Multiple linear regression analysis to estimate hydrological effects in soil Rn-222 at Ghuttu, Garhwal Himalaya, India: a prerequisite to identify earthquake precursors

Vishal Chauhan[1,3], Naresh Kumar[1,*], Vaishali Shukla[1,2] and Sanjay Kumar Verma[1]

[1]Wadia Institute of Himalayan Geology, 33 General Mahdeo Singh Road, Dehradun 248 001, India
[2]Department of Geology, Kumaun University, Sleepy Hollow, Nainital 263 001, India
[3]Department of Physics, Graphic Era Hill University, Dehradun 248 002, India

Various geophysical parameters including soil radon (^{222}Rn) are being continuously monitored at Ghuttu, Garhwal Himalaya, India since 2007 as a part of earthquake precursor studies. To analyse the earthquake precursory changes in soil radon, it is essential to clean the soil radon data from other effects. For this, we used data for the period of nine years from 2011 to 2019 and assessed the relationship of soil radon with five other parameters using regression analysis. These parameters are water level, atmospheric pressure, rainfall, air temperature and soil temperature at 10 m depth. We also added one more parameter, i.e. the difference of air temperature (T_{out}) and soil temperature at 10 m depth (T_{in}). From the observed six parameters, four showed strong correlation with soil radon. These are (i) water level (correlation coefficient (CC) = –0.9), (ii) atmospheric pressure (CC = 0.6), (iii) air temperature (CC = –0.6) and (iv) temperature difference ($T_{out} - T_{in}$; CC = 0.5). For regression analysis, data during the period 2011–2014 were used for training, while data during 2015–2019 were used for testing purpose. Based on different models, the one developed using all the six input parameters suggests lowest errors and highest correlation. The observed values of root mean square error, mean absolute error and CC were 0.332, 0.281 and 0.931 respectively. The regression coefficients obtained from this model were used to calculate the theoretical radon and residuals. By this approach, the effects of hydrological and atmospheric parameters were found to be reduced to a great extent.

Keywords: Earthquake precursors, hydrological effects, linear regression, soil radon.

A variety of geophysical and seismological parameters have been identified as potential earthquake precursors. These include unusual deviation in the time series of atmospheric/ionospheric, geodetic, geomagnetic, geoelectric and geochemical fields, and also seismological parameters. Among the precursors, radon was found to be a promising precursor due to its distinctive geochemical properties[1–4]. Radon (^{222}Rn), the daughter isotope of uranium–lead radioactive decay series is an inert gas with a half-life of 3.82 days. It is present as a trace element in soils and rocks. Radium (^{226}Ra) is the parent nuclide of radon and exhalation of radon depends on its presence in the mineral grains. After originating from the solid grains, radon gas propagates through the soil pores and then reaches the atmosphere by diffusion and advection[5]. A high rate of radon emanation is possible through fractured rocks. Radon concentration increases near tectonic faults[6]. Choubey *et al.*[6] measured radon concentration in soil gas and groundwater samples collected from different places of major tectonic zones in the area between Ghansali and Ghuttu in the Bhilangana valley, Garhwal Himalaya, Uttarakhand, India. They reported high radon concentration near the tectonic plane, including the present study region. The exhalation rate of radon depends on various factors which include the lithology of the area as well as physical properties of the rocks and soils such as elasticity, porosity, permeability, homogeneity and fragmentation. Radon concentration is also influenced by diurnal and seasonal effects, and the effects of environmental parameters like air temperature, soil temperature, soil moisture, barometric pressure and rainfall[5–8]. Most of the environmental parameters are inter-related and their major influence on the emanation rate of radon gas is widely reported. For example, Klusman and Jaacks[7] observed daily effects in the soil gases of Hg, Rn and He at a site for over 22 months. Simultaneously, they monitored other parameters to assess the effect of environmental parameters on the emanation rate of gases. By applying stepwise multiple regression, they found that the environmental parameters account for 83% of the total radon variance. Schery *et al.*[8] measured radon from a

*For correspondence. (e-mail: nkd@wihg.res.in)

