Reprinted from the Lawrence Berkeley Laboratory
Earth Sciences MASTER al Report, 1978(LBL-8648)

LBL-10139

STATISTICAL ANALYSIS OF THE CORRELATION OF EARTHQUAKES WITH RADON CONCENTRATION IN WATER FROM SHALLOW WELLS NEAR OROVILLE, CALIFORNIA

B. Berlin

April 1979

Prepared for the U.S. Department of Energy
under Contract W-7405-ENG-48

This article consists of pp. 193 to 198 as reprinted from the Lawrence Berkeley Laboratory Earth Sciences Division Annual Report, 1978 (LBL-8648).

STATISTICAL ANALYSIS OF THE CORRELATION OF EARTHQUAKES WITH RADON CONCENTRATION IN WATER FROM SHALLOW WELLS NEAR OROVILLE, CALIFORNIA

B. Berlin

INTRODUCTION

This report is an account of work performed from June 20 to September 25, 1978, funded by Lawrence Berkeley Laboratory Director's Development funds. This work is part of an investigation to determine whether a statistical relationship exists between varying radon concentration in well waters and the occurrence of nearby earthquakes. If such a dependence is verified, then certain patterns of changing radon activity may be useful for predicting earthquakes.

Radon concentration in two water wells near Oroville, California, the Prosise and the Gilley wells, were compared with periodic aftershocks of the August 1, 1975, Oroville earthquake for a period of about 600 days. The data are consistent with data being generated from a distribution of: (a) all noise, or; (b) long-term seismic fluctuations correlated with changes in the Prosise well, or short-term seismic fluctuations with changes in the Gilley well, or both. In both cases, the dependence must be expressed with different equations in different directions from the wells. Also the data are inconsistent with data being generated from a distribution where the dependence between seismic activity and radon activity is very strong in all directions.

DATA COLLECTION

The collection of radon data started a few days after the August 1, 1975, Oroville (California) earthquake of magnitude 6 on the Richter scale, and so coincides with the series of aftershocks. Sampling consisted of filling a pair of 500-ml-capacity polyethylene bottles at the wellhead, sealing them immediately against gas loss, transporting them to LBL within a few days of collection time, and making direct measurement of the radon content of the water by low-level gamma-ray spectrometry at the LBL Low Background Counting Facility.

One sample per day was collected from each of six wells in the region of aftershock occurrence, including wells drilled into poorly consolidated sediments and into bedrock formations. The location of these wells is given in Figure 1. Subsequent experience showed that only bedrock wells showed a significant radon variation. Sampling at other wells was then curtailed, and our efforts were concentrated on acquiring detailed data from three bedrock wells. The shallowest of these (the Norman well, 65-ft deep) is believed to have suffered occasional invasion of irrigation water applied to nearby pasture land. Data from this well are therefore of questionable value. Data from the two remaining bedrock wells (the Gilley and Prosise wells, each about 200-ft deep and dedicated to domestic use only) cover the time from August 12, 1975, to April 6, 1977—a stretch of 604 days. For each of these wells, there are fewer than 10 days of missing observations. Values were simulated for the missing days by linear or cubic interpolation from neighboring values. Typical data obtained from the Gilley and Prosise wells are illustrated in Figure 2.

The seismic data include earthquake time (to the second), epicenter coordinates, depth, and Richter magnitude. This information was obtained from lists compiled by the California State Department of Water Resources (SWR), Sacramento, California, and the United States Geological Survey National Center for Earthquake Research (USGS), Menlo Park, California.

