RADON ANOMALIES IN SOIL GAS CAUSED BY SEISMIC ACTIVITY

BORIS ZMAZEK, MLADEN ŽIVČIĆ, LJUBČO TODOROVSKI, SAŠO DŽEROSKI, JANJA VAUPOTIČ, IVAN KOBAL

About the authors

Boris Zmazek
The Jožef Stefan Institute,
Jamova 39, 1000 Ljubljana, Slovenia
Phone: + 386 1 477 35 80
Fax: + 386 1 477 38 11
E-mail: boris.zmazek@ijs.si

Mladen Živčić
The Jožef Stefan Institute,
Jamova 39, 1000 Ljubljana, Slovenia
Office of Seismology, Environmental Agency of the Republic of Slovenia,
Dunajska 47/VII, 1000 Ljubljana, Slovenia

Ljubčo Todoroski, Sašo Džeroski, Janja Vaupotič, Ivan Kobal
The Jožef Stefan Institute,
Jamova 39, 1000 Ljubljana, Slovenia

Abstract

At the Orlica fault in the Krško basin, combined barasol detectors were buried in six boreholes, two along the fault itself and four on either side of it, to measure and record radon activity, temperature and pressure in soil gas every 60 minutes for four years. Data collected have been analysed in a manner aimed at distinguishing radon anomalies resulting from environmental parameters (air and soil temperature, barometric pressure, rainfall) from those caused solely by seismic events. The following approaches have been used to identify anomalies: (i) ± 2σ deviation of radon concentration from the seasonal average, (ii) correlation between time gradients of radon concentration and barometric pressure, and (iii) prediction with regression trees within a machine learning program. In this paper the results obtained with regression trees are presented. A model has been built in which the program was taught to predict radon concentration from the data collected during the seismically inactive periods when radon is presumably influenced only by environmental parameters. A correlation coefficient of 0.83 between measured and predicted values was obtained. Then, the whole data time series was included and a significantly lowered correlation was observed during the seismically active periods. This reduced correlation is thus an indicator of seismic effect.

Keywords

radon in soil gas, environmental parameters, earthquakes, correlation, regression trees, forecasting

1 INTRODUCTION

Since the advent of the nuclear era in Slovenia in sixties, radon (^{222}Rn) and other radionuclides have been systematically monitored in ground and surface waters [1-7]. The first radon analyses, aimed at forecasting earthquakes [8-15], were carried out in Slovenia in 1982 [16]. In four thermal water springs, radon concentrations were determined weekly, while Cl^-, SO_4^{2-}, hardness and pH value, were determined monthly. In 1998, this study was extended to other thermal water springs [17-20] and also to soil gas [20-22] at selected, seismically relevant sites, and sampling frequency was increased from once a week to once an hour.

It is often difficult to distinguish a radon anomaly caused solely by a seismic event, from one resulting from meteorological or hydrological parameters, therefore the implementation of more advanced statistical methods in data evaluation [23-29] is important. We have found that among these methods, regression/decision trees may be very successful for this purpose and outperformed the others [30]. We teach the program to predict radon concentrations on the basis of environmental data (air and soil temperature, barometric pressure, rainfall) during seismically non-active periods, and then apply the hypothesis that the prediction is significantly worsened during seismically active periods.

In this paper we shall focus on the radon concentration in soil gas in the Krško basin. For earthquakes, Dobrovolsky's [31] equation was used to calculate the radius of the zone within which precursory phenomena may be manifested (so-called Dobrovolsky's radius R_D):

$$R_D = 10^{0.43M} \qquad (1)$$

where M is the earthquake magnitude. Earthquakes for which the distance R_E between the epicentre and our measuring site was equal or less than $2R_D$ have been taken into account.