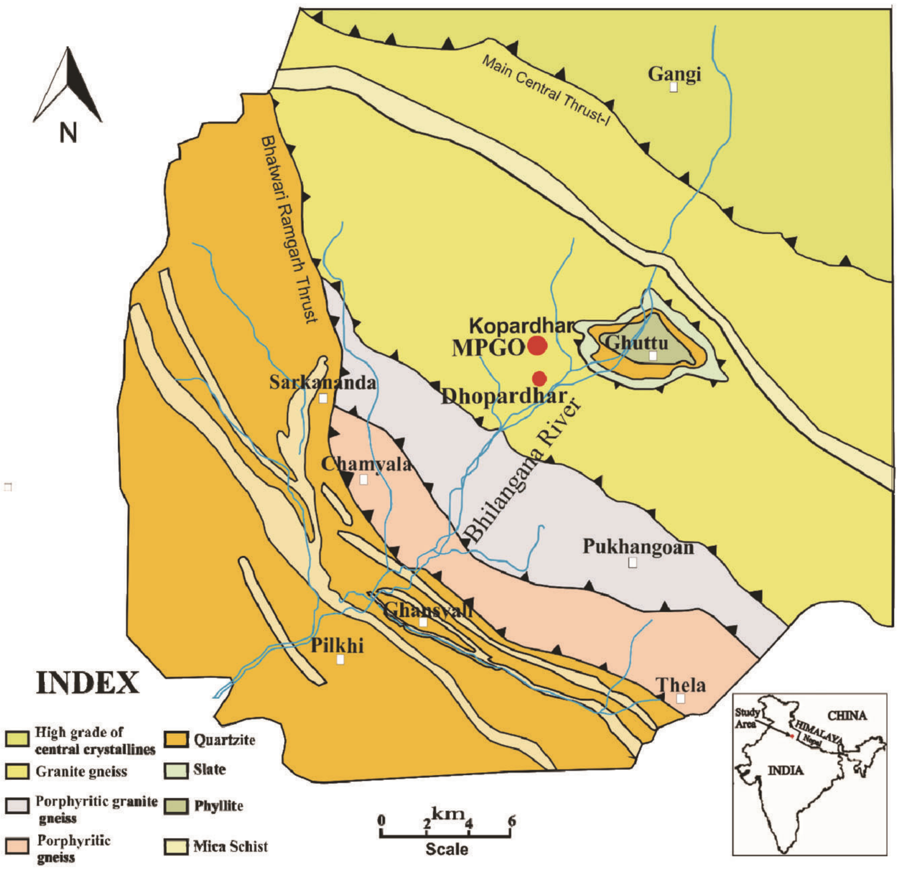

Figure 1. Location map of MPGO, Ghuttu, Garhwal Himalaya, India and surrounding geological features (after Vyshnavi *et al.*[17]).

gravely sandy loam in a semi-arid climate site, ~1 km west of the centre of the New Mexico Institute of Mining and Technology campus, New Mexico. They noticed a clear and major effect of atmospheric pressure and rain precipitation due to rainfall on the instantaneous values of radon, whereas effects due to other parameters such as wind and temperature were either comparatively small or undetectable.

We examined the data of soil radon measured at the Multi-Parametric Geophysical Observatory (MPGO) at Ghuttu, which was established in 2007 with an aim to carry out earthquake precursory research in an integrated manner using several parameters at a single site. Details of various equipment installed at MPGO, have been reported earlier along with the data[9–12]. Meteorological parameters like rainfall, air temperature, atmospheric pressure and temperature at a depth of 10 m are also being recorded simultaneously. MPGO, is located in a region of active geodynamics and high seismic activity. Tectonically, it is situated to the south of the Main Central Thrust (MCT-I) in the Himalayan Seismic Belt (HSB). In the recent past, two major earthquakes have struck near this region, i.e. Uttarkashi earthquake (*Mb* 6.5) in 1991 and Chamoli earthquake (*Mb* 6.3) in 1999. The study region is located in the middle of the Central Himalayan seismic gap of the great earthquake occurrences. These were the main aspects to establish this kind of state-of-the-art observatory at Ghuttu. However, this site has not witnessed any significant earthquake nearby since its installation; however, some minor to moderate-magnitude earthquakes have occurred. Choubey *et al.*[9] reported anomalous variations in radon emanation, measured at MPGO, prior to the occurrence of the *M* 4.9 Kharsali earthquake of 23 July 2007. Kumar *et al.*[10] reported large changes in radon emissions at MPGO prior to the occurrence of the Gorkha–Nepal earthquake (*M* 7.8) of 25 April 2015. In a recent study, Shukla *et al.*[11] assessed variation of radon emissions corresponding to 20 small to moderate magnitude earthquakes which occurred within an epicentral distance of 300 km. In addition the influence of seasonal, diurnal, semi-diurnal and multi-day recurring effects on radon data have also been estimated at MPGO[12]. This also includes changes in radon emanation due to atmospheric and soil temperature.

The above mentioned observations motivated us to study the long term effects of meteorological and hydrological parameters on the variation of radon exhalation to distinguish the anomalies related to seismological events. Different statistical methods are available to study the contribution of environmental parameters to radon exhalation rates[13–15]. Here, we have used multiple linear regression analysis to assess and remove the atmospheric and hydrological effects from the subsurface soil radon data.

