International Journal of Engineering Technology, Management and Applied Sciences

www.ijetmas.com March 2017, Volume 5 Issue 3, ISSN 2349-4476

A Study On Statistical Correlation Between Indoor Radon And Meteorological Parameters

Y. Sharma[1], D. Maibam[1], S. Ghosh[2], and A. Saxena[1]
[1]Department of Physics, North- Eastern Hill University, Shillong, India
[2]Department of Chemistry, Sankardev College, Shillong, India

ABSTRACT

Meteorological factors are considered important driving forces for the variation of indoor radon concentrations. The present investigation aims to correlate the changes in the radon concentration with these factors using standard statistical tools and thence elicit a relation between them; this again is only a preliminary exercise and gives a first simple statistical look into correlations of indoor radon and other possible covariates.

INTRODUCTION

The interest in the study of radon (Rn-222) concentration at homes and workplaces arises primarily from the fact that it has been declared a carcinogen by various reputed agencies including the World Health Organization[1].

The study of indoor radon concentration has brought to light the temporal variations of indoor radon (Rn-222), particularly short duration periodicities (diurnal and semi-diurnal) [2]. Naturally, the question of the driving mechanism causing the fluctuations follows; and since the major causes of radon transport are advection and diffusion from soil and construction material, which in turn are influenced by meteorological factors viz. ambient temperature, pressure, humidity, wind speed etc. we are provided with motivation to study the relationship between temporal variations of a few of the meteorological parameters and indoor radon concentrations.

METHODOLOGY

Instrumentation and measurement

AlphaGuard PQ 2000 Pro manufactured by Genitron Instruments, GmbH, Germany is used for the measurement of indoor radon concentration as well as ambient temperature, pressure and humidity via integrated sensors. The sensitivity for radon is 1 counts per minute at 20 Bq/m^3, the temperature resolution is 0.2 ^{o}C, the pressure resolution is 0.1 mbar and the humidity sensitivity is 0.2%rH. [1]. The measurements were recorded every 10 minutes in an undisturbed closed room for 10 days in the Department of Physics at North-Eastern Hill University, Permanent Campus, Mawlai, Shillong, India. After the exposure period, the data is transferred to a computer and then analysed using MATLAB software.

Correlation

The simplest way to look at the relationship between two time-series is to see what happens to the other series when there is an increase or decrease in the 'first', and the general method to get a single-number answer is to use the Spearman correlation coefficient(r_S)[3].

$$r_S = \frac{\sum_{i=1}^{n} R_{x_i} R_{y_i} - \frac{1}{n}\overline{R_{x_i}R_{y_i}}}{\sqrt{\left(\sum_{i=1}^{n} R_{x_i}^2 - \frac{1}{n}\overline{R_{x_i}}^2\right)\left(\sum_{i=1}^{n} R_{y_i}^2 - \frac{1}{n}\overline{R_{y_i}}^2\right)}} \quad \dots (1)$$

Where, $x_i$and y_i are the individual measurements of the two time-series; R_{x_i} and $R_{y_i}$are their ranks and $\overline{R_{x_i}}$ and $\overline{R_{y_i}}$ are the average value of their ranks respectively; n is the total number of measurements. The ranks are assigned by first sorting the $x_i$or y_i in descending values and then assigning serial order to them.

A positive value of $r_S$would indicate that increase in one of the series is likely to be accompanied by an increase in the other and vice-versa, while a negative value would mean that the opposite behavior is likely to be observed

The Spearman's correlation only gives information on the monotonic relationship between the two variables; if additionally one wants to find the degree of linear relationship between the two variable, then generally one uses the Pearson's correlation coefficient(r_P), defined by the following equation:

$$r_P = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad \dots (2)$$

where, $\bar{x}$ and $\bar{y}$ are the mean values of the two time series and the other symbols have the meaning assigned previously in equation 1. It may be mentioned here that the Spearman correlation measures the consistency of the relationship between two variables independent of its form and the Pearson correlation the consistency of linear relationship.

Cross correlation

The information provided by correlation coefficients is a very rudimentary indicator of the similarity between two signals. For periodic signal, another important situation may arise when one of the signal follow the pattern of the other signals but at some 'lag'; this condition may be tested by performing the correlation at various lags (called the cross-correlogram) and observing the resulting plot; for lag-correlated signals, the correlogram will be a damped-periodic curve, while for uncorrelated signals it will be a decreasing curve with random-small fluctuations (see figure 1).