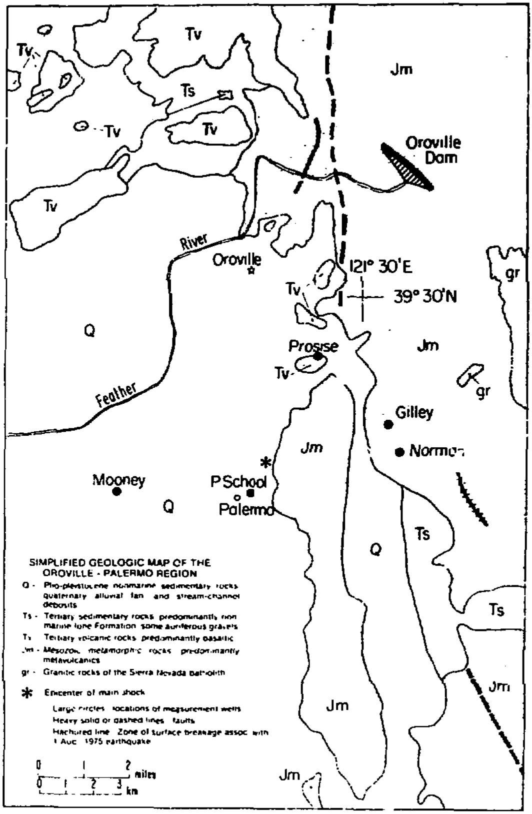
STATISTICAL ANALYSIS

The objective is to measure the extent to which variations in the earthquake process (time, magnitude, distance from wells) are correlated with changes in radon activity. The technique used is described in detail by Brillinger (1975). A more elementary introduction to this topic is given by Kendall (1973).

NOTATION AND MODEL

Time is discretized in days. That is, both radon sampling times and earthquake event times are truncated to the nearest day. The days are numbered: $t = 1, 2, 3, \ldots, N = 600$. The number of days studied, 600, was chosen because computations involved in the analysis are much less time-consuming when N can be factored in many small primes ($600 = 2^3.3.5^2$).

Earthquake data are thought of as the dependent variable where: $Y_1(t) = 1$ if an earthquake occurs on day t, and where $Y_1(t) = 0$, if not. It would have been very desirable to deal simultaneously with the epicenter-well distance, thus: $Y_2(t) =$ 1/distance, if the earthquake occurs on day t,



XBL 775-956

Figure 1. Simplified map of the Oroville, California, area, showing surface geology and the locations of sampled wells in relation to the August 1, 1975, earthquake.

and $Y_2(t) = 0$, if not. However, a problem arises here because $Y_2 = 0$ usually means that there was no earthquake. In the assumed model, $Y_2 = 0$ is confused with having a very distant earthquake; therefore, Y_2 is not used.

The appropriate model for simultaneous analysis of occurrence, location, and magnitude is a marked-point process in which the distribution of the location and magnitude is defined only when an earthquake occurs.

Well data are the independent variable where:

$X_1(t)$ = radon activity in Prosise on day t;

$X_2(t)$ - radon activity in Gilley on day t;

Many subsets of earthquakes have been fitted to models of the following type:

$Y(t) = Y_1(t)$, and $\underline{X}(t) = [X_1(t), X_2(t)]$.

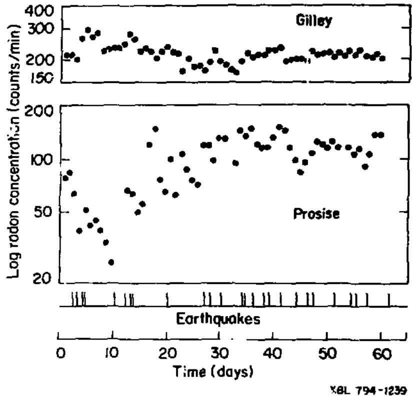


Figure 2. Typical data from the Gilley and Prosise wells for the period October to November 1975, showing daily radon activity and the occurrence times for all earthquakes with magnitudes greater than ~2 on the Richter scale.

These are assumed to be stationary time series. Therefore, we can conclude the following.

The expected values of $Y(t)$ and $\underline{X}(t)$ remain constant in time, so that fluctuations are observed to occur about a fixed mean level. This is, strictly speaking, false. Both the seismic activity and the radon measurements in the two wells show trends. Linear components of all trends have been removed as a first step to treat the data. Although the rate of seismic events decreases, as during the time interval studied, radon activity in the Prosise well shows a positive trend and for the Gilley well it is negative. The change in average radon level is considerable. The Prosise radon level is 40% higher toward the end of the 600-day period compared with the beginning; for the Gilley well the decrease is also 40% over the same time period. The removal of such a trend is a concession that only changes can be analyzed, for which several cycles are observed in the 600 days available.