Data and method of analysis

The radon observation site, MPGO (30.53°N, 78.74°E) is located in the Bhilangana valley, Uttarakhand (Figure 1). It is situated at an elevation of 1835 m amsl and has seasonally variable climate. In the Himalayan tectonic framework, Ghuttu is located just south of MCT-I on the northern margin of the outer Lesser Himalaya. On a very local scale, it is within the Ghuttu window, which consists of low-grade metamorphic rocks[16]. Carbonaceous phyllite forms the lower most part of the stratigraphic sequence in the Ghuttu window. It is overlain by sheared quartzites interspersed with thin limestone layers and slates. Further, rocks of the Ghuttu window are surrounded by those of the central crystallines sequence[16,17]. There are plentiful evidences for uranium content in the rocks of this region which are the main source of ^{222}Rn as a radioactive daughter element[16].

The radon measurement PM-11 probe of Ghuttu is a gamma-particle detecting instrument. It is installed at a depth of 10 m in a borehole to collect soil radon. Radon measurement is carried out by a 2″ × 2″ NaI(Tl) scintillation detector, based on total gamma counting system. The total gamma rays detected by the PM-11 probe are in the range 20 keV–3 MeV that are emitted during the radioactive decay, giving rise to radon as a daughter product of the uranium–lead series.

We applied multiple linear regression analysis to remove atmospheric and hydrological effects from the radon data. This is important to glean the earthquake-induced anomalous variations in the radon data. Regression analysis is used to compute correlation between two or more variables. Regression using one single independent variable is called univariate regression analysis and that using more than one independent variables is called multivariate analysis (Uyanik and Guler[18], and references therein). The equation of multivariate regression model is as follows

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon,$$

where y is the dependent variable, x_i the independent variables, β_i the parameters and ε is the error.

The processing of data based on the regression model equation is preceded by some pre-analysis steps. Apart from radon data, which are considered as the dependent variable, data of five other parameters, namely water level, atmospheric pressure, rainfall, air temperature and soil temperature at 10 m depth were also collected. We also calculated the difference of air temperature (T_{out}) and soil temperature at 10 m depth (T_{in}), and used it as one more parameter. One-to-one correlation of each parameter with radon and colinearity with each other, were measured using linear regression analysis. Further, different models were developed by changing the input variables. Finally, theoretical values of radon time series were generated using the model equation with least errors and highest correlation. The residual was also calculated for training and testing periods.

Results and discussion

Figure 2 shows plots of continuous time series of soil radon, atmospheric pressure, air temperature, soil temperature at 10 m depth, changes in groundwater level and rainfall measured at MPGO. Also, the average values of daily data are plotted for the period of nine years from 2011 to 2019. It is observed that the annual variation in soil radon varies between 2456 × 100 and 3260 × 100 radon counts with a standard deviation of 246 (Figure 2 *a*). It is to be noted here that the accuracy of the PM-11 probe to measure gamma counts is 1 count/h =0.003 Bq/m^3. The minimum concentration of soil radon is observed in the rainy season, mainly during June to September. This may be due to the saturation of rock pores with water and thus less transportation of radon to the surface. In the post-monsoon period, the radon concentration slowly increases as the groundwater depletes.

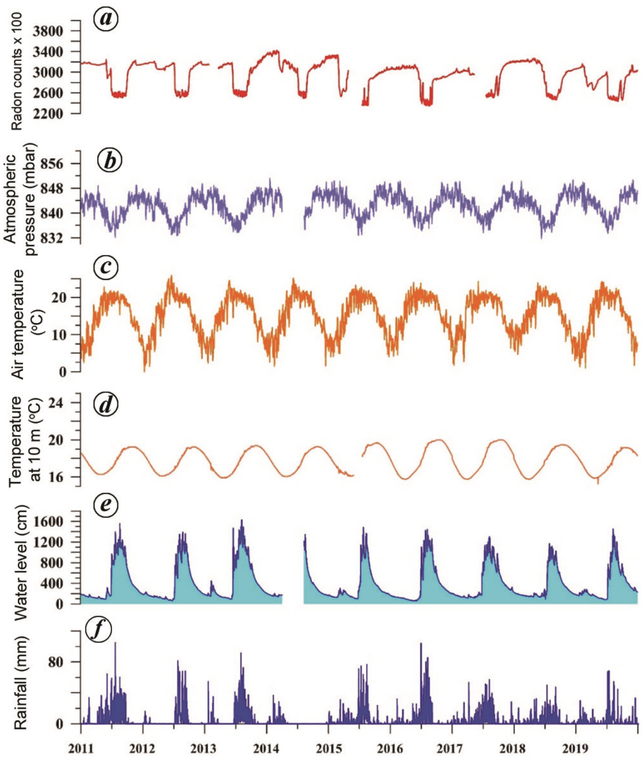

Figure 2. ***a***, Continuous time-series radon of soil at 10 m depth; ***b***, atmospheric pressure; ***c***, air temperature; ***d***, temperature at 10 m depth; ***e***, water level; ***f***, rainfall.

