sinusoids in the simplest case. If the two sinusoids are exactly in-phase, then their correlation coefficient will be 1 (maximum positive correlation) because both time-series are always changing in the same sense (positively or negatively) at the same time. Conversely, if they are exactly, i.e. a half-cycle, out-of-phase

then their correlation coefficient will be –1 (maximum negative correlation) because one time-series is always changing in the opposite sense to the other. Depending on their phase difference, their correlation coefficient will be between or equal to these two limiting values. However, if they are a quarter-cycle out of phase, their correlation coefficient will be zero. This is because in consecutive quarter-periods the time-series are alternately changing in the same or opposite sense, yielding alternate quarter-periods of positive and negative correlation, equal in magnitude, which sum to zero over a complete period. Thus, whilst a zero, or small, correlation coefficient can indicate a real lack of similarity between two time-series, it can be misinterpreted if there is no information with regard to their frequency content (and phase relationship).

This leads to the less familiar coherence technique. Coherence measures common frequency content, with or without phase information depending on the exact definition. Strictly, the coherence coefficient (i.e. "coherence") measures similarity of frequency content only, irrespective of relative phase, but the overall coherence analysis readily yields a phase-difference for each frequency considered for the coherence coefficient. In the hypothetical case of paired equal-frequency sinusoids considered above, the coherence would be maximal, i.e. coherence coefficient of 1, in all cases irrespective of the phase difference, although the coherence-derived phase-difference would vary exactly as the actual phase-difference. If there is no common frequency content, the coherence is zero.

Correlation is a time-domain technique, coherence is a frequency-domain technique: they are related via their Fourier transforms. Before considering correlation and coherence, therefore, a brief review of Fourier transforms – discrete Fourier transforms – will be given.

2.1 Fourier transform and Discrete Fourier Transform (DFT)

For a continuous function of time, $x(t)$, of infinite duration its continuous Fourier transform is:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t} dt \,, \tag{1}$$

and it is a frequency-domain representation of it (its spectrum). The inverse transform is:

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{2\pi i f t} df \,. \tag{2}$$

In practice, time-series are neither continuous nor infinite, and so the Discrete Fourier Transform (DFT) is used, generally via a Fast Fourier Transform (FFT) algorithm. Subject to constraints arising from the finite and discrete nature of the time-domain function $x(t)$, its DFT, $X(f)$, is a frequency-domain representation of it comprising a spectrum of discrete frequencies on a finite frequency interval. Thus, the forward DFT, $x(t) \to X(f)$, is:

$$X_n = \frac{1}{N} \sum_{k=0}^{N-1} x_k e^{-\left(\frac{2\pi i}{N}\right) nk} \text{, for } x(t) = x_0, x_1, ..., x_{N-1};\ x_k = x(t_k)\ ;\ N \text{ samples} \,, \tag{3}$$

The inverse DFT (IDFT), $X(f) \to x(t)$, is:

$$x_k = \sum_{n=0}^{N-1} X_n e^{\left(\frac{2\pi i}{N}\right)nk} \text{, for } X(f) = X_0, X_1, ..., X_{N-1};\ X_n = X(f_n)\ ;\ N \text{ elements .} \quad (4)$$

For further information on and fuller descriptions of Fourier transforms and their properties see, for example, Riley, 1974; Gabel & Roberts, 1986; Proakis & Manolakis, 2006.

2.2 Correlation analysis

Covariance and correlation are measures of similarity of shape, correlation is normalised covariance and so is independent of scale (e.g. Cramer, 1946; Mood *et al.* 1974). Generalising from the simple case outlined above, the correlation coefficient between two datasets (or samples) will be +1 or –1 if they have the same shape, in-phase or out-of-phase/sign-inverted respectively. The magnitude of the correlation coefficient will reduce according to differences in shape and, as noted above, mismatch in phase.

More formally, in summary, for the comparison of two N-sample time-series, $\{x_i\},\{y_i\}$, e.g. to ascertain if they contain similar features possibly occurring with a lag between them, there is lagged covariance, $\sigma_{xy}(k)$, and lagged cross-correlation, $R_{xy}(k)$, i.e.:

$$\sigma_{xy}(k) = \sum_{n=1}^{N} (x_n - \mu_x)(y_{n-k} - \mu_y) \text{ , for } k \le N \text{ ,} \quad (5)$$

$$R_{xy}(k) = \frac{\sigma_{xy}(k)}{\sigma_x \sigma_y} . \quad (6)$$

The straightforward unlagged covariance and (cross-) correlation of two time-series are (5) and (6) evaluated at zero-lag, i.e.:

$$\sigma_{xy} = \sigma_{xy}(0) = \sum_{n=1}^{N} (x_n - \mu_x)(y_n - \mu_y) \text{ ,} \quad (7)$$

$$R_{xy} = R_{xy}(0) = \frac{\sigma_{xy}(0)}{\sigma_x \sigma_y} . \quad (8)$$

Where, in equations (5) – (8):

x_i, y_i are the i^{th} members of time-series $\{x_i\},\{y_i\}$ respectively

μ_x, μ_y are the mean values of $\{x_i\},\{y_i\}$ respectively

σ_x, σ_y are the standard deviations of $\{x_i\},\{y_i\}$ respectively

For the comparison of a time-series against lagged versions of itself, e.g. to find the period(s) of any cyclic features, there is (lagged) autocovariance, $\sigma_{xx}(k)$, and (lagged) autocorrelation $R_{xx}(k)$, i.e.:

$$\sigma_{xx}(k) = \sum_{n=1}^{N} (x_n - \mu_x)(x_{n-k} - \mu_x) \text{ ,} \quad (9)$$

$$R_{xx}(k) = \frac{\sigma_{xx}(k)}{\sigma_{xx}(0)} = \frac{\sigma_{xx}(k)}{\sigma_{xx}} \text{ ,} \quad (10)$$

which are (5) and (6) with both time-series the same, and clearly both have their maximum values (i.e. autocorrelation is +1) at zero lag.