Any dependence within and between processes relates only to the length of time between the two points considered, rather than to their absolute location on the time scale. Thus, February and April of 1976 are assumed to be as strongly related as September and November of 1977 (two months apart in both cases).

These assumptions are expressed in terms of the covariances:

$$\mathrm{Cov}\{X_i(t),X_j(t+u)\} = c_{X_i,X_j}(u) \quad \text{for all } t,\ i,\ j = 1,\ 2$$

$$\mathrm{Cov}\{Y(t),Y(t+u)\} = C_{YY}(u) \quad \text{for all } t,\ i,\ j = 1,\ 2 \ .$$

$$\mathrm{Cov}\{Y(t),X_j(t+u)\} = c_{X_jY}(u) \quad \text{for all } t,\ i,\ j = 1,\ 2 \ .$$

The model commonly used to relate Y and $\underline{X}$ is linear:

$$Y(t) = \mu + \sum_{u=-\infty}^{+\infty} \underline{a}(u)\ \underline{X}(t-u) + \varepsilon(t) \quad , \quad (1)$$

meaning that the probability of an earthquake on day t, given a particular pattern of $\underline{X}$ around t, is:

$$\mu + \sum_{u=-\infty}^{+\infty} \underline{a}(u)\ \underline{X}(t-u)$$

Here, μ is the long-term probability of an earthquake, and $\underline{a}(u)$ is a function of the time lag u. For each $u = 0, \pm 1, \pm 2, \ldots$, $a(u)$ is a vector with two elements of unknown constants and $\varepsilon(t)$ is the error series, which also is stationary. If the variations in Y are well accounted for by equation (1) for some values of μ and $\underline{a}(u)$, then the variation in $\varepsilon(t)$ will be considerably less than in $Y(t)$. Note that when u is allowed to assume values smaller than zero, *the future of X (beyond t) is involved*, so a good fit for equation (1) does not necessarily imply that a prediction based on past $\underline{X}$ only will be successful.

Equation (1) is reminiscent of multiple linear regression, but here observations are correlated even when they are made at different times. This complicates the analysis quite a bit; statisticians prefer to work with Fourier transforms of the series and their covariance functions because it is much easier to derive criteria to check if any patterns in the data are statistically significant, that is, if they are unlikely to have arisen only from random noise. In the covariance functions given above, dependence is described for observations u days apart. The Fourier transforms of these functions tell the same message, but the argument is a frequency λ, rather than the time lag u. In Fourier analysis, the time series is decomposed in a linear combination of many trigonometric functions (sines and cosines) of varying amplitudes. The different components have frequencies varying between 0 and π, where the highest frequency corresponds to $u = 1$, or cycles of 1 day. Here u and λ correspond to the same wave and are inversely proportional.

DISCUSSION OF RESULTS

Spectra for radon data alone reveal two things. First, the long-term variations (periods greater than 15 days) have much larger amplitude than components with periods of a few days. This phenomenon appears despite the fact that a linear trend has been removed. Second, the correlation between the two wells is negligible. This suggests that any effect that increased stress has on radon activity is very local. The wells are 3.7 km apart.

It is more difficult to find a consistent pattern in the earthquake spectra. Whereas each attempt to fit the data involves the complete set of well observations, the set of selected earthquakes changes. This is done by including all earthquakes within some distance of the well considered.

Also, only earthquakes with a magnitude greater than 1.5 are included. For some selections, there is a tendency for the events to occur in cycles; in other cases the estimated dependence is weak, more like a Poisson process. This is a matter to which little attention has been given during this search for a relation between earthquakes and radon data. It is worth further investigation.