2 EXPERIMENTAL

Since April 1999, in 60-90 cm deep boreholes at six locations in the Krško basin, radon concentration in soil gas, barometric pressure and soil temperature have been measured and recorded once an hour, using barasol probes (MC-450, ALGADE, France). Other meteorological data, such as air temperature and rainfall, have been provided by the Office of Meteorology of the Environmental Agency of the Republic of Slovenia, and seismic data by the Office of Seismology of the same agency. Boreholes 1 and 4 are located in the Orlica fault zone, at a distance about 4000 m from each other, while the other boreholes are at distances from 150 to 2500 m on either side of the fault zone (Fig. 1). Experimental details are described elsewhere [22]. Air temperature and rainfall were measured at the meteorological station Bizeljsko, approximately 14 km from the boreholes. Data recorded are shown in Fig. 2. In this paper, only data collected from stations 1 (Krško-1) are evaluated because they have the longest time series.

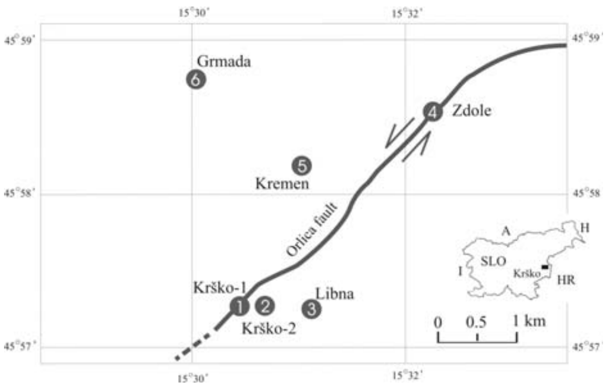

Figure 1. Map of the Krško basin with locations of radon monitoring stations at the Orlica fault with strike-slip displacement. The insert shows the position of Krško (SLO - Slovenia, I - Italy, A - Austria, H - Hungary, HR - Croatia).

3 RESULTS AND DISCUSSION

Since radon concentration is a numerical variable, we have approached the task of predicting radon concentration from meteorological data using regression (or function approximation) methods. We used regression trees [32], as implemented with the WEKA data mining suite [33]. Details of our procedure are described elsewhere [30].

Regression trees are a representation for piece-wise constants or piece-wise linear functions. Like classical regression equations, they predict the value of a dependent variable (called class) from the values of a set of independent variables (called attributes). Data presented in the form of a table can be used to learn or automatically construct a regression tree. In this table, each row (example) has the form $(x_1, x_2,..., x_N, y)$, where x_i are values of the N attributes (e.g., air temperature, barometric pressure, etc.) and y is the value of the class (e.g., radon concentration in soil gas). Unlike classical regression approaches, which find a single equation for a given set of data, regression trees partition the space of examples into axis-parallel rectangles and fit a model to each of these partitions. A regression tree has a test in each inner node that tests the value of a certain attribute and, in each leaf, a model for predicting the class. The model can be a linear equation or just a constant. Trees having linear equations in the leaves are also called model trees (MT).

A number of systems exist for inducing regression trees from examples, such as CART [32] and M5 [34]. M5 is one of the best known programs for regression tree induction. We used the system M5' [35], a reimplementation of M5 within the WEKA data mining suite [33]. The parameters of M5' were set to their default values, unless stated otherwise.

To test the hypothesis that the predictability of radon concentration in periods with seismic activities is worse than in periods without seismic activities, the following procedure was applied. Firstly, the value of the class - daily radon concentration, the values of attributes - daily average barometric pressure, daily average air temperature, daily average soil temperature, difference between daily soil and daily air temperatures, daily amount of rainfall, and difference in daily barometric pressure were selected. The difference between the pressures on day i +1 and day i is related to day i. Secondly, this dataset was split into two parts. In the first part (labelled SA and amounting to 23 % of the data), data for the periods with seismic activity were included. As a first estimate, periods of seven days before and after an earthquake were taken. Data for the remaining days were included in the second part, belonging to the seismically non-active periods (labelled non-SA and amounting to 77 %). Then, to evaluate the predictability of radon concentration in the non-SA periods, we estimated the performance of model trees on the non-SA data with cross-validation.