2.2.1 Rolling/sliding correlation

Whilst lagged cross-correlation reveals any lag between two time-series and cross-correlation of whole time-series gives an overall measure of similarity, neither reveals any time-dependence of similarity during the time-series, i.e. sections where the time-series correlate to greater or lesser extents. One means of achieving this is to window the paired time-series, starting at the beginning, cross-correlate across the window, roll/slide the window forwards a number of samples and repeat the cross-correlation, repeating the roll/slide-and-correlate procedure until the end of the time-series is reached. This yields a time-series of correlation coefficients which reveals those sections of the paired time-series which are varying exactly in phase (R = 1), those which are varying exactly out-of-phase (R = –1) and all intermediate values. 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 to cross-coherence, as described below (e.g. Crockett *et al.*, 2006a).

2.3 Coherence (cross-spectral) analysis

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 (Brockwell & Davis, 2009; Penny, 2009; Proakis & Manolakis, 2006; Venables & Ripley, 2002). It is a normalised measure of power cross-spectral density and is a frequency-domain measure of correlation of the two signals (time-series).

2.3.1 Power spectral density

The concept of "signal power" is not itself useful in this context but the terminology is established and is retained herein for reasons of simplicity. In signals which transmit energy, power, it is significant and is determined by the square of the amplitude.

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 (Proakis & Manolakis, 2006), the power spectral density, G_{xx}, of a signal (time-series) is the Fourier transform of the autocovariance, i.e.:

$$G_{xx}(k) = \sum_{n=0}^{N-1} \sigma_{xx}(n) e^{-i\frac{2\pi}{N}nk} \quad . \tag{11}$$

Also, the power cross-spectral density, G_{xy}, of two signals (time-series) is the Fourier transform of their cross-covariance, i.e.:

$$G_{xy}(k) = \sum_{n=0}^{N-1} \sigma_{xy}(n) e^{-i\frac{2\pi}{N}nk} \quad . \tag{12}$$

2.3.2 Coherence (magnitude-squared coherence)

The coherence (coherence-coefficient), C_{xy}, is the cross-spectral density (complex) normalised by the product of the individual spectral densities (real), i.e.:

$$C_{xy}(k) = \frac{\left|G_{xy}(k)\right|^2}{G_{xx}(k)G_{yy}(k)}, \tag{13}$$

and the phase difference, Φ_{xy}, is given by:

$$\tan(\Phi_{xy}(k)) = \frac{\mathrm{Im}(G_{xy}(k))}{\mathrm{Re}(G_{xy}(k))}. \tag{14}$$

The coherence, C_{xy}, is real and varies between zero, i.e. no common components of the signals (time-series) at frequency k, and 1, i.e. equal proportional components of the signals (time-series) at frequency k.

2.4 Summary

This section has outlined the mathematics of correlation and coherence and also summarised their strengths and weaknesses. Both techniques have their strengths and weaknesses but complement each other. Correlation compares shape (envelope), effectively in-phase periodic features, but gives no information regarding frequency content and so can be misleading with regard to out-of-phase features. Conversely, coherence compares periodic structure (frequency composition and phase difference) and so informs with regard to similarity of composition and structure, accounting for phase difference, but not with regard to overall shape. Thus, used in combination, a much more complete comparison can be obtained than by using either in isolation.

3. Case-study: July – November 2002

The techniques outlined above have been investigated, developed and adapted using primarily one paired time-series dataset: this 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.*, 2006a).

In effect, this is an investigation of three time-series, one of earthquake incidence and two of hourly-sampled radon concentrations, for common radon responses to earthquake stimuli. The hypothesis for this investigation is that a big disturbance, such as an earthquake, occurring at a relatively large distance compared to the detector separation could be expected to produce simultaneous similar radon anomalies (Crockett *et al.*, 2006a). Conversely, any anomalies which are neither simultaneous nor similar are more likely to arise from stimuli local to individual monitoring locations. This hypothesis is appropriate under the circumstances for the radon time-series considered herein, as described below.

However, under different circumstances, e.g. where the detector separation is greater in comparison to the distance from the stimulus, or where one detector location is significantly closer to the stimulus than the other, the analysis might have to modified appropriately to account for a time lag, e.g. moving one time-series relative to the other in the time domain as indicated by lagged cross-correlation. Also, where the individual time-series arise from different monitored substances or processes (e.g. radon and another soil-gas or rock property), pre-processing as indicated in section 1.3 (or otherwise) might be appropriate.

3.1 The time-series

The radon data were collected using Durridge RAD7s, operated 2.25 km apart in Northampton (Phillips *et al.*, 2004; Crockett *et al.*, 2006a,b). A central 20-week extract, 14 July – 30 November 2002, of the 5.5-month paired time-series is shown in Figure 1.

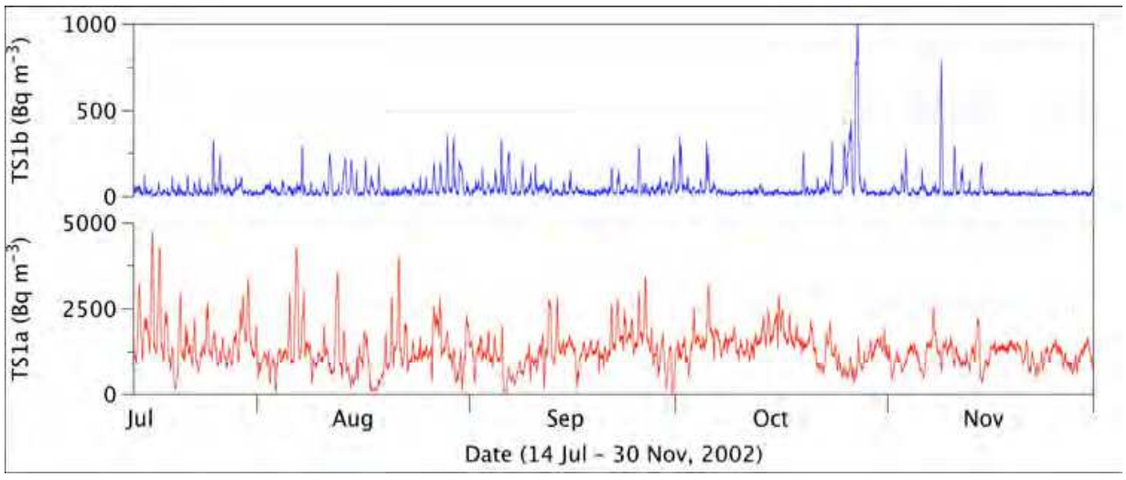


Fig. 1. The paired radon time-series, central 20-week period.