Figure 1 shows a correlogram of two sets of signals, the first set consists of sinusoidal signals A, B and C; the signal B has a phase difference of $\frac{\pi}{4}$ with A, while the signal C additionally also has a trend. The other set also contains three signals D, E and F made up of uniform random numbers generated by computer, additionally, F also has a linear trend. Since, the presence of trend (frequency components with periodicities much larger than the signal length) has an adverse effect on the value of the correlation coefficient being computed as can be seen from the figure, we have estimated and removed linear trends (called detrending) from all signals prior to the correlgram calculations.

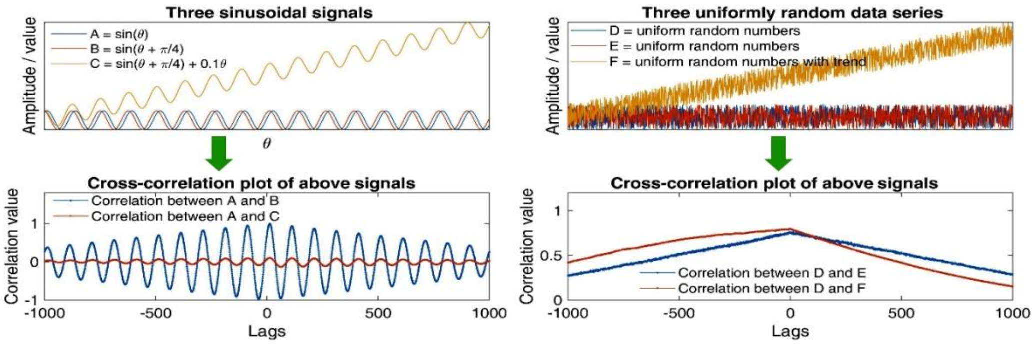


Figure 1: Cross-correlation between two highly correlated signals but with a phase difference and that between two completely random signals. The effect of trend in the correlated signals are also shown.

RESULTS AND CONCLUSION

Figure 2 shows the time series of the data analyzed in the present work viz. indoor radon concentration, temperature, pressure and relative humidity. That the signals are periodic are evidently visible in the time-series plot.

Figure 3, shows the correlation between the various parameters. The graphs in the diagonal sub-graphs are the histogram of the particular variable and the off-diagonal sub-graphs show scatter plots between the two variable listed in the row and column. The value of the correlation coefficients (Spearman and Pearson) are printed inside the scatter plots (red font indicates that the coefficients are statistically significant at 95% significance, the others are not statistically significant at this significance level). The red line indicates the best-fit line between the two variables i.e. the Pearson correlation coefficient.

A cursory visual inspection of the histograms (Figure 3) reveals that the indoor radon values are lognormally distributed, whereas the other parameters are slightly right-skewed. The lognormality of spatial distribution of indoor radon values is well documented [4, 5] but here we find that its temporal distribution is also lognormal. A thorough discussion on the distribution pattern of the variables is beyond the scope of the present work, we however would like to point to one factor (apart from the smallness of the sample length) which might bias the distribution viz. the range of fluctuations of the variables compared to the precision of the instrument – the best range to resolution ratio (in the actual experiment) was that for radon, followed by temperature, humidity and pressure in decreasing order of magnitude.

The Pearson's correlation coefficient show a positive correlation with pressure and negative with temperature and relative humidity. The Spearman correlation coefficient shows the same pattern and the coefficient values are not considerably different (slightly higher) from those of Pearson's, except for the fact that Pearson coefficient for the small positive correlation between radon and pressure is not statistical significant whereas the Spearman's is.

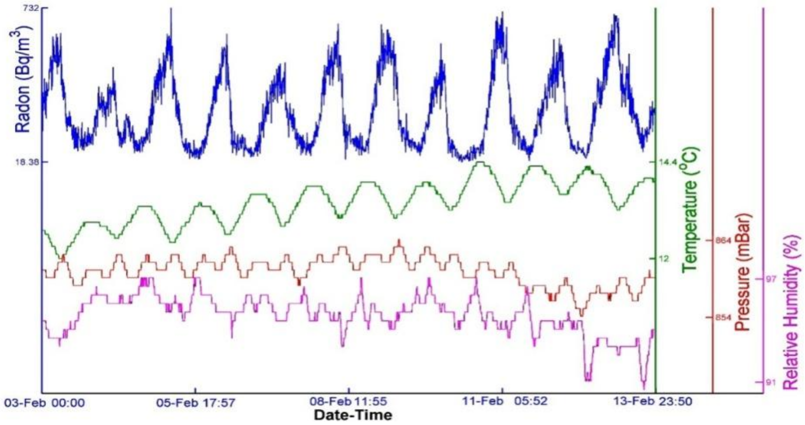


Figure 2: The radon time-series (top) and the concurrent time-series of ambient temperature, pressure and relative humidity used in the present study.