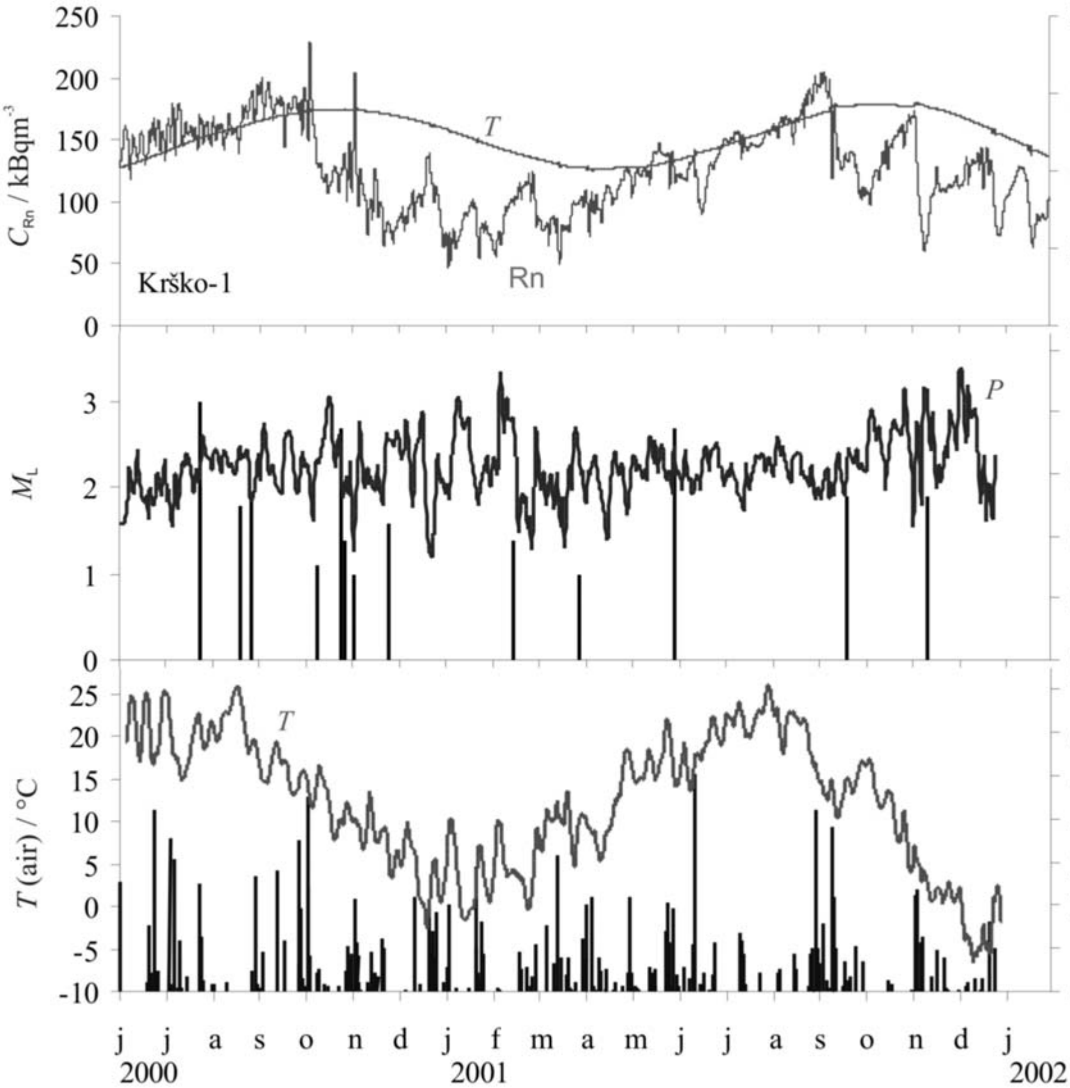

Figure 2. Time run of daily average radon concentration in soil gas and of soil temperature recorded with barasol probes in 60 cm deep boreholes at the Krško-1 station at the Orlica fault in the Krško basin during the period from June 2000 to January 2002. Local earthquakes with R_E/R_D equal to or less than 2 (Dobrovolsky et al., 1979), barometric pressure, air temperature and rainfall at the nearby meteorological station Bizeljsko are also shown.