In summary, both time-series are lognormally distributed, with correlation coefficients to lognormal distributions of > 0.91, although TS1a recorded radon levels typically 30-40 times greater than TS1b (Crockett *et al.*, 2006a). Both time-series are characterised by weak, noisy, non-stationary 24 h cycles having short autocorrelation times, *ca.* 1-2 days, as shown in the spectrograms and autocorrelograms in Figures 2 and 3 respectively. There is also weak evidence of 7 day cycles, typical of anthropogenic influences, and in TS1b there is evidence of some longer-period variations, durations *ca.* 15 and 30 days, which are possibly attributable to lunar-tidal influences (Crockett *et al.*, 2006a, 2006b). With the exception of rainfall, behind which TS1a and TS1b lag by 14 and 10 days respectively, there are no observed meteorological dependencies (Crockett *et al.*, 2006a).

The spectrograms are for 6 day windows, stepped at 1 day intervals. This window duration gives the best balance between resolution in the time and frequency domains in light of the durations of the periods of high correlation and coherence discussed below.

The earthquake data for the monitoring period, for earthquakes occurring within 250 km of Northampton, are presented in Table 1 (BGS (http://www.bgs.ac.uk) 2003; USGS/ANSS (http://quake.geo.berkeley.edu/anss), 2011). The next nearest earthquakes during this period occurred at distances greater than 420 km, i.e. there is a distance interval of 170 km between the earthquakes listed and the next nearest earthquakes. As well as the Dudley and Manchester earthquakes, there were other earthquakes of interest, these being an English

Channel earthquake on 26 August and a North Sea earthquake on 22 November (not of identified interest in previous stages of this research).

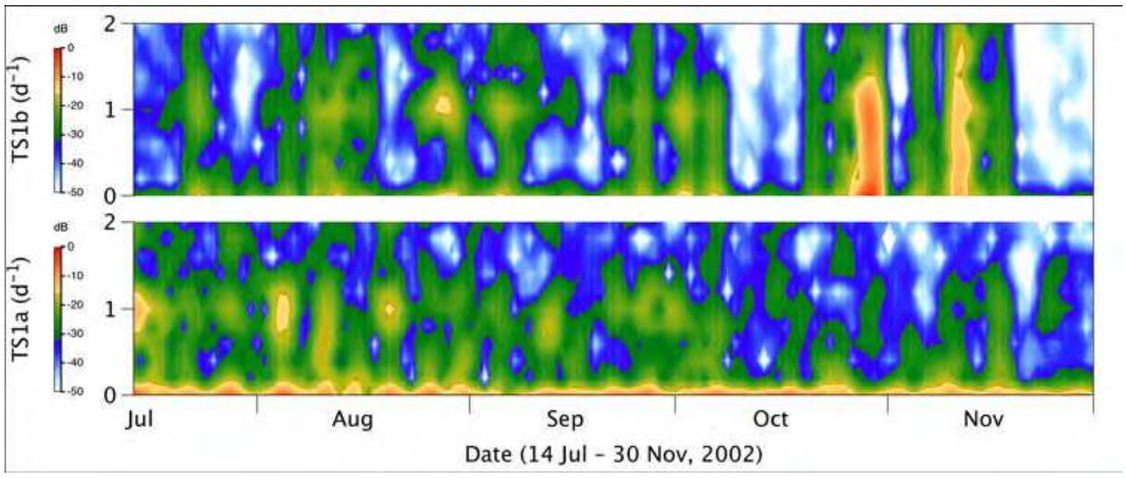


Fig. 2. Spectrograms of the radon time-series, central 20-week period. The amplitude is calibrated in dB, from maximum 0 dB (red) to –50 dB (blue-white). The vertical axes are cycles per day.

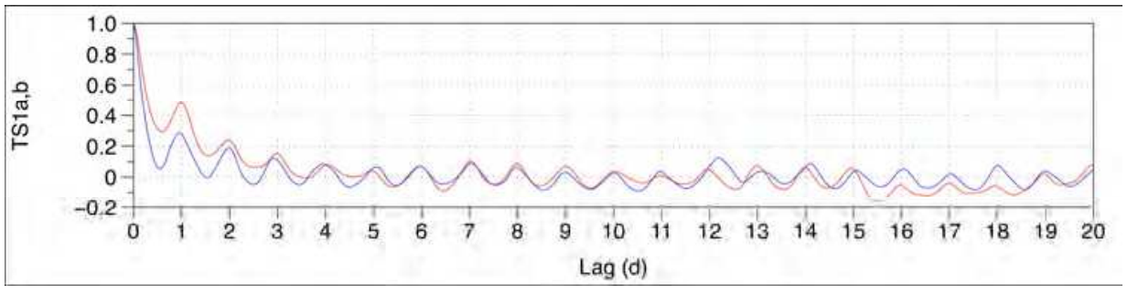


Fig. 3. Autocorrelograms of the radon time-series, central 20-week period (TS1a in red, TS1b in blue).

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

Table 1. Earthquakes ($M_L \geq 3$) within 250km of Northampton, July – December 2002.

3.2 Correlation results

The rolling/sliding windowed correlation, for windows of duration 1-10 days, is shown contour-plotted in the upper plot of Figure 4 (vertical axis is window duration), with the radon time-series and earthquake timings also shown in the lower plot. In this and subsequent figures the time-series are shown normalised to unit mean, to assist visual comparison. This is a simple scaling of amplitudes: in terms of the waveforms of the time-series, the shape and phase are preserved unaltered by this normalisation.