Finally, we considered the dependence between seismic events and radon activity. The extent to which the data fit the model (equation 1) is measured by the coherence $R^2_{YX}(\lambda)$, a generalization of the usual correlation coefficient used in simple linear regression. For a particular λ, it measures the dependence of the trigonometric wave in earthquake occurrence (frequency) on the linear expression of radon data, as in equation (1).

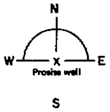
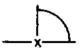
Looking first at the vicinity of the Prosise well, all earthquakes were picked within radii of 2, 3, and 4 km. In each case, the estimated coherence was low enough to be consistent with no dependence at all. Thereafter, semicircles and quadrants were tried with 3-km radii, or a total of eight different tests. In two cases the coherence assumed significant values. These are presented in Figure 3.

Of course, the criteria for a significance are calculated for one particular run. Where many statistical tests are tried, the significance statements lose their power for drawing conclusions. Rather, attention is drawn here to the subsets of earthquakes with the highest correlation, without stating that it will hold up in the long run.

Likewise for the Gilley well, the earthquakes within circles show faint dependence with the radon data. The same is true for semicircles.

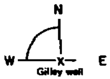
Quadrants look better, as shown in Figure 4. Note that $R^2_{xy}(\lambda) = 0.16$ is the 5% rejection limit. Why do all four quadrants show some dependence, but none of the semicircles? The reason is that the best fitting constants [$\underline{a}(u)$ in equation (1)] are quite different for the four quadrants, and compromise values necessary for semicircles fail to explain the variation in seismic activity.

Figure 5 is a plot of $\underline{a}(u)$ vs. u for the four quadrant zones around the Gilley well that showed some promise with respect to coherence. The horizontal u-axis has been reversed, putting negative u's to the right. The advantage is that the positive u's (corresponding to preceding days) will be to the left of the origin. The plots of $\underline{a}(u)$ all show strong oscillations from high to low values. What does this mean? A first idea would be to reconstruct what pattern of radon changes makes the right-hand

Geographical region for earthquake selection	max $R^2_{XY}(\lambda)$	Approximate period with largest coherence; days
	0.18	100
	0.17	20

XBL 794-1240

Figure 3. Geographical selections of seismic events within a 3-km radius of the Prosise well showing the correlation between seismicity and radon activity.

Geographical region for earthquake selection	max $R^2_{XY}(\lambda)$	Approximate period with largest coherence; days
	0.23	1 1/4
	0.27	1 1/4
	0.33	1 2/3 to 2
	0.27	1 1/2

XBL 794-1241

Figure 4. Geographical selections of seismic events within a 3-km radius of the Gilley well showing the correlation between seismicity and radon activity.

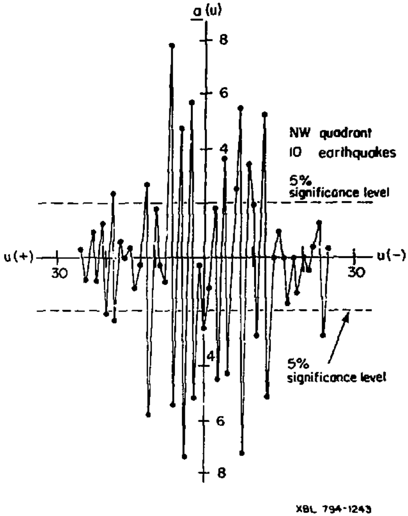


Figure 5. Profile of values for the constant $\underline{a}(u)$ in equation (1), computed for the 3-km-radius quadrant northwest of the Gilley well. Previous days are plotted to the left of the origin; future days, to the right. The ordinate scale is in terms of the estimated standard deviation on $\underline{a}(u)$ values.

side of equation (1) large for the given estimated $\underline{a}(u)$. That is, what makes the chance for an earthquake large? Also, what pattern makes the chance small?

Obviously, a day with a large positive $\underline{a}(u)$ calls for a positive radon count to contribute to a large probability. If $\underline{a}(u)$ has a large negative value, the radon count should be unusually small. Therefore, a strongly oscillating radon pattern would serve as a precursor. Further, an oscillating pattern out of phase with the former type would be an "antiprecursor" and make the chance small.