Furthermore, we induced a model tree on the total non-SA data and measured its performance on the SA data in order to evaluate the practicability of predicting radon concentration in the SA periods. If our hypothesis is true, the first prediction should be better than the second.

In order to facilitate the visualisation of radon anomalies found by this technique, the quantity $\{(C_{Rn})_m/(C_{Rn})_p - 1\}$ was plotted *versus* the time elapsed, as shown in Fig. 3 for selected periods. Here, $(C_{Rn})_m$ is the measured radon concentration and $(C_{Rn})_p$ is the radon concentration predicted with decision trees. In the plots, in addition to the $\{(C_{Rn})_m/(C_{Rn})_p - 1\} = 0$ line, the ± 0.2 regions are indicated by dashed lines. Values falling beyond the dashed lines are considered as anomalies. We see that some earthquakes are preceded and accompanied by radon anomalies (denoted as CA case: *correct anomaly* related to seismic events), some are not (denoted as NA case: *no anomaly* observed for an earthquake), and, also, that there are anomalies during seismically non-active periods (denoted as FA case: *false anomaly* appearing without a seismic event). Sometimes a single, short anomaly appears, but more often swarms of anomalies are observed over longer periods. The duration period of a swarm, also called *total time of anomalies*, is defined as the time from the beginning of the first to the stopping of the last anomaly in the swarm. On the other hand *net time of anomalies* in a swarm is called the sum of duration times of all anomalies in the swarm.

All the anomalies found over the total period of observation are collected in Table 1. Several earthquakes occurring within a few days (such as 14.04.00, 16.04.00 and 17.04.00, 24.08.00 and 31.08.00, 29.10.00, 31.04.00 and 06.11.00) are considered as one seismic event. The area of an anomaly is the area between the – 0.2 and + 0.2 regions and the $(C_{Rn})_m/(C_{Rn})_p - 1$ *versus t* curve. For every earthquake the anomaly was observed at all stations (if in operation), though not at the same time.

Data from Table 1 are summarized in Table 2. The number of CA cases largely outweighs the number of FA cases. The average surface area per anomaly is more than 2-fold greater for CA than for FA. A positive anomaly (+) is one with $\{(C_{Rn})_m/(C_{Rn})_p - 1\} > +0.2$, and negative (–), with $\{(C_{Rn})_m/(C_{Rn})_p - 1\} < -0.2$. For CA, the number of '+' cases is higher than the number of '–'cases. The numbers of '+' and '–' cases for FA are practically the same at all stations. In Table 3, the results for different threshold of $(C_{Rn})_m/(C_{Rn})_p - 1$ are shown, and we see that for $\{(C_{Rn})_m/(C_{Rn})_p - 1\} > \pm 0.2$ optimal results were obtained.

Table 1. Krško-1 station: earthquakes listed with (1) the date of occurring, (2) M_L magnitude, and (3) R_E/R_D value (R_E, distance of the measuring site from the epicentre; R_D, Dobrovolsky's radius (Dobrovolsky et al., 1979)), and radon anomalies defined with $\{(C_{Rn})_m/(C_{Rn})_p - 1\} < -0.2$ and $> +0.2$ ($(C_{Rn})_m$ is the measured radon concentration and $(C_{Rn})_p$ is radon concentration predicted by decision trees, cf. Fig. 3), and characterised by, (4) period of the anomaly, (5) type (CA – *correct anomaly*, FA – *false anomaly*, NA - *no anomaly*), (6) how many days the anomaly appeared before the seismic event, (7) duration of the anomaly in days (net time of anomalies / total time from the start of the first to the end of the last anomaly in the swarm), (8) number of anomalies in a swarm, and (9) surface area of the anomaly in day