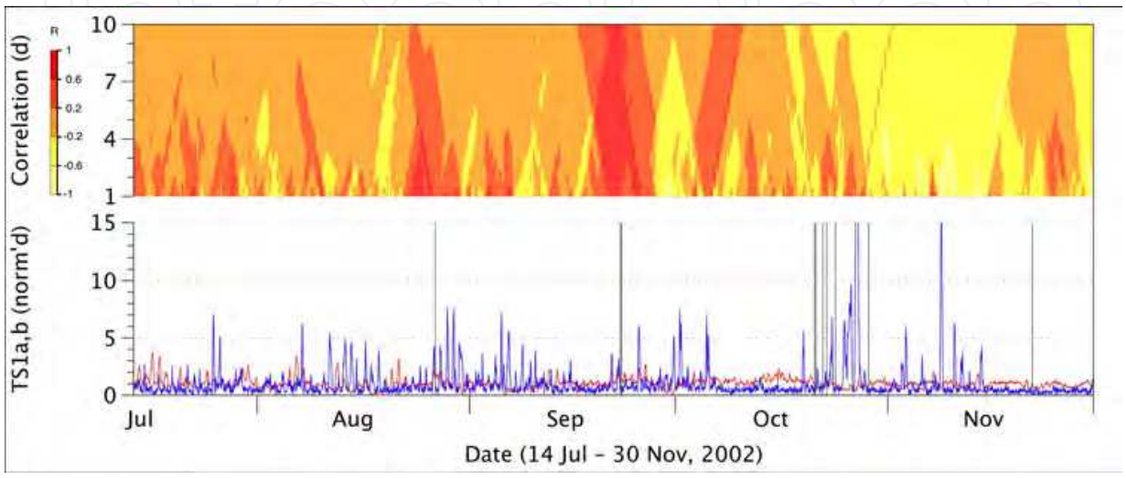


Fig. 4. Rolling/sliding windowed cross-correlation shown contour-plotted (upper plot, strong red and pale yellow areas indicate high positive and negative correlation respectively) with time-series and earthquake incidence (lower plot, TS1a and TS1b in red and blue respectively, earthquake timings as vertical black lines).

Figure 4 (upper) shows two distinct periods of high positive correlation (i.e. red); around (i) 21-23 September, across window durations of up to 10 days, and (ii) 25-27 August, across window durations of up to 5-6 days. The correlation coefficient for the whole period is $R = -0.08$ and the mean values of the correlation coefficient across all ten window durations is $|R| \leq 0.036$. Therefore, the paired time-series typically do not correlate and periods of high correlation are anomalous: e.g. the maximum values of the correlation coefficient for 1-5 day windows occur less than 5% of the time and do so predominantly in these two periods. The more significant of these two periods, 21-23 September, corresponds temporally to the Dudley earthquake of 22 September. The other period, 25-27 August, corresponds temporally to the English Channel earthquake of 26 August. Each period lasts *ca.* 5-7 days and, in each, correlation peaks prior to the earthquakes by *ca.* 1 day. There is no similar period of high positive correlation at the time of the Manchester or North Sea earthquakes.

3.3 Coherence results

The coherence, for windows of duration 6 days stepped at 1 day intervals, is shown in Figure 5. The corresponding phase coherence is shown in Figure 6 (phase difference of TS1b relative to TS1a). In these (upper) figures, the coherence information is contour-plotted to show the variation in coherence (Fig. 5) and phase-difference (Fig. 6) at frequencies of 0-3 cycles per day (vertical axes). The 6 day window duration corresponds to the window

duration used for the spectrograms and also to the durations of the two periods of high positive correlation.

In both coherence and phase-coherence, it is the 24 h cycles, as revealed by the spectrograms and autocorrelograms, which are the focus as these constitute the main frequency component in each time series and, thus, the main frequency at which the time-series could cohere. The figures also show coherence and phase-coherence for 12 h cycles, i.e. twice the frequency of the main cycles, because coherence at this frequency also reveals some information. There is no significant coherence information outside this frequency range.

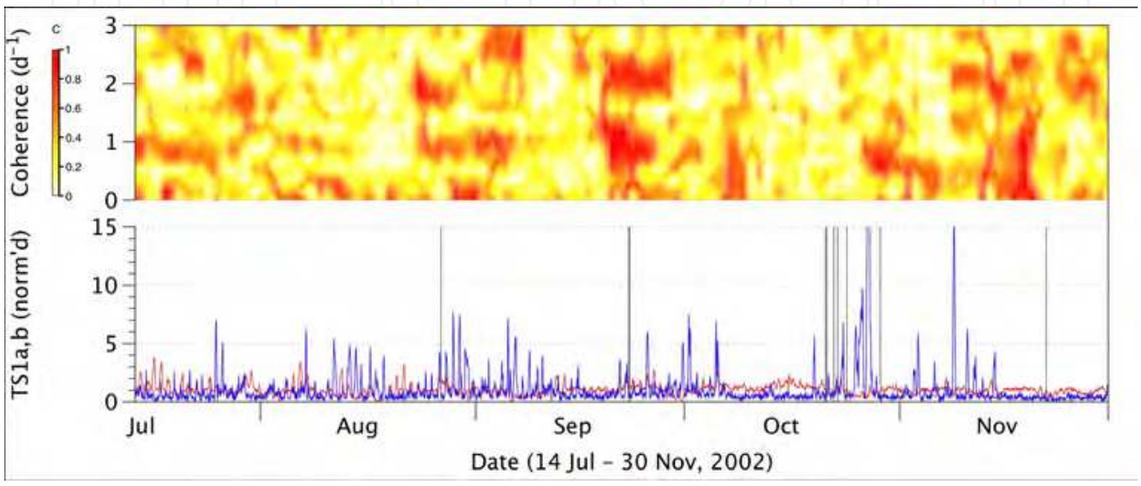


Fig. 5. Coherence shown contour-plotted (upper plot, strong red and pale yellow areas indicate high and low coherence respectively) with time-series and earthquake incidence (lower plot, TS1a and TS1b in red and blue respectively, earthquake timings as vertical black lines).