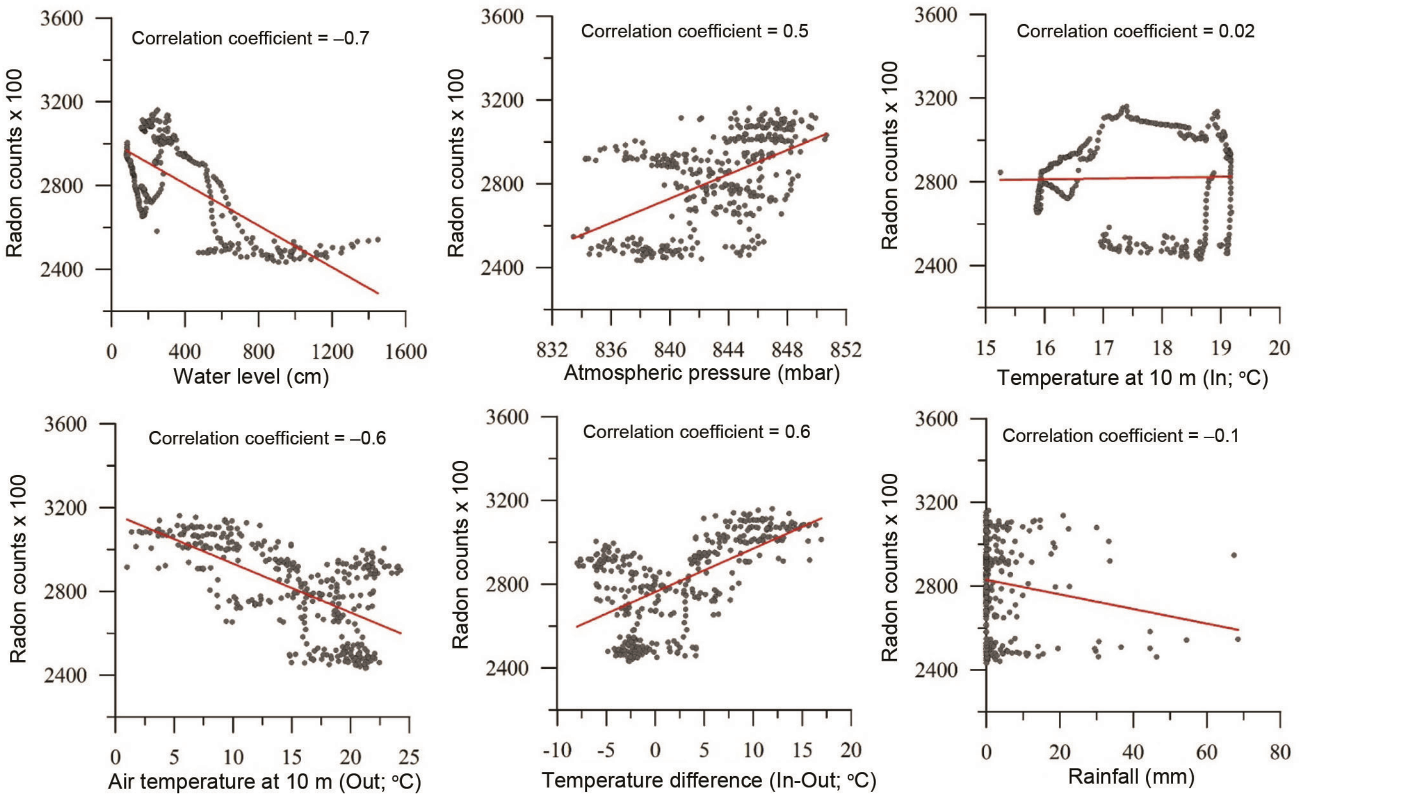

Figure 3. Linear regression plots between Rn-222 and other time series (hydrological and meteorological) for the duration of one year, i.e. 2019.

Therefore, radon concentration is highest during peak summer, before the start of the rainy season. Atmospheric pressure is found to vary between 833 and 842 mbar (Figure 2 *b*). The air temperature varies in the range 2°–24°C, whereas temperature at 10 m depth changes annually between 16.1°C and 19.7°C (Figure 2 *c*). The water level has a variation of 1328 cm (Figure 2 *e*), whereas the rainfall has maximum value of 74 mm (Figure 2 *f*). Seasonal variation and inter-dependence of each parameter is clearly visible.