However, this idea is a dead end because the radon record simply does not have any such rapid oscillations. As mentioned earlier, the variation in radon activity is dominated by slow oscillations.

A more believable interpretation of the *rapid fluctuations of $\underline{a}(u)$ is that the predic*tive effect of radon changes is very short term, about 1 to 2 days.

From the point of view of earthquake prediction, the long-term variations with large amplitude are noise that must be filtered away by the coefficients $\underline{a}(u)$. This is actually achieved with high and low $\underline{a}(u)$ following each other.

An object for further investigation is to study the quickly varying radon signal that remains after filtering, which actually serves as a precursor.

The combined predictive ability for the two wells was also tried by selecting earthquakes within ellipses that had the wells as foci, as illustrated in Figure 6(a).

As expected from the weak dependence of the two radon series, it turned out that for these subsets, the Prosise well had negligible dependence with the earthquakes, while the largest coherence observed so far occurred between the

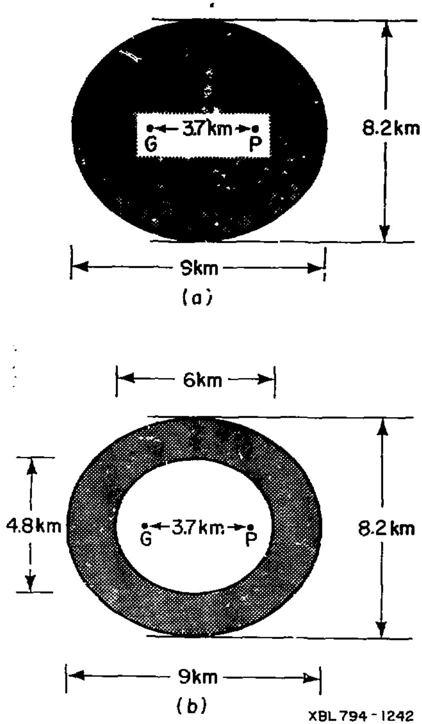


Figure 6. Configurations to test the combined predictive ability for Prosise and Gilley wells (shaded areas indicate regions in which earthquakes occurred).

Gilley well and the earthquakes. Ellipses of different sizes were tried and the fit improves further if an inner ellipse is excluded, as shown in Figure 6(b).

Further work should be done to identify a more reasonable looking zone where the Gilley well is sensitive. The substantial coherence appeared at a frequency corresponding to a period of 2-1/2 days.

RECOMMENDATIONS

To put the above inconclusive findings to a test, we recommend the following.

More data are needed, particularly from the Gilley well, which appears most promising. In the present data, the radon samples are not exactly 24 hours apart; they were taken at different times in the afternoon. To check the one- to four-day fluctuations, it would be useful to have radon activity recorded at exact 12-hr intervals.

With the statistical methods used here (continuous time series) there is no satisfactory way to take epicenter-to-well distance and magnitude into account simultaneously, as pointed out earlier with regard to the variable Y_2. A statistical method relating a marked-point process to a continuous time series should be developed.

It is useful to test the radon data against a model in which two types of earthquake precursors are postulated:

1. Signals whose appearance in prior-time are related to the magnitude of an impending earthquake--signals that may encompass the time domain of days to years
2. Signals whose appearance in prior-time are unrelated to the magnitude of an impending earthquake--signals that may encompass the time domain of hours to a few days (foreshocks)

The statistical method employed here should be well suited for identifying precursors of the second kind--those that occur at a (relatively) fixed time before an earthquake. However, in its present form, the method is not well suited for identifying precursors of the first kind--those that may occur over a few days to a few 10s of days for the aftershock magnitudes encountered during the Oroville study. Hence, adapting the present method or adopting some other approach, is important in order to include time/magnitude/distance parameters for individual earthquakes in the analysis procedure.

REFERENCES CITED

Brillinger, D. R., 1975. Time series, data analysis, and theory. New York, Holt, Rinehart and Winston

Kendall, M. G., 1973. Time series. New York, Griffin.

Work supported by Department of Energy.