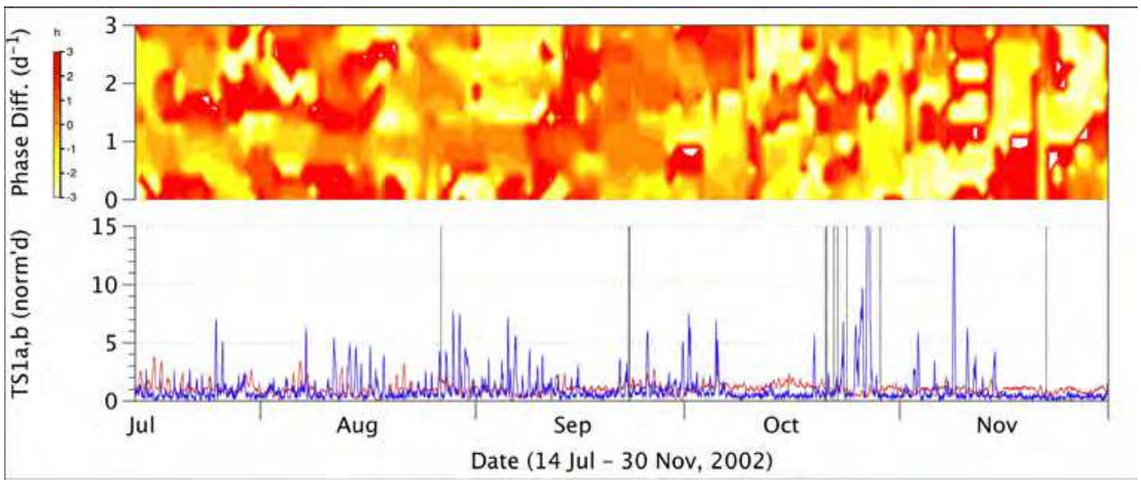


Fig. 6. Phase coherence shown contour-plotted (upper plot, strong red and pale yellow areas indicate positive and negative phase difference respectively, mid-orange indicates small phase difference) with time-series and earthquake incidence (lower plot, TS1a and TS1b in red and blue respectively, earthquake timings as vertical black lines).

Figure 5 (upper) shows two conspicuous periods of high coherence (i.e. red) for both 24 h and 12 h cycles; i.e. late August, late September, and another less conspicuous period in mid-late November. The late August and late September periods correspond to the periods of high positive correlation shown in Figure 4 and correspond temporally to the English Channel and Dudley earthquakes respectively. The mid-late November period occurs a few days before the North Sea earthquake of 22 November. Additionally, there is a period of high coherence for 24 h cycles in late October, which corresponds temporally to the end of the Manchester earthquake swarm, and also a period of high coherence for 24 h cycles in late July with no apparent correspondence to any recorded earthquakes in the region.
Figure 6 (upper) shows two periods where the time-series are in phase (i.e. approximately zero phase difference, medium orange) for both 24 h and 12 h cycles, one in late August and the other in late September. These two periods confirm the two periods of high correlation: if the principal periodic content (24 h) and a principal harmonic (12 h) are in phase then there will be underlying in-phase similarities in the envelopes, giving rise to (high) positive correlation. However, across the time-series as a whole, the daily maxima typically occur at *ca.* 18:00 and 16:00 GMT for TS1a and TS1b respectively, i.e. typically the time-series are not in-phase. This is confirmed directly in Figure 6, which shows variations in phase-difference throughout the period, and also confirmed by the short autocorrelation times shown in Figure 3 and the small average cross-correlation coefficients shown in Figure 4.

3.4 Summary of results

The rolling/sliding windowed correlation, as reported in the initial investigation (Crockett *et al*. 2006a), reveals strong evidence that the two radon time-series correlate positively at the time of the Dudley earthquake, and also evidence that they correlate positively at the time of the English Channel earthquake, but that in general they do not correlate. These results are confirmed by the coherence results, both coherence coefficients and phase-difference.
Seven day details from the time-series are shown in Figure 7, for the period around the Dudley earthquake and Figure 8, for the period around the English Channel earthquake.

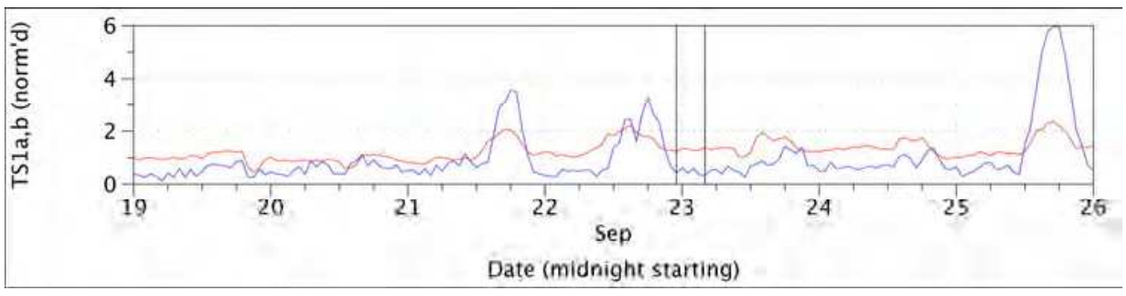


Fig. 7. Late September Anomalies (TS1a in red, TS1b in blue, earthquake timings as vertical black lines).

The two periods are similar in showing the daily maxima, i.e. spikes of *ca.* 6 h duration, occurring simultaneously (i.e. in-phase), and are dissimilar to the remainder of the time-series. Some of these simultaneous maxima occur before the earthquakes – and are potentially earthquake precursory phenomena.
It is clear from these figures, however, that the relative magnitudes of these daily maxima are 2-4 times greater in TS1b than TS1a, indicating that the radon emission characteristics in

the two locations are different. The reasons for this will be differences in the rocks and soils at the two locations and possibly different ventilation characteristics in the two basements where the RAD7s were placed. However, if the time-series are represented in terms of SRIs, effectively standard normal variables, then these different responses are (partially) equalised in terms of probability of occurrence. This is shown in Figure 9 for the September anomalies.