To understand the relationship of soil radon with other parameters, we performed regression analysis and calculated the correlation coefficients (CCs). Figure 3 shows the results for 2019. It is clearly visible that radon concentration has a negative but strong correlation (CC = –0.7) with water-level changes. The soil radon also has a good, but negative correlation (CC = –0.6) with air temperature. However, it has a positive correlation with atmospheric pressure (CC = 0.5). We also obtained the difference of atmosphere temperature (outer temperature) with soil temperature (10 m depth). The temperature difference between inside and outside of the borehole shows a positive correlation with soil radon (CC = 0.6). Temperature at 10 m depth and rainfall do not show any correlation with soil radon. A strong correlation of radon with the groundwater changes and negligible effect of rainfall indicate that hydrological effect on radon is not instantaneous. Continuous spells of rainfall during the monsoon season slowly and incrementally lead to percolation of water in the ground. Gradually, the pores of the soil and fractures of subsurface rocks are saturated with water. The movement of radon through the soil is controlled by moisture content, porosity and permeability of soil. Sun *et al.*[19] pointed out that when water content is low, the emanation of soil radon increases. However, when water content has reached a certain level, the radon diffusion and advection process is largely weakened by the water in the soil pores, which leads to a decrease in soil radon.

Further, to understand the interdependency and colinearity among the above-mentioned parameters, we calculated the CC for each parameter with other parameters. Figure 4 shows yearly matrix plots (2011–2019) with CC values for all the parameters. Table 1 gives the average values of these CCs. Four independent variables are found to depict strong correlation with the dependent variable radon. These are water level, atmospheric pressure, air temperature and temperature difference ($T_{out} - T_{in}$). A colinearity is also observed among the independent variables. For example, air temperature is inversely correlated to water level (CC = –0.5) and atmospheric pressure (CC = –0.6), and positively to temperature difference (CC = 1.0). The temperature difference also shows a good correlation with atmospheric pressure (CC = 0.6). The inter-dependency indicates that a model can be formulated using any two variables, i.e. water level and atmospheric

Table 1. Averages of correlation coefficients (2011–2019) among all the borehole parameters, i.e. radon, atmospheric pressure, air temperature, temperature at 10 m depth, water level, rainfall and temperature difference

	Atmospheric pressure	Air temperature	Temperature at 10 m	Water level	Rainfall	Temperature difference
Radon in soil	0.6	–0.6	–0.2	–0.9	–0.4	0.5
Atmospheric pressure		–0.6	0.2	–0.5	–0.3	0.6
Air temperature			–0.1	0.5	0.1	–1
Temperature at 10 m				0.4	0	0.3
Water level					0.4	–0.4
Rainfall						–0.1

Table 2. Statistics for the models developed by checking the input variables

Inputs	Period	RMSE	MAE	R	R^2
WL	2011–2014	0.4271	0.3584	0.8840	0.7815
AP	2011–2014	0.7327	0.6179	0.5973	0.3568
AT	2011–2014	0.7598	0.6371	0.6076	0.3692
TD	2011–2014	0.8346	0.6726	0.4888	0.2389
PT	2011–2014	0.9508	0.7844	0.1110	0.0123
RA	2011–2014	0.8459	0.6457	0.4671	0.2182
WL + AP	2011–2014	0.3623	0.3036	0.9180	0.8427
WL + AP + AT	2011–2014	0.3458	0.3028	0.9256	0.8567
WL + AP + AT + TD	2011–2014	0.3331	0.2825	0.9312	0.8671
WL + AP + AT + TD + PT	2011–2014	0.3330	0.2824	0.9312	0.8671
WL + AP + AT + TD + PT + RA	2011–2014	0.3325	0.2817	0.9314	0.8676

Note: WL, Water level; AP, Atmospheric pressure; AT, Atmospheric temperature; TD, Temperature difference; PT, Temperature at 10 m; RA, Rainfall; RMSE, Root mean square error; MAE, Mean absolute error; R, Correlation coefficient; R^2, coefficient of determination.

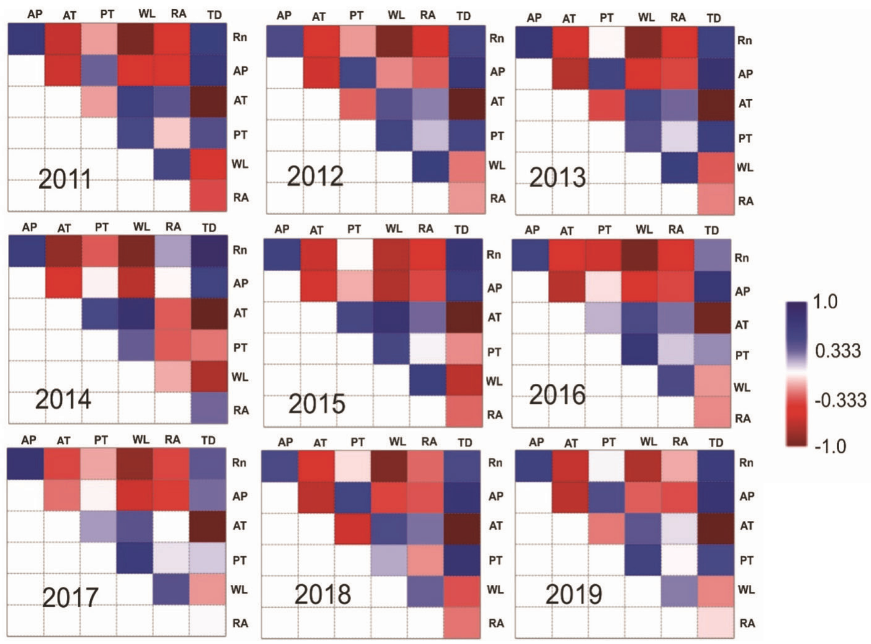

Figure 4. Linear regression correlation coefficients computed for different borehole data, i.e. radon (Rn), atmospheric pressure (AP), air temperature (AT), temperature at 10 m depth (PT), water level (WL), rainfall (RA) and temperature difference (TD).

pressure. However, to assess all possibilities, we developed different models by changing the input variables. Initially for a training procedural model, we selected data for the four period from 2011 to 2014. After normalizing all the data, we developed different models and calculated errors and CCs. Table 2 shows the statistics of different models developed for the testing period by changing the input variables. These include root mean square error (RMSE), mean absolute error (MAE), correlation coefficient (R) and coefficient of determination (R^2;

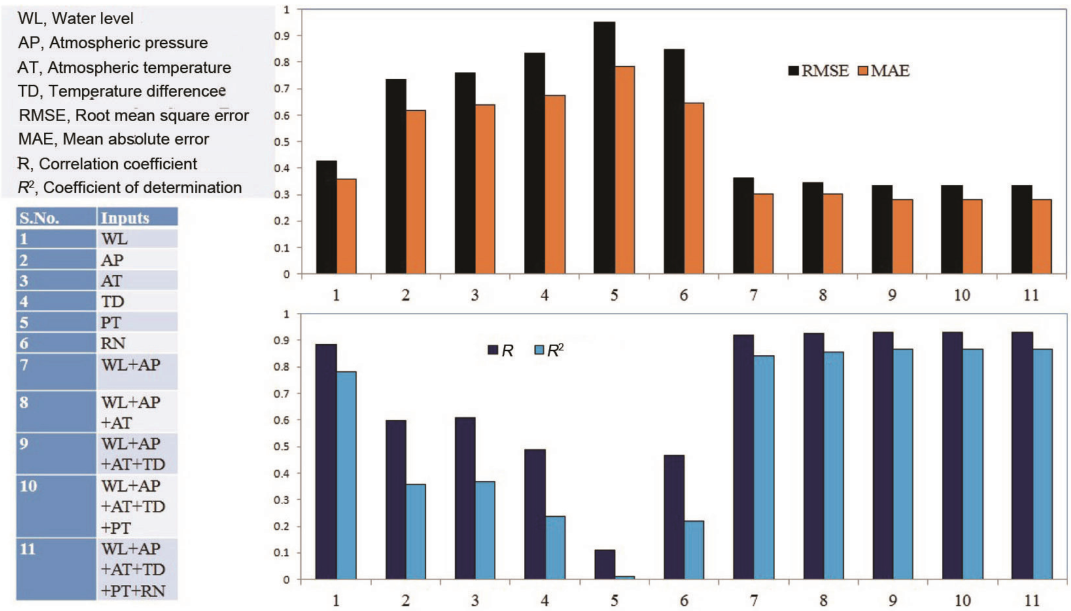

Figure 5. (Upper panel) Errors and (lower panel) correlation coefficients estimated in different models by varying independent input parameters of testing data (2011–2014).

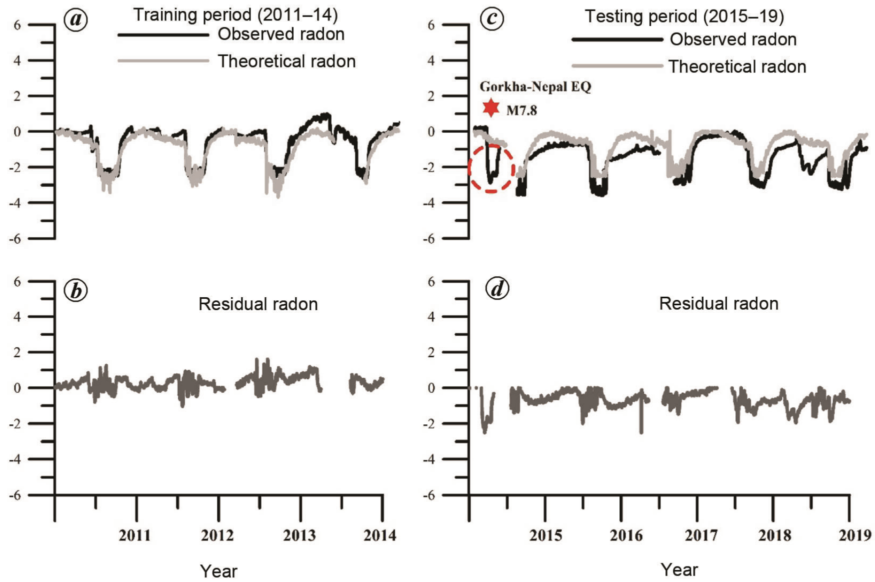

Figure 6. Plots of (***a***) observed radon and theoretical radon for the training period (2011–2014), (***b***) residual radon for the training period, (***c***) observed radon and theoretical radon for the testing period (2015–2019) (star indicates the date of occurrence of the Gorkha–Nepal earthquake (M 7.8), viz. 25 April 2015) and (***d***) residual radon for the testing period.

Table 2 and Figure 5). The models developed using individual parameters show lowest error (RMSE = 0.4271, MAE = 0.3584) and high CC (R = 0.8840) in the case of water level, large high errors are observed with other individual inputs (RMSE ranging between ~0.7 and 0.9). In the case of multiple parameters, five combinations are used to develop the models and all the models show almost similar results. The reason for these similarities is the colinearity among some of the individual parameters. However, the highest correlation (R = 0.9314) and lowest errors (RMSE = 0.3325, MAE = 0.2817) are observed in the model developed using all the six parameters. Further the regression equation generated using six input parameters was employed to calculate the theoretical values of soil radon for both the training and testing periods. The testing period was chosen from 2015 to 2019. Figure 6 shows the variation of observed soil radon, theoretical soil radon and the residuals. The figure shows that the hydrological and meteorological effects are minimized from soil radon data at a certain limit. The annual decrease in soil radon is observed when the values of water level are high during the rainy season. However, a large decrease before the rainy season is also observed between 26 February and 27 April in 2015, even after removing all effects. It is to be noted here that the devastating Gorkha–Nepal earthquake (*M* 7.8) occurred in the central part of the Nepal Himalaya on 25 April 2015. The epicentre of this earthquake was 636 km away from MPGO. Another earthquake of magnitude 5.1 also occurred on 1 April 2015 in the Chamoli region of Uttarakhand, and its epicentre was only 73 km away from MPGO. Therefore, the observed anomalous decrease in radon in which the other effects are minimized, may be the precursory signature of these earthquakes. The observed anomaly can be interpreted in terms of the dilatancy–diffusion model, which suggests the build-up of strain, dilatancy of the rocks, diffusion and influx of water in the rocks during the earthquake preparation period. This may be a possible reason for decreased emanation rate of soil radon.

It can be noticed that the multiple linear regression analysis can considerably minimize different kinds of effects from radon data. However, there is scope for applying more rigorous modelling. Therefore, work on more elongated modelling techniques like artificial neural networks is in process.

Conclusion

Regression analysis performed on soil radon data of MPGO, suggests major effect of water level on soil radon (CC = –0.9). The atmospheric pressure, air temperature and the difference between inner (soil) and outer (air) temperatures also show effects on the variation of soil radon. By considering all these effects, different models are developed using multiple regression analysis. The model with maximum input parameters, least errors (RMSE = 0.332, MAE = 0.281) and high correlation ($R = 0.931$) is chosen for the calculation of theoretical values of radon. The residuals obtained from observed and theoretical radon enhance the anomalous changes for seismic activity. Using this method, the hydrological and environmental effects are reduced from radon data to a fair extent. These results can be enhanced by employing more arduous statistical methods.

1. Verma, M. and Bansal, B. K., Earthquake precursory studies in India: scenario and future perspectives. *J. Asian Earth Sci*., 2012, **54–55**, 1–8.
2. Arora, B. R. *et al.*, Assessment of the response of the meteorological/hydrological parameters on the soil gas radon emission at Hsinchu, northern Taiwan: a prerequisite to identify earthquake precursors. *J. Asian Earth Sci.*, 2017, **149**, 49–63.
3. Walia, V. *et al.*, Temporal variation of soil gas compositions for earthquake surveillance in Taiwan. *Radiat. Meas*., 2013, **50**, 154–159.
4. Oh, Y. H. and Kim, G., A radon–thoron isotope pair as a reliable earthquake precursor. *Nature*, 2015, doi:10.1038/srep13084.
5. Kulali, F., Akkurta, I. and Özgürb, N., The effect of meteorological parameters on radon concentration in soil gas. *Acta Phys. Pol. A*, 2017, **132**(3-II), 999–1001.
6. Choubey, V. M., Ahmad, I., Kamra, L. and Ramola, R. C., Radon variations in soil and groundwater of Bhilagana valley, Garhwal Himalaya, India. *Jpn. J. Health Phys*., 2010, **45**(3), 278–283.
7. Klusman, R. W. and Jaacks, J. A., Environmental influences upon mercury, radon and helium concentrations in soil gases at a site near Denver, Colorado. *J. Geochem. Explor*., 1987, **27**, 259–280.
8. Schery, S. D., Gaeddert, D. H. and Wilkening, M. H., Factors affecting exhalation of radon from a gravelly sandy loam. *J. Geophys. Res*., 1984, **89**, 7299–7309.
9. Choubey, V. M., Kumar, N. and Arora, B. R., Precursory signatures in the radon and geo-hydrological borehole data for *M* 4.9 Kharsali earthquake of Garhwal Himalaya. *Sci. Total Environ*., 2009, **407**, 5877–5883.
10. Kumar, N., Chauhan, V., Dhamodharan, S., Rawat, G., Hazarika, D. and Gautam, P. K. R., prominent precursory signatures observed in soil and water radon data at multi-parametric geophysical observatory, Ghuttu for *M* 7.8 Nepal earthquake. *Curr. Sci*., 2017, **112**(5), 907–909.
11. Shukla, V., Chauhan, V., Kumar, N. and Hazarika, D., Assessment of Rn-222 continuous time series for the identification of anomalous changes during moderate earthquakes of the Garhwal Himalaya. *Appl. Radiat. Isot*., 2020, **166**, 109327(1-10).
12. Choubey, V. M., Arora, B. R., Barbosa, S. M., Kumar, N. and Kamra, L., Seasonal and daily variation of radon at 10 m depth in borehole, Garhwal Lesser Himalaya, India. *Appl. Radiat. Isot*., 2011, **69**, 1070–1078.
13. Gelman, A., Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics*, 2006, **48**(3), 432–435.
14. Owusu-Amponsah, N. Y., Using GIS and regression analysis to evaluate physical factors of radon concentration. *Resour. Anal.*, 2015, **18**, 13.
15. Singh, S., Jaishi, H. P., Tiwari, R. P. and Tiwari, R. C., Time series analysis of soil radon data using multiple linear regression and artificial neural network in seismic precursory studies. *Pure Appl. Geophys*., 2017, **174**, 2793–2802.
16. Islam, R. and Thakur, V. C., Geology of Bhilangana Valley, Garhwal Himalaya. *Geosci. J.*, 1988, **9**(2), 143–152.
17. Vyshnavi, S., Islam, R. and Sundriyal, Y. P., Comparative study of soil profiles developed on metavolcanic (basaltic) rocks in two different watersheds of Garhwal Himalaya. *Curr. Sci*., 2015, **108**(4), 699–707.
18. Uyanik, G. K. and Guler, N., A study on multiple linear regression analysis. *Proc. Soc. Behav. Sci*., 2013, **106**, 234–240.
19. Sun, K., Guo, Q. and Cheng, J., The effect of some soil characteristics on soil radon concentration and radon exhalation from soil surface. *J. Nucl. Sci. Technol*., 2004, **41**(11), 1113–1117.

ACKNOWLEDGEMENTS. We thank the Director, Wadia Institute of Himalayan Geology, Dehradun, for providing the necessary facilities to carry out this study, and the Ministry of Earth Sciences, Government of India for providing the funds to run the MPGO station through a sponsored project (MoES/P.O. (Seismo)/NPEP/15/2009 and MoES/P.O. (Seismo)/1(256)/2015). We also thank the entire team of MPGO for help with data collection, the staff working at remote site of Ghuttu for assistance. This work has Wadia Institute contribution number WIHG/0083.

Received 17 September 2020; revised accepted 16 March 2021

doi: 10.18520/cs/v120/i12/1905-1911