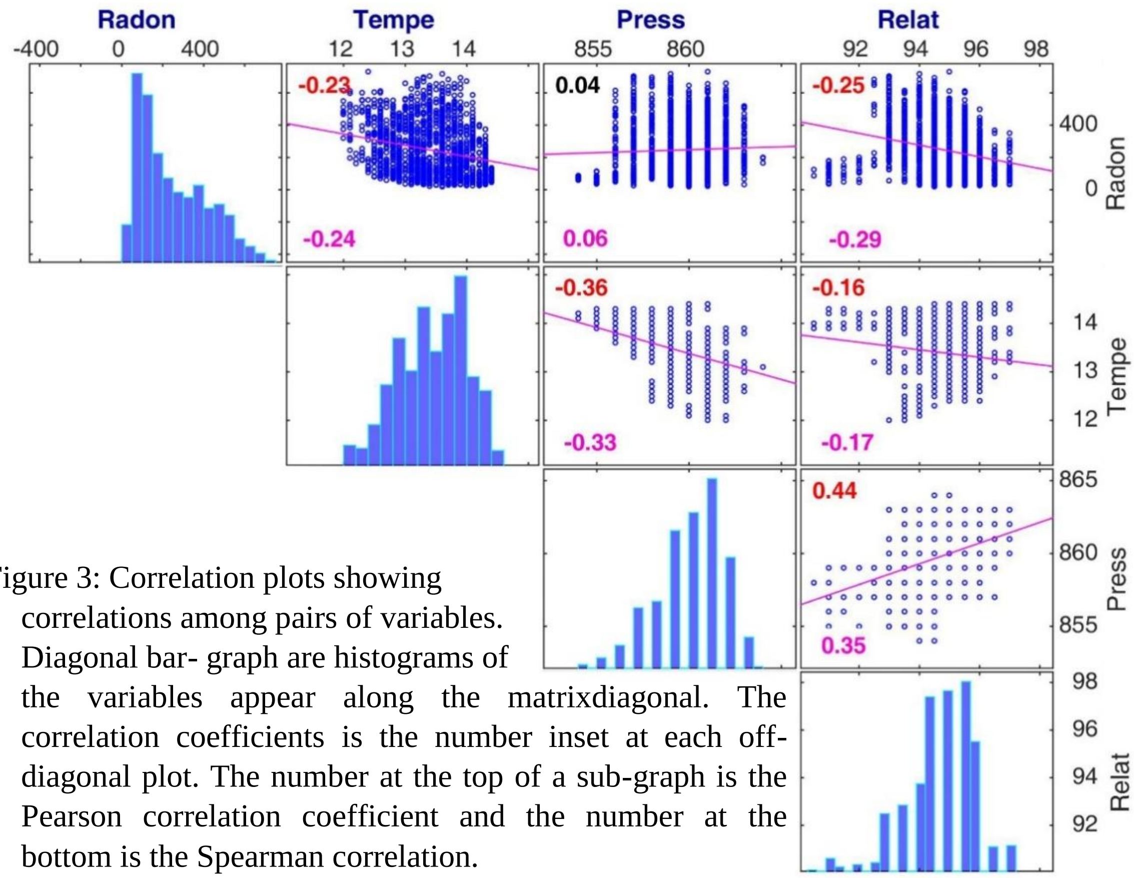
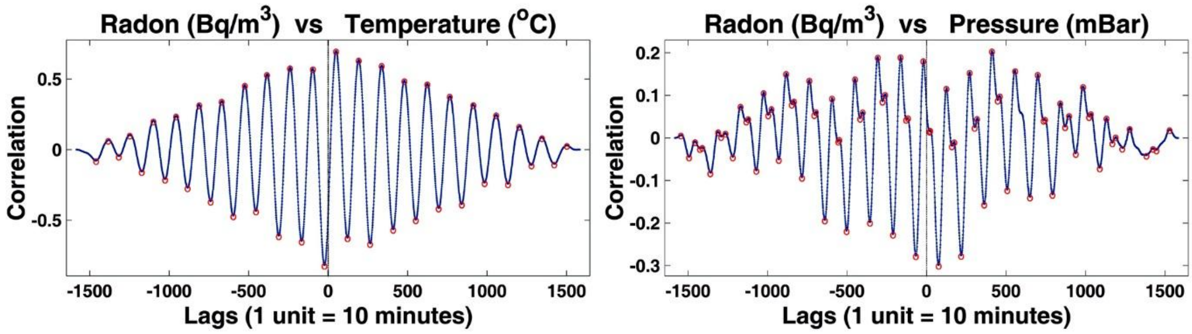


Figure 3: Correlation plots showing correlations among pairs of variables. Diagonal bar- graph are histograms of the variables appear along the matrixdiagonal. The correlation coefficients is the number inset at each off-diagonal plot. The number at the top of a sub-graph is the Pearson correlation coefficient and the number at the bottom is the Spearman correlation.

The cross-correlogram patterns (Figure 4) show that there are significant positive correlations between radon and other parameters at various lags i.e. cross-correlation itself oscillates significantly above and below zero indicating that strong correlation exist between the given pair of variables (cf. figure 1). The first peak for cross-correlation between radon and temperature occurs at lag 50 (8.3 hours) on the positive lag direction and the second peak occurs at lag 192 (32 hours) about 24 hours from the first; similarly, the first peak on the negative lag direction occurs at -96 (-16 hours) and the second peak is at lag -240 (-40 hours) again 24 hours before the first peak. The positive lag direction indicates by what amount the temperature time-series is being shifted forward in relation to the radon time-series, while the negative lag direction indicates the shifting backwards.



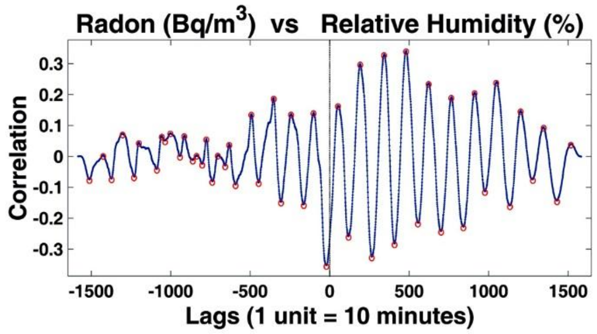


Figure 4: Cross-correlogram between radonand other meteorological factors. The red circles indicate the location of the local extrema.

The first trough (minima) occurs at lag 122 (20.3 hours) followed by lag 264 (44 hours) in the positive lag direction while for the negative lag direction the first peak occurs at lag -23(-3.8 hours) followed by lag -166 (-27.7 hours). This translates to the fact that the two series align constructively on shifting temperature time-series 8 hours forward (or 16 hours backwards); thus, on a typical day, the increase in radon is observed first followed by increase in ambient temperature at a separation of 8 hours.

For radon and relative humidity, the relationship is similar: the first peak towards the positive lag direction is observed at lag 52 (8.7 hours) followed by lag 192 (32 hours) about 24 hours from the first. Towards the negative lag direction, the first peak is observed at lag -102 (-17 hours) and the second peak 24 hours later at lag -243 (-40.5 hours). In this case also, the first trough occurs at lag 119 (19.8 hours) later than the first peak, followed by lag 265 (44.2 hours) on the positive lag direction, while in the negative lag direction the first trough is observed earlier at lag -20 (-3.3 hours) than the first peak, followed by lag -163 (-27.2 hours).

For radon and ambient pressure, the relationship is different: the trough occurs before the peak in the correlogram indicating that on gradually shifting the pressure-time series forward, the two-time series align oppositely first and then align constructively. The first peak in the positive lag direction is obtained at lag 23 (3.8 hours) followed by lag 125 (20.8 hours) whereas the first trough occurs at lag 19 (3.2 hours) followed by lag 76 (12.7 hours). The time-lag between the peaks is 17 hours and also as can been seen from the correlogram, there are two peaks within one day – one small and the other comparatively larger. This may be interpreted as follows: that on shifting the pressure time-series gradually for a period of 24 hours there are two points within the time-interval when the two series align constructively – the first alignment being the smaller one.

We can summarize the above results as confirming repeating patterns in the radon and other meteorological time-series and that the patterns have a similar trend but at some time-lags and following the well-known adage "correlation does not imply causation", we are forced to put a caveat to our conclusion that although patterns have been detected but which of the covariates is the primary driving force is hard to conclude from the limited data and analyses done here. A more controlled set of experiments are required to shed further light into the actual strength of the relationships that are identified in the present work.

REFERENCES

1. Deveshwori Maibam, *Studies on radon activity in soil-air, indoors and in water at selected locations in meghalaya*, in *Department of Physics*. 2014, North-Eastern Hill University: Shillong. p. 234.
2. Yubaraj Sharma, Deveswori Maibam, and Atul Saxena, *Frequency and time-frequency analysis of indoor radon time-series*, in *Climatic Aberrations: Contemporary dilemma of the anthropogenic climate in socio and ecological issues*. 2015: Sankerdev College, Shillong.
3. Frederick J Gravetter and Larry B Wallnau, *Statistics for the behavioral sciences*. 2015: Cengage Learning.
4. Jon Miles, *Mapping radon-prone areas by lognormal modeling of house radon data.* Health Physics, 1998. **74**(3): p. 370-378.
5. P. Bossew, *Radon: Exploring the log-normal mystery.* Journal of Environmental Radioactivity, 2010. **101**(10): p. 826-834.