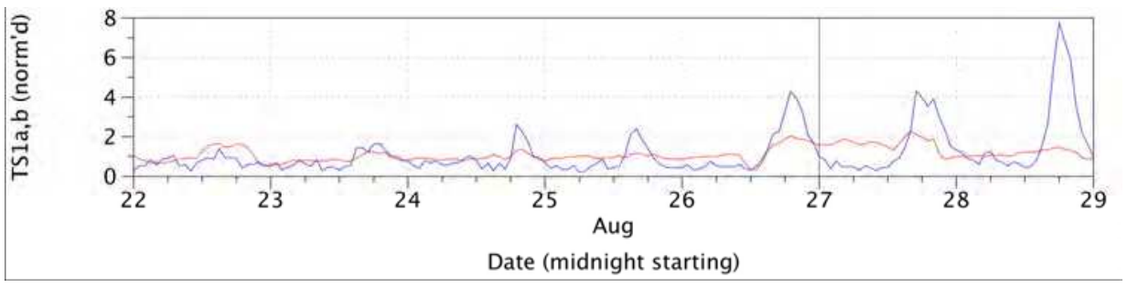


Fig. 8. Late August Anomalies (TS1a in red, TS1b in blue, earthquake timing as a vertical black line).

Figure 9 shows that these simultaneous daily maxima are more similar in probability than in relative magnitude. In SRIs, these maxima have magnitudes *ca.* 1.5-2.5, i.e. 1.5-2.5 standard deviations from the mean. Thus, some of these would 'pass' a plus-or-minus 2 standard deviations criterion for magnitude anomaly whilst others would not, and none would 'pass' a plus-or-minus 3 standard deviations criterion. Therefore, what the correlation and coherence analyses have identified are anomalies in the time domain, i.e. anomalies in frequency composition and phase, rather than in the magnitude domain.

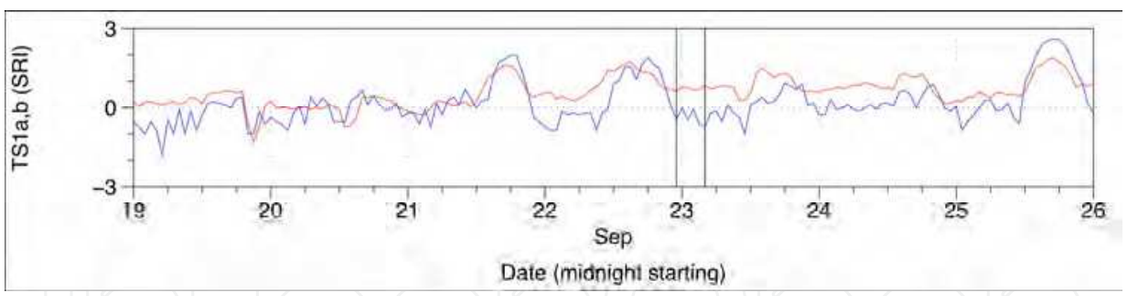


Fig. 9. Late September Anomalies (TS1a in red, TS1b in blue, earthquake timings as vertical black lines).

The late September and late August periods are identified as time-domain anomalies by correlation, coherence and phase-coherence (at 24 h and 12 h cycles), and these periods temporally correspond to the Dudley and English Channel earthquakes. The mid-late November period of coherence (at 24 h and 12 h cycles), but without significant correlation or phase-coherence, is another time-domain anomaly, which has similarities to (coherence) and differences from (correlation, phase-coherence) the late September and late August time-domain anomalies. This mid-late November anomaly occurs a few days prior to the 22 November earthquake but any temporal association with that earthquake is weakened by the absence of corresponding correlation or phase-coherence information. The two periods of 24 h coherence (without 12 h coherence) are clearly ambiguous: the late October period

corresponds temporally to the later Manchester earthquakes but the late July period does not correspond temporally to any recorded earthquake in the region.

Noting the relative magnitudes and proximities of the Manchester earthquakes compared to the English Channel earthquake, it is perhaps surprising the there is no similarly well-defined time-domain anomaly in the radon time-series at the time of Manchester earthquakes. The reasons for this are not currently understood but there are essentially two possibilities. First, any such temporal correspondence between earthquake and time-domain anomalies in the radon time-series is coincidental, as discussed below. Second, the temporal correspondence is not coincidental but the natures of the geologies at and between Northampton and Manchester are such that any earthquake-related radon-stimuli are blurred and attenuated. For information on geology see, for example, Boulton, 1992; Hains & Horton, 1969; Poole *et al.*, 1968; Smith *et al.*, 2000 and Toghill, 2003. However, having also identified time-domain radon anomalies which temporally correspond to the Market Rasen earthquake of 27 February 2008 (Crockett & Gillmore, 2010), the second reason is arguably more probable but more data are required, both earthquake and radon, to investigate this more fully.

In these data, another potential geophysical explanation is possible: i.e. lunar-tidal influences, which have been reported for TS1b in terms of cyclic variations in radon concentration (Crockett *et al.*, 2006b). Tidal influences might account for the periods of 24 h coherence which temporally correspond to tidal maxima associated with the full moons of 24 July, 22 August, 21 September and 20 November, but not the absence of any coherence around the 21 October full moon. This potential explanation is further weakened by the absence of (a) consistent coherence around the new-moon maxima and (b) consistent correlation or phase coherence around both sets of maxima. Also, a tidal-maximum explanation does not account for the anomalies in the February 2008 time-series, which temporally correspond to the Market Rasen earthquake but not a tidal maximum. Lastly, despite the apparent similarities in timing, it is unknown whether there is any lunar-tidal influence on the English Channel, Dudley, Manchester and North Sea earthquakes and no such influence has been reported for UK earthquakes in general.

4. Conclusions

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. In the radon data used to illustrate the techniques, the paired time-series typically neither correlate nor cohere but do so anomalously for short periods. In other data, the emphasis might be different, e.g. the time series might typically both correlate and cohere but contain anomalous periods where they do not or, the time-series might typically cohere at some frequencies but contain anomalous periods where the cohering frequencies change.