earthquakes			radon anomalies					
1	2	3	4	5	6	7	8	9
13.04.99	0.8	1.4	13.04.-16.04.99	CA	1	3/4	2	0.34
20.05.99	0.6	1.6	06.05.-23.05.99	CA	14	12/18	3	1.73
-	-	-	11.08-12.08.99	FA	-	1/1	1	0.02
06.10.99	2.1	2.0	04.10.-10.10.99	CA	2	5/7	2	0.23
-	-	-	21.02.-09.03.00	FA	-	3/18	2	0.04
14.04.00	1.8	1.6						
16.04.00	3.2	0.5	01.04.-11.04.00	CA	13	10/11	2	1.93
17.04.00	2.2	1.2						
28.07.00	3.0	0.4	10.07.-15.07.00	CA	17	4/6	2	0.29
24.08.00	1.8	1.0	25.08.-26.08.00	CA	7	1/1	1	0.01
31.08.00	1.9	2.0						
13.10.00	1.1	1.4	08.10.-12.10.00	CA	5	4/5	2	1.27
29.10.00	2.7	1.5						
31.10.00	1.3	1.0	02.11.-07.11.00	CA	4	4/6	2	1.39
06.11.00	1.0	2.0						
29.11.00	1.6	0.8	15.11.-16.12.00	CA	14	24/32	5	2.60
-	-	-	02.01-12.01.01	FA	-	10/11	2	0.80
19.02.01	1.4	0.4	26.01.-04.03.01	CA	24	15/38	6	1.14
-	-	-	09.03.-29.04.01	FA	-	12/50	8	0.63
04.06.01	2.7	2.0	01.06.-12.06.01	CA	3	2/12	2	0.15
-	-	-	20.06.-24.06.01	FA	-	5/5	1	0.39
25.09.01	1.9	1.4	03.09.-11.10.01	CA	22	18/39	7	3.50
-	-	-	09.11.-28.12.01	FA	-	11/50	5	1.24

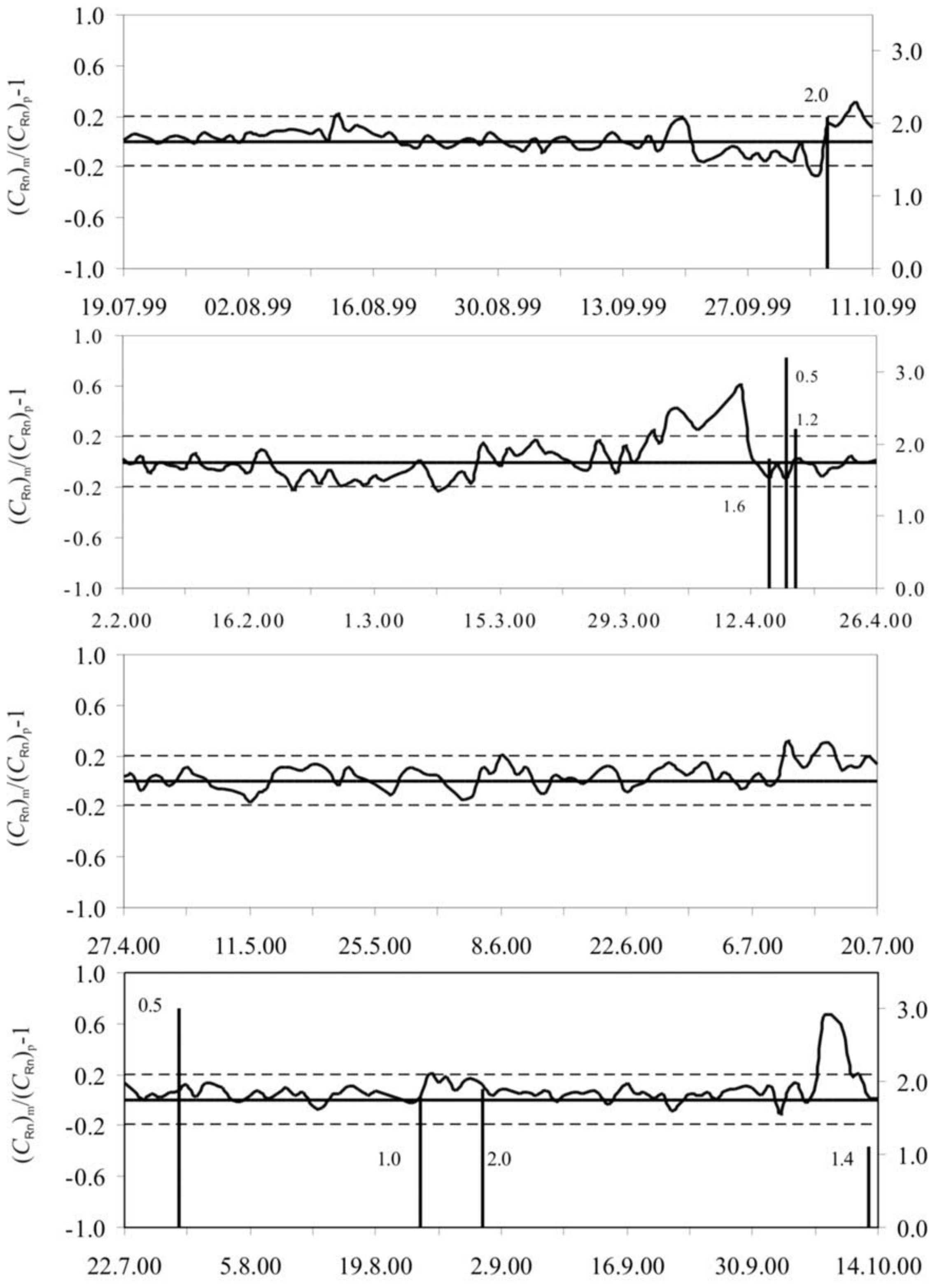

Figure 3. Time run of expression $(C_{Rn})_m/(C_{Rn})_p - 1$ (C_{Rn}, radon concentration in soil gas, m – measured, p - predicted with decision trees) for selected periods at the Krško-1 stations. The solid line is drawn at $\{(C_{Rn})_m/(C_{Rn})_p - 1\} = 0$, and dashed lines at – 0.2 and + 0.2. Numbers attached to the earthquake bars are R_E/R_D values. Radon anomalies are the $(C_{Rn})_m/(C_{Rn})_p - 1$ values outside the – 0.2 and 0.2 regions.

Table 2. Summary of characteristics of anomalies from Table 1.

	CA	FA	NA
total number of anomalies[a]	36/12	19/6	0
total duration of anomalies[b] / d	102/179	42/135	-
average duration time / d	2.83	2.21	-
total surface area of anomalies / d	14.58	3.12	-
average surface area per anomaly / d	0.41	0.16	-
number of '+' anomalies	22	11	-
number of '-' anomalies	14	8	-

[a] number of anomalies / number of swarms
[b] net time of duration of anomalies / total time of duration of swarms

Table 3. Different threshold for $(C_{Rn})_m/(C_{Rn})_p - 1$, marked as A.

Krško-1	A > ± 0,15	A > ± 0,20	A > ± 0,25
PA	12	12	9
LA	8	4	2
NA	0	0	3

4 CONCLUSIONS

The analysis has shown that regression trees are reliable in identifying radon anomalies caused by earthquakes, i.e., radon anomalies were observed for all earthquakes of $R_E/R_D < 1$. Unfortunately, the approach has shown a number of false anomalies (FA), that is, ones not related to seismic events. We hope to reduce this number by including additional environmental parameters such as humidity of soil [36-37], direction and velocity of wind [38], and snow coverage [37], and by extending the time with continued measurements. This will improve the machine learning and hence increase predictability of radon levels. Therefore, a great deal of further effort will be devoted to this approach.

Acknowledgments – The study was funded by the Slovenian Ministry of Education, Science and Sport.

REFERENCES

[1] Kobal, I., Kristan, J., Škofljanec, M., Jerančič, S. and Ančik, M. (1978). Radioactivity of spring and surface waters in the region of the uranium ore deposit at Žirovski vrh. *J. Radioanal. Chem.* 44, 307-315.

[2] Kobal, I. (1979). Radioactivity of thermal and mineral springs in Slovenia. *Health Phys.* 37, 239-242.

[3] Kobal, I. and Renier, A. (1987). Radioactivity of the Atomic spa at Podčetrtek, Slovenia, Yugoslavia. *Health Phys.* 53, 307-310.

[4] Kobal, I. and Fedina, Š. (1987). Radiation doses at the Radenci health resort. *Radiat. Prot. Dosim.* 20, 257-259.

[5] Kobal, I., Vaupotič, J., Mitić, D., Kristan, J., Ančik, M., Jerančič, S. and Škofljanec, M. (1990). Natural radioactivity of fresh waters in Slovenia, Yugoslavia. *Environ. Int.* 16, 141-154.

[6] Vaupotič, J. and Kobal, I. (2001). Radon exposure in Slovenian spas. *Radiat. Prot. Dosim.* 97, 265-270.

[7] Vaupotič, J. (2002). Radon exposure at drinking water supply plants in Slovenia. *Health Phys.* 83, 901-906.

[8] Ulomov, V.I. and Mavashev, B.Z. (1971). Forerunners of the Tashkent earthquake. *Izv. Akad. Nauk Uzb. SSR*, 188-200.

[9] King, C.Y. (1978). Radon emanation on San Andreas fault. *Nature* 271, 516-519.

[10] Ui, H., Moriuchi, H., Takemura, Y., Tsuchida, H., Fujii I. and Nakamura, M. (1988). Anomalously high radon discharge from the Atotsugawa fault prior to the western Nagano Prefecture earthquake (M 6.8) of September 14, 1984. *Tectonophys.* 152, 147-152.

[11] Ohno, M. and Wakita, H. (1996). Coseismic radon changes of the 1995 Hyogo-ken Nanbu earthquake. *J. Phys. Earth* 44, 391-395.

[12] Planinić, J., Radolić, V. and Čulo, D. (2000). Searching for an earthquake precursors: temporal variations of radon in soil and water. *Fizika B* (Zagreb) 9, 75-82.

[13] Planinić, J., Radolić, V. and Lazanin, Ž. (2001). Temporal variations of radon in soil related to earthquakes. *Appl. Radiat. Isot.* 55, 267-272.

[14] Virk, H.S., Walia, V. and Kumar, N. (2001). Helium/radon precursory anomalies of Chamoli earthquake, Garhwal Himalaya, India. *J. Geodyn.* 31, 201-210.

[15] Yasuoka, Y. and Shinnogi, M. (1997). Anomaly atmospheric radon concentration: a posible precursor of the 1995 Kobe, Japan, earthquake. *Helth Phys.* 72, 759-761.

[16] Zmazek, B., Vaupotič, J., Živčić, M., Premru, U. and Kobal, I. (2000). Radon monitoring for earthquake prediction in Slovenia. *Fizika B* (Zagreb) 9, 111-118.

[17] Zmazek, B., Vaupotič, J., Živčić, M., Martinelli, G., Italiano, F. and Kobal, I. (2000). Radon, temperature, electrical conductivity and $^3He/^4He$ measurements in three thermal springs in Slovenia. In: *Book of Abstracts: New Aspects of Radiation Measurements, Dosimetry and Spectrometry*, 2nd Dresden Symposium on Radiation Protection, September 10-14, Dresden, Germany.

[18] Zmazek, B., Italiano, F., Živčić, M., Vaupotič, J., Kobal, I. and Martinelli, G. (2002). Geochemical monitoring of thermal waters in Slovenia: relationships to seismic activity. *Appl. Radiat. Isot.* 57, 919-930.

[19] Zmazek, B., Vaupotič, J. and Kobal, I. (2002). Radon, temperature and electric conductivity in Slovenian thermal waters as potential earthquake precursors. In: *Book of Abstracts: 1st Workshop on Natural Radionuclides in Hydrology and Hydrogeology*, Centre Universitaire de Luxembourg, September 4-7.

[20] Zmazek, B. (2004). *Influence of Seismic Activity on Radon Levels in Thermal Waters and Soil Gas at Selected Sites in Slovenia.* PhD Thesis, University of Maribor, Faculty for Civil Engineering.

[21] Zmazek, B., Vaupotič, J., Bidovec, M., Poljak, M., Živčić, M., Pineau, J. F. and Kobal, I. (2000). Radon monitoring in soil gas tectonic faults in the Krško basin. In: *Book of Abstracts: New Aspects of Radiation Measurements, Dosimetry and Spectrometry*, 2nd Dresden Symposium on Radiation Protection, September 10-14, Dresden, Germany.

[22] Zmazek, B., Živčić, M., Vaupotič, J., Bidovec, M., Poljak, M. and Kobal, I. (2002). Soil radon monitoring in the Krško basin, Slovenia. *Appl. Radiat. Isot.* 56, 649-657.

[23] Di Bello, G., Ragosta, M., Heinicke, J., Koch, U., Lapenna, V., Piscitelli, S., Macchiato, M. and Martinelli, G. (1998). Time dynamics of background noise in geoelectrical and geochemical signals: an application in a seismic area of Southern Italy. *Il Nuovo Cimento* 6, 609-629.

[24] Cuomo, V., Di Bello, G., Lapenna, V., Piscitelli, S., Telesca, L., Macchiato M. and Serio, C. (2000). Robust statistical methods to discriminate extreme events in geoeletrical precursory signals: implications with earthquake prediction. *Nat. Hazard* 21, 247-261.

[25] Biagi, P.F., Ermini, A., Kingsley, S.P., Khatkevich, Y.M. and Gordeev, E.I. (2001). Difficulties with interpreting changes in groundwater gas content as earthquake precursors in Kamchatka, Russia. *J. Seismol.* 5, 487-497.

[26] Belyaev, A.A. (2001). Specific features of radon earthquake precursors. *Geochem. Int.* 12, 1245-1250.

[27] Negarestani, A., Setayeshi, S., Ghannadi-Maragheh, M. and Akashe, B. (2001). Layered neural networks based analysis of radon concentration and environmental parameters in earthquake prediction. *J. Environ. Radioact.* 62, 225-233.

[28] Planinić, J., Vuković, B., Radolić, V., Faj, Z. and Stanić, D. (2003). Deterministic chaos in radon time variations. In: *Proceeding of the 5th Symposium of the Croatian Radiation Protection Association*, HDZZ-CRPA, Zagreb, 349-354.

[29] Steinitz, G., Begin, Z.B. and Gazit-Yaari, N. (2003). Statistically significant relation between radon flux and weak earthquakes in Dead Sea rift valley. *Geology* 31, 505-508.

[30] Zmazek, B., Todorovski, L., Džeroski, S., Vaupotič, J. and Kobal, I. (2003). Application of decision trees to the analysis of soil radon data for earthquake prediction. *Appl. Radiat. Isot.* 58, 697-706.

[31] Dobrovolsky, I.P., Zubkov, S.I. and Miachkin, V.I. (1979). Estimation of the size of earthquake preparation zones. *Pure Appl. Geophys.* 117, 1025-1044.

[32] Breiman, L., Friedman, J.H., Olshen, R.A and Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont.

[33] Witten, I.H and Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco.

[34] Quinlan, J.R. (1992). Learning with continuous classes, In: *Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence*, *World Scientific*, Singapore, 343-348.

[35] Wang, Y. and Witten, I.H. (1997). Induction of model trees for predicting continuous lasses. In: *Proceedings of the Poster Papers of the European Conference on Machine Learning*, University of Economics, Faculty of Informatics and Statistics, Prague.

[36] Ioannides, K.G., Papachristodoulou, C., Karamanis, D.T., Stamoulis, K.C. and Mertzimekis, T.J. (1996). Measurements of ^{222}Rn migration in soil. *J. Radioanal. Nucl. Chem.* 208, 541-547.

[37] Fujiyoshi, R., Morimoto, H. and Sawamura, S. (2002). Investigation of soil radon variation during the winter months in Sapporo, Japan. *Chemosphere* 47, 369-373.

[38] Riley, W.J., Gadgil, A.J. and Nazaroff, W.W. (1996). Wind-inudced ground-surface pressures around a single-family house. *J. Wind Engin. Ind. Aerodyn*. 61, 153-167.