Correlation does not imply causality, is not proof of causality: at most, correlation might be evidence to support causality. In the dataset analysed above, despite the clear temporal correspondence of the late September and late August time-domain anomalies to earthquakes, and the temporal correspondence of the mid-late November less well-defined time-domain anomaly to another earthquake, this is all that is shown, i.e. temporal

correspondence. The analysis does not prove that the anomalies are related to the earthquakes, i.e. does not demonstrate that an earthquake-stimulus radon-response relationship exists, but such analysis does provide necessary evidence towards the demonstration that such a relationship might exist.

With regard to magnitude anomalies, care must be taken to apply criteria used for identifying anomalies correctly dependent upon the probability distribution(s) of the data being investigated. Noting the simplicity and familiarity of the normal distribution, and associated *de facto* standard criteria for determining anomalies, a technique such as the SRI which maps data onto standard normal variables is useful, but this technique is also useful in effectively equalising different, generally non-linear radon-emission characteristics and facilitating comparison in terms of probability of occurrence.

5. Acknowledgements

The author gratefully acknowledges members of the University of Northampton Radon Research Group for the collection of the data and also DEFRA (UK) for funding the research under which the 2002 data were collected (EPG 1/4/72, RW 8/1/64). The author also acknowledges UNESCO / IUGS / IGCP Project 571 for facilitating preparation and dissemination of earlier stages of this and related research.

The data-analysis was performed using the open-source R (http://www.r-project.org)) and Scilab (http://www.scilab.org) software packages, with the EMD and Seewave R libraries.

6. References

Asada, T. (1982). *Earthquake Prediction Techniques: Their Application in Japan*, University of Tokyo Press, Japan.

Baykut, S., Akgul, T. & Seyis, C. (2010). Observation and removal of daily quasi-periodic components in soil radon data, *Rad. Meas.*, Vol.45, No.7, pp. 872-879, doi:10.1016/j.radmeas.2010.04.002.

Bella, F. & Plastino, W. (1999). *Radon time series analysis at LNGS, II*, LNGS Annual Report 1999, Gran Sasso National Laboratory, Italy, pp.199-203.

Bolt, B.A. (2004). *Earthquakes* (5th edition). W.H. Freeman & Co, New York, USA. ISBN 0-7167-5618-8.

Boulton, G.S. (1992). Quaternary, In: *Geology of England and Wales*, Duff, P.McL.D. & Smith, A.J. (Eds.),The Geological Society, London, pp.413-444.

Brockwell, P.J. & Davis, R.A (2009). *Time Series: Theory and Methods*, Springer. ISBN 978-1-4419-0319-8.

Chyi, L., Chou, C-Y., Yang, F. & Chen, C-H. (2001). *Continuous Radon Measurements in Faults and Earthquake Precursor Pattern Recognition*, Western Pacific Earth Sciences, Vol.1, No.2, pp.227-246.

Climent, H., Tokonami, S. & Furukawa, M. (1999). Statistical Analysis Applied to Radon and Natural Events, *Proc. Radon in the Living Environment*, Athens, Greece, April 1999, pp.241-253.

Cramer, H. (1974, repub. 1999). *Mathematical Methods of Statistics*, Princeton University Press, ISBN 978-06-910-0547-8.

Crockett, R.G.M., Gillmore, G.K., Phillips, P.S., Denman, A.R. & Groves-Kirkby, C.J. (2006a). Radon Anomalies Preceding Earthquakes Which Occurred in the UK, in Summer and Autumn 2002. *Science of The Total Environment*, Vol.364, pp. No.1-3, pp.138-148. doi:10.1016/j.scitotenv.2005.08.003.

Crockett, R.G.M., Gillmore, G.K., Phillips, P.S., Denman, A.R. and Groves-Kirkby, C.J. (2006b). Tidal Synchronicity of Built Environment radon levels in the UK. *Geophys. Res. Lett.* Vol.33, No.5, L0538, doi:10.1029/2005GL024950.

Crockett, R.G.M. & Gillmore, G.K. (2010). Spectral-decomposition techniques for the identification of radon anomalies temporally associated with earthquakes occurring in the UK in 2002 and 2008. *Nat. Hazards Earth Syst. Sci.*, 10, 1079–1084, doi:10.5194/nhess-10-1079-2010.

Crockett, R.G.M. & Holt, C.P. (2011). Standardised Radon Index (SRI): a normalisation of radon datasets in terms of standard normal variables. *Nat. Hazards Earth Syst. Sci.*, 11, 1839-1844, doi:10.5194/nhess-11-1839-2011.

Feng, Z. (2011). The seismic signatures of the 2009 Shiaolin landslide in Taiwan, *Nat. Hazards Earth Syst. Sci.*, Vol.11, 1559-1569, doi:10.5194/nhess-11-1559-2011.

Finkelstein, M., Brenner, S., Eppelbaum, L. & Ne'eman, E. (1998). Identification of Anomalous Radon Concentrations Due to Geodynamics Processes by Elimination of Rn Variations Caused by Other Factors. *Geophys. J. Int.*, Vol.133, No.5, pp.407-412, doi:10.1046/j.1365-246X.1998.00502.x.

Fu, G., Viney, N.R. & Charles, S.P. (2010). Evaluation of various root transformations of daily precipitation amounts fitted with a normal distribution for Australia. *Theor. Appl. Climatol.* Vol.99, No.1-2, pp.229–238, doi:10.1007/s00704-009-0137-6.

Gabel, R.A. & Roberts, R.A. (1986). *Signals and Linear Systems* (3rd edition), Wiley. ISBN 978-04-718-2513-5

Hains, B.A. & Horton, A. (1969). *British Regional Geology Central England.* (3rd edition), HMSO.

Huang. N.E., Shen. Z., Long. S.R., Wu, M.L., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H. (1998). The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. Roy. Soc. A*, Vol.454, No.1971, pp.903–995, doi:10.1098/rspa.1998.0193.

Igarashi, G., Saeki, S., Takahata, N., Sumikawa, K., Tasaka, S., Sasaki, Y., Takahashi, M. & Sano, Y. (1995). Ground-Water Radon Anomaly Before the Kobe Earthquake in Japan, *Science*, Vol.269, pp.60-61, DOI: 10.1126/science.269.5220.60.

Kerr, R.A. (2009). After the Quake, in Search of the Science – or even good prediction. *Science*, Vol.324, No.5925, p.322, doi:10.1126/science.324.5925.322.

Koch, U. & Heinicke, J. (1994). Radon Behaviour in Mineral Spring Water of Bad Bramburgh (Vogtland, Germany) in the Temporal Vicinity of the 1992 Rörmond Earthquake, the Netherlands, *Geologie en Mijnbouw*, Vol.73, pp.399-406.

McKee, T.B., Doesken, N.J. & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. Preprints, *8th Conference on Applied Climatology*, pp.179–184. January 17–22, Anaheim, California.

Meyer, L.L. (1977). *California Quake*, Sherbourne Press, Nashville, USA.

Mood, A.M., Graybill, F.A. & Boes, D.C. (1974). *Introduction to the Theory of Statistics* (3rd edition), McGraw-Hill. ISBN 978-00-708-5465-9

Musson, R. (1996). British earthquakes and the seismicity of the UK. *Geoscientist*, Vol. No.2, pp. 24-25.

Penny, W.D. (2009). Signal Processing Course, April 2011, Available from http://www.fil.ion.ucl.ac.uk/~wpenny/course/course.html

Planinic, J., Radolic, V. & Culo, D. (2000). Searching for an Earthquake Precursor: temporal variations of radon in soil and water, *Fizika B*, Vol.9, No.2, pp.75-82. ISSN 1330–0008.

Plastino, W., Bella, F., Catalano, P. & Di Giovambattista, R. (2002). Radon groundwater anomalies related to the Umbria-Marche, September 26, 1997, Earthquakes, *Geofisica Internacional*, Vol.41, No.4, pp.369-375. ISSN 0016-7169.

Phillips, P.S., Denman, A.R., Crockett, R.G.M., Groves-Kirkby, C.J. & Gillmore, G.K. (2004). *Comparative analysis of weekly vs. three-monthly radon measurements in dwellings*. DEFRA commissioned research for radioactive substances division. Report DEFRA/RAS/03.006. ISBN 1-900868-44-x.

Poole, E.G., Williams, B.J. & Hains, B.A. (1968). *Geology of the Country around Market Harborough*. Institute of Geological Sciences Memoirs of the Geological Survey of Great Britain, England and Wales, HMSO.

Proakis, J.G. & Manolakis, D.K. (2006). Digital Signal Processing (4th Edition), Prentice Hall. ISBN 978-01-318-7374-2

Riley, K.F. (1974 [and subsequent editions]). *Mathematical Methods for the Physical Sciences: An Informal Treatment for Students of Physics and Engineering*, Cambridge University Press, ISBN 978-05-210-9839-7.

Rilling, G., Flandrin, P. & Goncalves, P. (2003). On Empirical Mode Decomposition and its Algorithms, *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*, Grado (I)

Smith, K.A., Gillmore, G.K. & Sinclair, J.M. (2000). Sediments and Ostracoda from Courteenhall, Northamptonshire, U.K. and their implications for the depositional environment of the Pleistocene Milton Formation, *Proc. Geologists' Association*, Vol.111, No.3, pp.253-263, doi:10.1016/S0016-7878(00)80018-8.

Toghill, P. (2003). *The Geology of Britain: an Introduction*, Airlife, Wiltshire, UK. ISBN 1-84037-404-7.

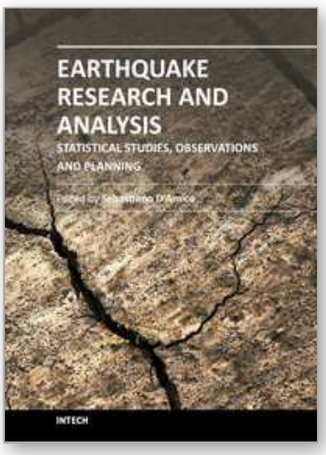
Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S.* (4th Edition), Springer. ISBN 0-387-95457-0 and online complement http://www.stats.ox.ac.uk/pub/MASS4/

Wakita, H. (1996). Geochemical Challenge to Earthquake Prediction, Proc. *USA National Academy Science*, Vol.93, No.9, pp.3781-3786.

Walia, V., Virk, H.S., Yang, T.F., Mahajan, S., Walia, M. & Bajwa, B.S. (2005). Earthquake prediction studies using radon as a precursor in N-W Himalayas, India: A case study. *Terrestrial, Atmospheric and Oceanic Sciences*, Vol.16, No.4, pp.775-804. ISSN 1017-0839.

Walia, V., Virk, H.S. & Bajwa, B.S. (2006). Radon precursory signals for some earthquakes of magnitude > 5 occurred in N-W Himalaya. *Pure and Applied Geophysics*, Vol.163, No.4, pp.711-721. ISSN 0033-4553.

Zmazek, B., Vaupotic, J., Zivcic, M., Premru, U. & Kobal, I. (2000). Radon Measurements for Earthquake Prediction in Slovenia, *Fizika B*, Vol.9, No.3, pp.111-118. ISSN 1330-0016 / 1333-9133.



Earthquake Research and Analysis - Statistical Studies, Observations and Planning

Edited by Dr Sebastiano D'Amico

ISBN 978-953-51-0134-5
Hard cover, 460 pages
Publisher InTech
Published online 02, March, 2012
Published in print edition March, 2012

The study of earthquakes plays a key role in order to minimize human and material losses when they inevitably occur. Chapters in this book will be devoted to various aspects of earthquake research and analysis. The different sections present in the book span from statistical seismology studies, the latest techniques and advances on earthquake precursors and forecasting, as well as, new methods for early detection, data acquisition and interpretation. The topics are tackled from theoretical advances to practical applications.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen