Available online at www.sciencedirect.com
SCIENCE DIRECT
Applied Radiation and Isotopes 64 (2006) 725–734
Applied Radiation and Isotopes
www.elsevier.com/locate/apradiso

Radon in a thermal spring: Identification of anomalies related to seismic activity

B. Zmazek[a,*], L. Todorovski[a], M. Živčić[a,b], S. Džeroski[a], J. Vaupotič[a], I. Kobal[a]

[a]*Jožef Stefan Institute, 1000 Ljubljana, Slovenia*
[b]*Office of Seismology, Environmental Agency of the Republic of Slovenia, 1000 Ljubljana, Slovenia*

Received 16 January 2005; accepted 2 December 2005

Abstract

Anomalies have been observed in the radon content of thermal spring water at the Italian–Slovenian border. To distinguish the anomalies caused by environmental parameters (air and water temperature, barometric and hydrostatic pressure, rainfall) from those ascribed solely to earthquakes with M_L from 1.2 to 2.5 and epicentres, R_E, within $2R_D$ (R_D—Dobrovolsky's radius), two approaches have been used: (i) correlation between time gradients of radon concentration and hydrostatic pressure, and (ii) regression trees within machine learning programs. The regression trees approach has been improved by introducing additional environmental parameters and prolonging the measuring period.

Keywords: Radon in thermal water; Anomalies; Environmental parameters; Earthquakes; Correlation; Regression trees; Forecasting

1. Introduction

Radon (^{222}Rn) can be transported effectively from deep layers of the Earth to the surface by carrier gases and by water (Kristiansson and Malmqvist, 1982; Etiope and Martinelli, 2002). This transport is affected by phenomena accompanying seismic events (Ulomov and Mavashev, 1971; Scholz et al., 1973; Mjachkin et al., 1975; King, 1978; Ui et al., 1988; Ohno and Wakita, 1996; Pulinets et al., 1997; Touten and Baubron, 1999; Planinić et al., 2000; Belyaev, 2001; Virk and Walia, 2001). If radon is therefore monitored at a thermal water spring, shortly before or during an earthquake, an anomaly, i.e. a sudden increase or decrease in radon level, may be observed (Teng, 1980; Heinecke et al., 1995; Virk and Singh, 1995; Fedeli et al., 2001; Heinecke et al., 1997; Virk and Sharma, 1997; Singh et al., 1999; Italiano et al., 2001; Virk and Walia, 2001). Thermal springs and ground waters in Slovenia have been systematically surveyed for radon (Kobal et al., 1978; Kobal, 1979; Kobal and Renier, 1987; Kobal and Fedina, 1987; Kobal et al., 1990; Vaupotič and Kobal, 2001; Vaupotič, 2002; Popit et al., 2002). In a study in 1982 aimed at forecasting earthquakes, radon concentration was determined weekly in four thermal springs, and Cl^-, SO_4^{2-}, hardness and pH, monthly (Zmazek et al., 2000a). In 1998, the study was extended to other thermal springs (Zmazek et al., 2000b, 2002a, b) as well as to soil gas (Zmazek et al., 2000c, 2002c; Zmazek et al., 2003; Zmazek et al., 2004) at selected seismically relevant sites, and the sampling frequency was increased to once an hour. We shall focus here on radon anomalies in a thermal spring.

As a general practice (Yasuoka and Shinogi, 1997; Singh et al., 1999; Virk and Walia, 2001), we have tried to identify anomalies in radon concentration and then relate them to seismic activity (Zmazek et al., 2002a). While some anomalies may be observed easily, it is most often impossible to distinguish an anomaly caused solely by a seismic event, from one related to meteorological or hydrological parameters. This distinction was additionally obscured by the rapid change of the water level above the spring, caused by the flooding of the Tolminka river. The implementation of more advanced statistical methods for data evaluation (Di Bello et al., 1998; Cuomo et al., 2000; Biagi et al., 2001; Belyaev, 2001; Negarestani et al., 2001; Planinić et al., 2003; Steinitz et al., 2003, Planinić et al.,

*Corresponding author. Tel.: +386 1 477 35 80; fax: +386 1 477 38 11.
E-mail address: boris.zmazek@ijs.si (B. Zmazek).

0969-8043/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apradiso.2005.12.016

2004) was expected to improve the results. In this paper, we have applied, for the first time, regression trees for earthquake prediction, based on radon content in a thermal spring. Data mining and machine learning methods used for that purpose have already been applied successfully to many environmental problems, as reviewed by Džeroski (2002), and also to predicting radon levels in soil gas (Zmazek et al., 2003, 2004). In this paper two approaches will be used. Firstly, radon anomalies will be sought on the basis of the relationship between the time gradients of radon concentration and hydrostatic pressure. Then radon concentration will then be predicted from meteorological and hydrological data during seismically non-active periods, followed by testing the hypothesis that the prediction is significantly worse for seismically active (SA) than for seismically non-active periods.

2. Experiment

Experiments are described in detail elsewhere (Zmazek et al., 2000c, 2002b, c). The thermal spring is located in a SA area at the Slovenian–Italian border in the Devil's Bridge Cave near Zatolmin village (Fig. 1). The spring, whose average temperature was $20.6 \pm 1.0\,^{\circ}C$ during the experiment, is at the bottom of a small pool in a cave, which is connected to the nearby Tolminka river. Since the end of 1999, radon concentration, temperature and pressure (hydrostatic pressure, which is related to the water level above the spring) in the water at the spring have been measured and recorded once per hour with a barasol probe (MC-450, ALGADE, France). Meteorological data, such as air temperature, barometric pressure, rainfall and water level of the Tolminka river 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. Using Dobrovolsky's equation, $R_D = 10^{0.43M}$, (Dobrovolsky et al., 1979), the radius R_D of the zone within which precursory phenomena may be manifested (so-called Dobrovolsky's radius) was calculated, M being the magnitude. The earthquakes for which the distance R_E between the epicentre and the thermal spring was equal to or less than $2R_D$ have been taken into account.

3. Methodology of data analysis

Experimental data have been analysed by searching for radon anomalies (i) expressed as time intervals when time gradients of hydrostatic pressure and radon concentration have the same sign, and (ii) by using regression trees.

3.1. Dependence of radon concentration on hydrostatic pressure

Hydrostatic pressure, as measured with the barasol probe at the spring, is equivalent to the water level in the pool. The relationship between radon concentration in the water, C_{Rn}, and the water level, h, is very complex. The overflow of the Tolminka river increases water level and decreases the temperature of the water in the pool. This enhances the dissolution of radon gas in the water. On the other hand, radon in the pool is diluted by river water, reducing C_{Rn}. In addition, an increase in h may reduce the flow rate of the spring and thus reduce the inflow of radon into the pool. Therefore, in this study we considered the periods when the time gradient of radon concentration in thermal water, $\Delta C_{Rn}/\Delta t$, and the time gradient of water level, $\Delta h/\Delta t$, have simultaneously the same sign, as radon anomalies, attributed to the seismic activity and not to environmental parameters.

3.2. Regression trees

We used regression trees (Breiman et al., 1984), as implemented with the WEKA data mining suite (Witten

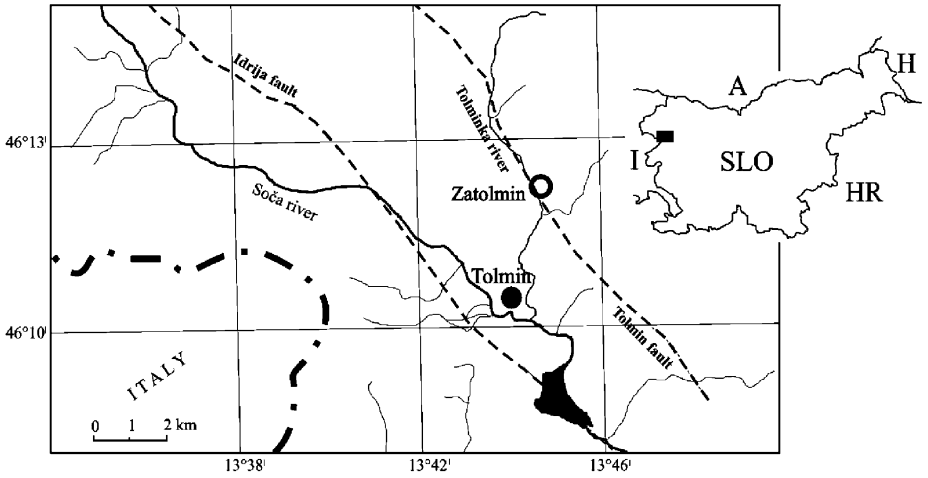

Fig. 1. Map of the region with the location of the Devil's Bridge Cave near Zatolmin. Shown also are the Idrija and Tolmin faults, and the Soča and Tolminka rivers. The inset shows the position of this region (SLO—Slovenia, I—Italy, A—Austria, H—Hungary, HR—Croatia).

and Frank, 1999). Details of our approach are described elsewhere (Zmazek et al., 2003).

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 that 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 and water temperature, hydrostatic pressure, etc.) and y is the value of the class (e.g., radon concentration in water). 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).

The predictive performance of the regression methods was determined using two different measures, the correlation coefficient and the root mean squared error (RMSE). The correlation coefficient (r) expresses the level of correlation between the measured and predicted values of radon concentration, higher values of the correlation coefficient signify a better correlation. The RMSE measures the discrepancy between measured and predicted values of radon concentration, smaller RMSE values indicate lower discrepancies.

In order to estimate the performance of predictors on measurements that were not used for training the predictor, a standard 10-fold cross validation method was applied. We shall base our analysis on the hypothesis that, during SA periods, r will be lower and RMSE higher than during seismically non-active (non-SA) periods.

4. Results and discussion

Experimental results for the thermal spring at Zatolmin are shown in Fig. 2. In the following sections, these raw data will be analysed by applying (i) time gradients of hydrostatic pressure and radon concentration, and (ii) regression trees. Some earthquakes may be 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); there are also 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 a number of anomalies over longer periods have been observed, hereafter called swarms of anomalies. From the point of view of earthquake forecasting, only anomalies starting day or more before occurrence of an earthquake are relevant. 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 end of the last anomaly in the swarm. The *net time of anomalies* in a swarm denotes the sum of duration times of all anomalies in the swarm.

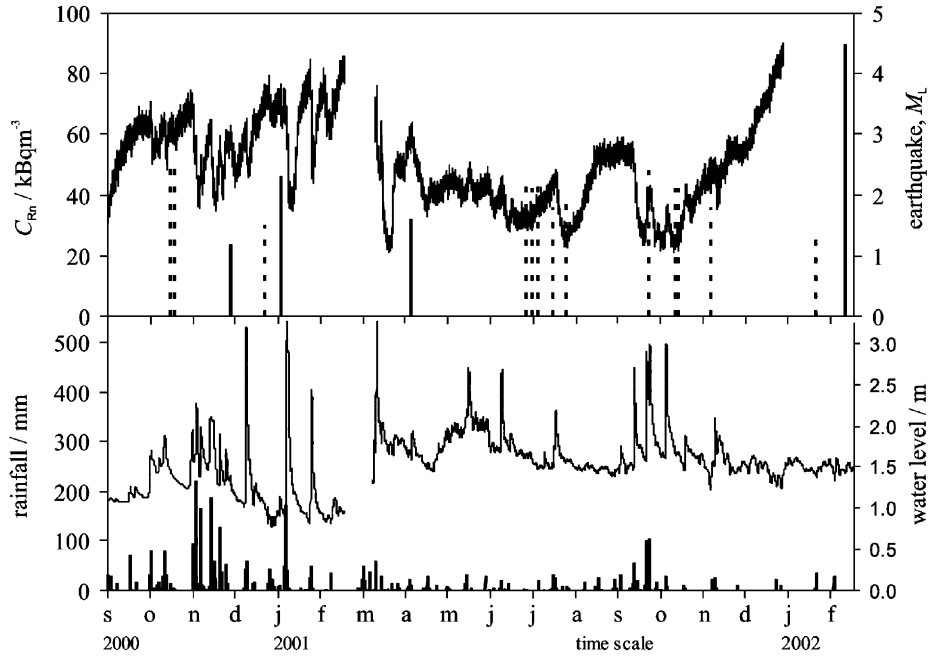

Fig. 2. Temporal variations of radon concentration in thermal spring at the Devil's Bridge Cave, earthquake events (solid lines, $R_E/R_D \leqslant 1$, dashed lines, $1 < R_E/R_D \leqslant 2$), level of the Tolminka river and rainfall for September 2000–February 2002.

4.1. Anomalies based on the correlation between $\Delta C_{Rn}/\Delta t$ and $\Delta h/\Delta t$

From the data of the measured radon concentration in water, C_{Rn}, and the height of the water level in the pool above the barasol probe, h, time gradients $\Delta C_{Rn}/\Delta t$ and $\Delta h/\Delta t$ were calculated. Values of 1, 2, 6, 12 and 24 h have been taken for Δt. Because of radon data fluctuations, values $\Delta t < 6$ h are not used, while $\Delta t = 24$ h appeared to be optimal and has been eventually used in this paper. Time intervals with both gradients having the same sign, either positive or negative, i.e., $(\Delta C_{Rn}/\Delta t/C_{Rn})/(\Delta h/\Delta t/h) > 0$, were considered as radon anomalies, possibly related to seismic activity (Zmazek et al., 2002c). Taking the same value of Δt for both gradients $(\Delta C_{Rn}/\Delta t/C_{Rn})/(\Delta h/\Delta t/h) = (\Delta C_{Rn}/C_{Rn})/(\Delta h/h)$. In addition, threshold values for both gradients were required. Thus, if a positive $(\Delta C_{Rn}/C_{Rn})/(\Delta h/h)$ value was observed below the threshold fixed at $(\Delta C_{Rn}/\Delta t)/C_{Rn} < 0.1\,\mathrm{d}^{-1}$ and $\Delta h/\Delta t < 2\,\mathrm{cm\,d}^{-1}$, a negative sign was assigned to it and it was not considered as an anomaly. Values of $(\Delta C_{Rn}/C_{Rn})/(\Delta h/h)$ thus obtained are plotted versus elapsed time in Fig. 3. In the case of several earthquakes occurring on the same day the magnitude of the largest is given. A series of several earthquakes occurring within a few days (such as those on 18.04 and 26.04.2000; 15.10 and 18.10.2000; 29.06, 03.07 and 07.07.2001; 15.10 and 17.10.2001) are considered as one seismic event.

Anomalies were observed (Table 1) for all earthquakes (CA cases) except for that on 18.07.2000 (NA case). In fact, this has magnitude $M_L = 1.8$ and $R_E/R_D = 1.8$, and hence failure to forecast it does not appear to be critical. The CA anomaly on 17.–18.05.2001 is presumably a precursor for the earthquake on 18.07.2001. A swarm of FA anomalies was observed between the two earthquakes on 26.04.2000 and 16.08.2000. We have tried to reduce the number of FA cases by augmenting the threshold of $\Delta h/\Delta t$ from 2 to $5\,\mathrm{cm\,d}^{-1}$ and to $10\,\mathrm{cm\,d}^{-1}$. It is evident from Table 2, which also summarises the main findings from Table 1, that increasing this threshold results in decreasing the number of FA anomalies but, at the same time, it decreases the number of CA and increases the number of NA cases. Thus, with $\Delta h/\Delta t = 10\,\mathrm{cm\,d}^{-1}$, anomalies were observed for only 8 earthquakes and not for the remaining 7. This is undesirable from the earthquake forecasting point of view, so we have mostly based our analysis on $\Delta h/\Delta t > 2\,\mathrm{cm\,d}^{-1}$ (Tables 1 and 2).

Fig. 4 presents how (a) time delay between the appearance of the anomaly and occurrence of the earthquake, (b) net duration time of anomalies, (c) total duration time of anomalies and (d) number of anomalies in the swarm all depend on the R_E/R_D ratio. Though with widely scattered points, all plots show negative trends with increasing R_E/R_D, as expected.

Table 2 also shows that the average duration time is longer for FA than for CA anomalies, and is not much dependent on the $\Delta h/\Delta t$ threshold. For CA, the number of

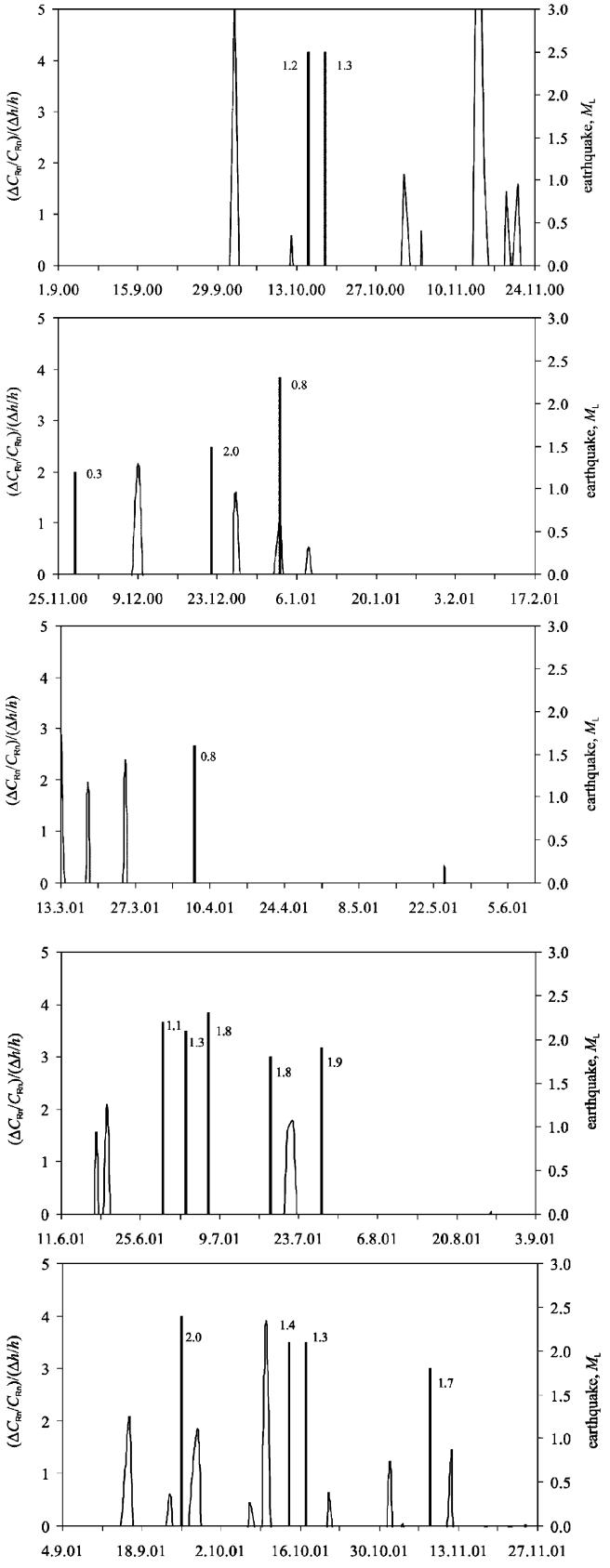

Fig. 3. Time run of $(\Delta C_{Rn}/C_{Rn})/(\Delta h/h)$ at the thresholds fixed at $(\Delta C_{Rn}/\Delta t)/C_{Rn} > 0.1\,\mathrm{d}^{-1}$ and $\Delta h/\Delta t > 2\,\mathrm{cm\,d}^{-1}$, for selected periods. Also earthquakes are shown by bold vertical lines, the attached numbers are R_E/R_D.

Table 1
Gradients method: 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 $(\Delta C_{Rn}/C_{Rn})/(\Delta h/h)>0$ under condition $(\Delta C_{Rn}/\Delta t)/C_{Rn}>0.1\,d^{-1}$ and $\Delta h/\Delta t>2\,cm\,d^{-1}$, and characterised by: (4) period of the anomaly, (5) type (CA—correct anomaly, FA—false anomaly, NA—no anomaly), (6) start time (how many days the anomaly appeared before the seismic event), (7) duration of the anomaly swarm 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

Earthquakes (EQ)			Radon anomalies				
1 Date	2 M_L	3 R_E/R_D	4 Duration period	5 Type	6 Start time before EQ/d	7 Duration time/d	8 Number of anomalies
15.10.2000	2.5	1.2	02.10.–12.10.2000	CA	13	2/11	2
18.10.2000	2.5	1.3					
—	—	—	01.11.–04.11.2000	FA	—	2/4	2
28.11.2000	1.2	0.3	14.11.–21.11.2000	CA	14	4/8	3
22.12.2000	1.5	2.0	09.12.–10.12.2000	CA	13	1/1	1
03.01.2001	2.3	0.8	03.01.–08.01.2001	CA	1	2/6	3
07.04.2001	1.6	0.8	13.03–25.03.2001	CA	25	3/13	3
29.06.2001	2.2	1.1	17.06.–19.06.2001	CA	12	2/3	2
03.07.2001	2.1	1.3					
07.07.2001	2.3	1.8					
18.07.2001	1.8	1.8	—	NA	—	—	—
27.07.2001	1.9	1.9	21.07.–22.07.2001	CA	6	2/2	1
25.09.2001	2.4	2.0	15.09.–28.09.2001	CA	10	5/14	3
14.10.2001	2.1	1.4	07.10.–10.10.2001	CA	7	2/4	2
17.10.2001	2.1	1.3					
—	—	—	21.10.–22.10.201	FA	—	1/1	1
08.11.2001	1.8	1.7	01.11.–12.11.2001	CA	7	2/12	2

Table 2
Summary of main features from Table 1, and their comparison for different values of the $\Delta h/\Delta t$ ($cm\,d^{-1}$) threshold

	$\Delta h/\Delta t>2$			$\Delta h/\Delta t>5$			$\Delta h/\Delta t>10$		
	CA	FA	NA	CA	FA	NA	CA	FA	NA
Total number of anomalies[a]	22/10	3/2	—/1	18/6	3/2	—/5	10/4	3/2	—/7
Total duration of anomalies[b]/d	26/74	3/5	—	20/66	3/5	—	11/34	3/5	—
Average duration time/d	1.18	1.0	—	1.11	1.0	—	1.10	1.0	—
Number of '+' anomalies	10	1	—	8	1	—	5	1	—
Number of '−' anomalies	13	2	—	10	2	—	5	2	—

[a]Number of anomalies/number of swarms.
[b]Net time of duration of anomalies/total time of duration of swarms.

'−'anomalies (when both gradients are negative) always outweighs the number of '+' anomalies (when both gradients are positive) while for FA, the '−'/'+' ratio decreases with increasing $\Delta h/\Delta t$.

4.2. Anomalies found with regression trees

The following procedure was applied. Firstly, the value of the class, i.e., daily radon concentration in water, and the values of attributes, i.e., average daily height of water level, average daily air and water temperatures and daily amount of rainfall were selected. Secondly, this dataset was split into two parts. As in our work on soil gas radon (Džeroski et al., 2003a, b; Zmazek et al., 2003, 2004), in the first part, data for the periods with seismic activity (labelled SA) were included, i.e., periods of 7 days before and after an earthquake. Data for the remaining days were included in the second part, belonging to the seismically non-active periods (labelled non-SA). Then, to evaluate the predictability of radon concentration in the seismically non-active periods, we estimated the performance of MT on the non-SA data with cross-validation. Furthermore, we induced a model tree on the whole 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 measured performance should be higher than the second.

In order to facilitate the visualisation of radon anomalies found by this technique, we plotted the value of $(C_{Rn})_m/(C_{Rn})_p-1$ Versus the time elapsed, as shown in Fig. 5. Here, $(C_{Rn})_m$ is the measured radon concentration and $(C_{Rn})_p$ is the radon concentration predicted with MT. 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. These are threshold

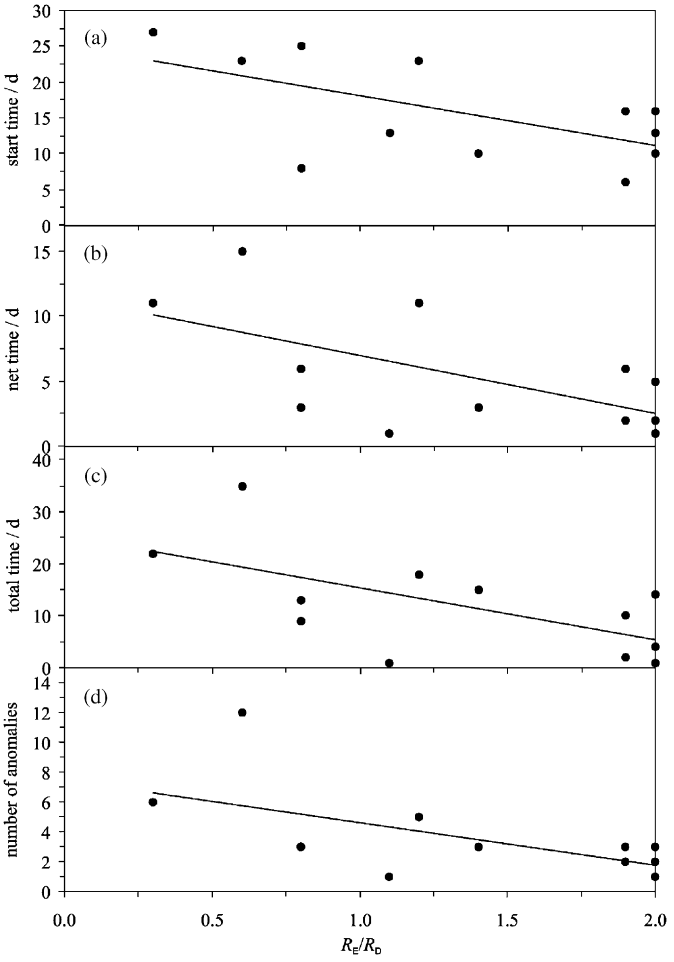

Fig. 4. Gradients method: dependence of (a) time delay between start time of anomalies and earthquake occurrence, (b) net duration time of anomalies, (c) total duration time of anomalies, and (d) number of anomalies, on R_E/R_D.

regions and values of $(C_{Rn})_m/(C_{Rn})_p$–1 falling outside them were considered as anomalies. All the anomalies found are collected in Table 3. In the case of several earthquakes occurring on the same day the magnitude of the largest is given. Also here, a series of earthquakes occurring within a few days (such as those on 15.10 and 18.10.2000; 23.12.2000 and 04.01.2001; 29.06, 03.07. and 07.07.2001; 18.07 and 28.07.2001; 15.10 and 17.10.2001) have been considered as one seismic event. The area of an anomaly is the area between the threshold line at −0.2 or +0.2 and the $(C_{Rn})_m/(C_{Rn})_p$–1 Versus *time* curve. Results are collected in Table 3. With this approach, no anomaly was observed for the earthquake with $M_L = 1.8$ and $R_E/R_D = 1.8$ on 18.07.2001, unless we connect this earthquake to the preceding series of earthquakes from 29.06.2001 to 07.07.2001 with the CA anomaly on 17.–19.06.01. The number of FA cases is higher than with the gradient approach. Reducing the threshold of $(C_{Rn})_m/(C_{Rn})_p$–1 from ±0.20 to ±0.15 does not eliminate the NA event on 18.07.2001, as seen in Table 4. Table 4 also shows that, on increasing the threshold of $(C_{Rn})_m/(C_{Rn})_p$–1 from

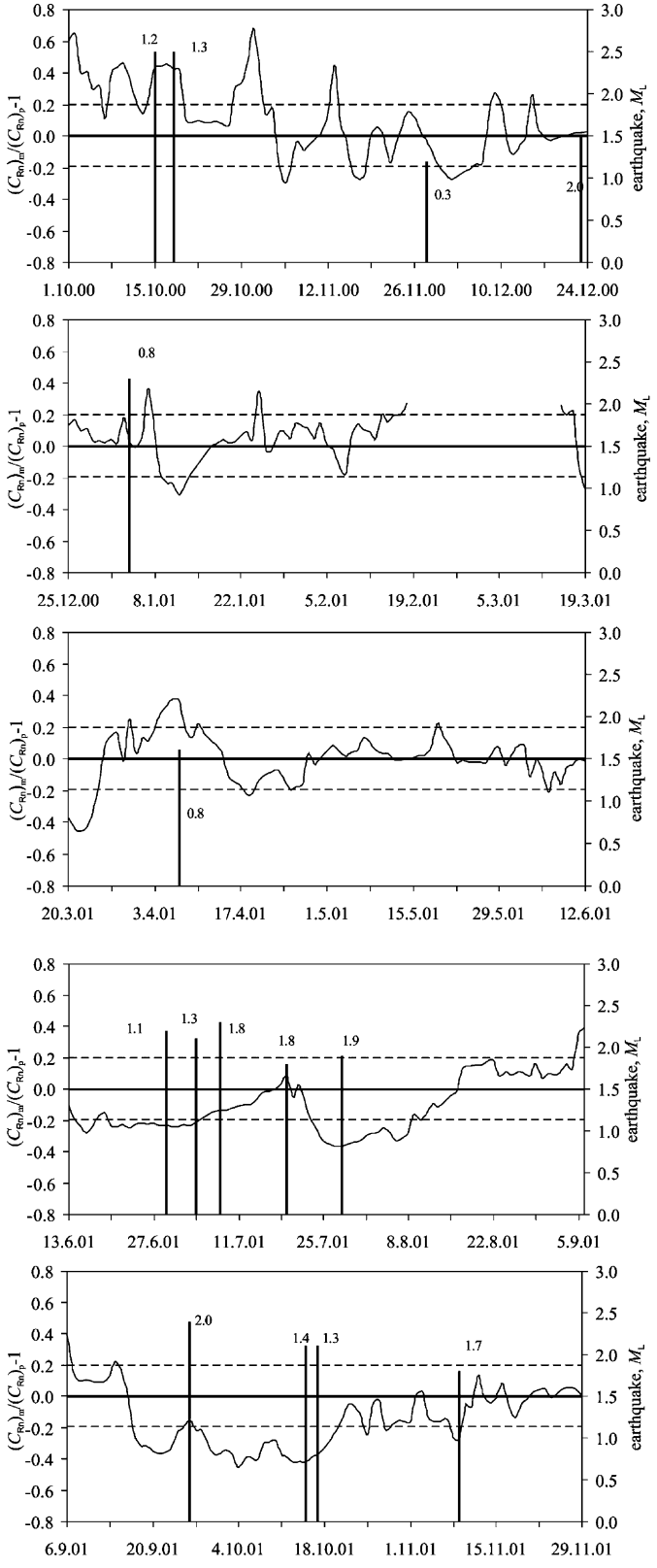

Fig. 5. Time run of $(C_{Rn})_m/(C_{Rn})_p$–1 for selected periods. The solid line is drawn at $(C_{Rn})_m/(C_{Rn})_p - 1 = 0$, and dashed lines at –0.2 and 0.2 Numbers attached to the earthquake bars are R_E/R_D. Radon anomalies are the $(C_{Rn})_m/(C_{Rn})_p$–1 values outside the −0.2 and 0.2 regions.

Table 3
Trees method: 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 $((C_{Rn})_m/(C_{Rn})_p-1)<-0.2$ or >0.2 ($(C_{Rn})_m$ is the measured radon concentration and $(C_{Rn})_p$ is radon concentration predicted by decision trees), 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, (9) surface area of the anomaly, i.e., the area between the –0.2 or 0.2 line and the $((C_{Rn})_m/(C_{Rn})_p-1)$—time curve

Earthquakes (EQ)			Radon anomalies					
1 Date	2 M_L	3 R_E/R_D	4 Duration period	5 Type	6 Start time before EQ/d	7 Duration time/d	8 Number of anomalies	9 Surface area/d
15.10.2000	2.5	1.2	01.10.–19.10.2000	CA	14	17/19	3	3.50
18.10.2000	2.5	1.3						
28.11.2000	1.2	0.3	28.10.–05.12.2000	CA	31	15/39	5	1.84
22.12.2000	1.5	2.0	09.12.–13.01.2001	CA	14	7/36	4	0.53
03.01.2001	2.3	0.8						
—	—	—	25.01.–18.02.2001	FA	—	4/25	3	0.24
07.04.2001	1.6	0.8	14.03.–19.04.2001	CA	24	18/37	7	2.20
—	—	—	19.05.–06.06.2001	FA	—	2/19	2	0.03
29.06.2001	2.2	1.1	15.06.–04.07.2001	CA	14	18/20	2	0.58
03.07.2001	2.1	1.3						
07.07.2001	2.3	1.8						
18.07.2001	1.8	1.8	—	NA	—	—	—	—
27.07.2001	1.9	1.8	24.07.–08.08.2001	CA	4	16/16	1	1.79
—	—	—	05.09.–06.09.2001	FA	—	2/2	1	0.35
25.09.2001	2.4	2.0	14.09.–25.09.2001	CA	12	10/12	2	1.79
14.10.2001	2.1	1.4	27.09.–20.10.2001	CA	18	24/24	1	3.50
17.10.2001	2.1	1.3						
—	—	—	25.10.–28.10.2001	FA	—	2/4	2	0.06
08.11.2001	1.8	1.8	08.11.–09.11.2001	CA	1	2/2	1	0.13

Table 4
Summary of main features from Table 3, and their comparison for different values of the $(C_{Rn})_m/(C_{Rn})_p$–1 threshold

	$>\pm0.15$			$>\pm0.20$			$>\pm0.25$		
	CA	FA	NA	CA	FA	NA	CA	FA	NA
Total number of anomalies[a]	26/9	8/4	—/1	26/9	8/4	—/1	21/9	3/2	—/1
Total duration of anomalies[b]/d	127/205	10/50	—	127/205	10/50	—	85/161	4/27	—
Average duration time/d	4.88	1.25	—	4.88	1.25	—	4.85	3.0	—
Total surface area of anomalies/d	21.3	2.3	—	21.3	2.3	—	9.8	0,4	—
Average surface area per anomaly/d	0.82	0.28	—	0.82	0.28	—	0.47	0.31	—
Number of '+' anomalies	14	5	—	14	5	—	10	3	—
Number of '−' anomalies	12	3	—	12	3	—	11	0	—

[a]Number of anomalies/number of swarms.
[b]Net time of duration of anomalies/total time of duration of swarms.

±0.20 to ±0.25, the number of FA cases is reduced without the appearance of new NA cases.

In the analysis, those anomalies were also accepted which appeared more than 7 days before an earthquake, provided they continued without interruption, or the interruptions were not longer than 2–3 days. The longest period of continuously appearing anomalies was a month. About 52% of anomalies were of such a kind.

The dependence on the R_E/R_D ratio of (a) time delay between the appearance of the anomaly and occurrence of the earthquake, (b) net duration time of anomalies, (c) total duration time of the swarm and (d) number of anomalies in the swarm is shown in Fig. 6. Though with widely scattered points, all plots show decreasing trends with increasing R_E/R_D, as expected.

Table 4 summarises the main features obtained by regression trees. The number of CA swarms outweighs the number of FA swarms by a factor of about 2. Also, the average duration time and average area of an anomaly is much bigger for CA than for FA. Both for CA and FA, the number of '+' anomalies (when the value of $(C_{Rn})_m/(C_{Rn})_p$–1 is positive) is always higher than the number of '−' anomalies (when the value of $(C_{Rn})_m/(C_{Rn})_p$–1 is negative).

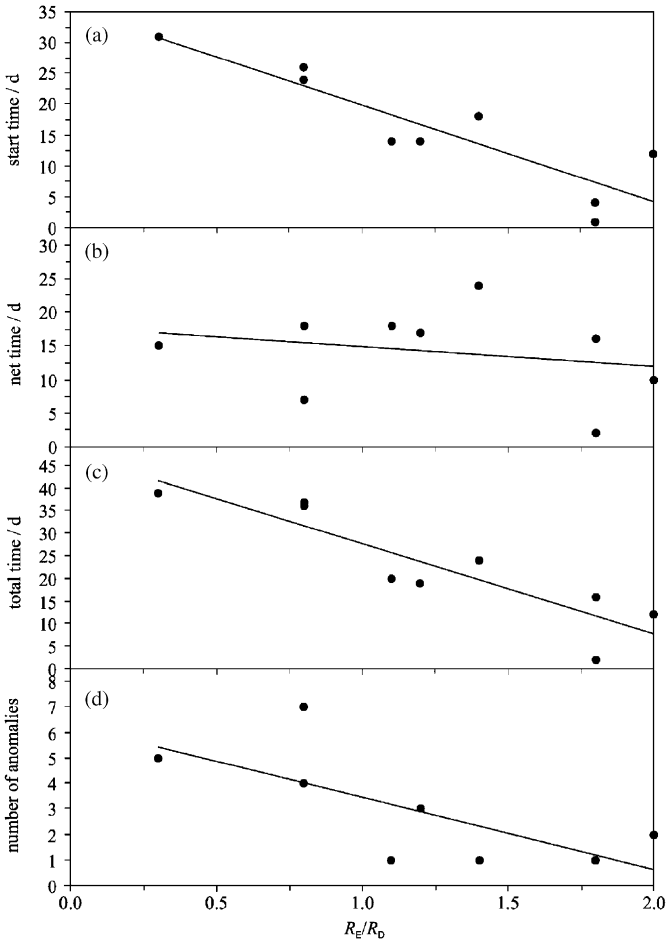

Fig. 6. Trees method: dependence of (a) time delay between start time of anomalies and earthquake occurrence, (b) net duration time of anomalies, (c) total duration time of anomalies, and (d) number of anomalies, on R_E/R_D.

Fig. 7 shows earthquakes and CA and FA anomalies, as observed by gradients at different values of the $\Delta h/\Delta t$ threshold and by regression tress at different values of the $(C_{Rn})_m/(C_{Rn})_p$–1 threshold. For gradients, $\Delta h/\Delta t = 2\,cm\,d^{-1}$ appears to be optimal, while for regression trees, the same results are obtained with $(C_{Rn})_m/(C_{Rn})_p - 1 = \pm 0.15$ or ± 0.20, but with ± 0.25 some of the CA and FA anomalies are lost.

5. Conclusion

The analysis of temporal variation of radon concentration in a thermal spring has shown, by applying both (i) the correlation of time gradients of radon concentration and hydrostatic pressure and (ii) regression trees within the machine-learning programs, radon anomalies caused by environmental parameters can be distinguished from those ascribed to seismic activity for earthquakes with M_L between 1.2 and 2.5. Both approaches have shown radon anomalies for all earthquakes occurring within $R_E/R_D < 1$ (R_D-Dobrovosky's radius), but both approaches failed for an earthquake with $M_L = 1.8$ and $R_E/R_D < 1.8$. A number

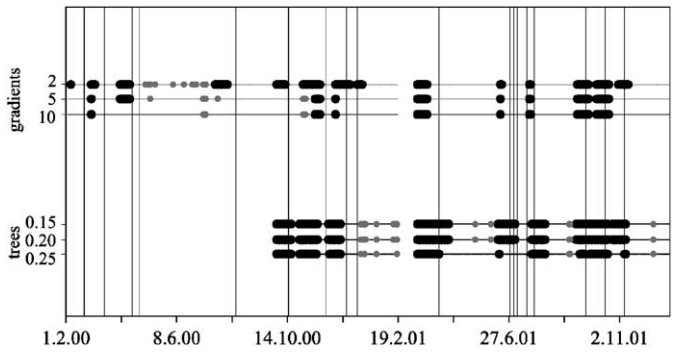

Fig. 7. CA-correct (black point) and FA-false (gray point) anomalies as observed by applying C_{Rn} and P gradients for $\Delta h/\Delta t > 2$, 5 and 10 cm d^{-1}, and regression trees for $(C_{Rn})_m/(C_{Rn})_p$ –1 beyond $\pm$ 0.15, $\pm$0.20 and $\pm$0.25. Vertical lines represent earthquakes with R_E/R_D lower than 2. Horizontal dotted lines represent the periods when the instrument was in operation and no anomalies were observed.

of false anomalies (those not accompanying earthquakes) are also seen. This number could not be effectively reduced in approach (i) because it is based on the assumption that the effect of hydrostatic pressure on radon level overwhelms other environmental parameters, which is often not necessarily the case. The situation is more promising with decision trees. Here, the possibility exists that, by including additional environmental parameters such as barometric pressure and its time gradient, temperatures of air and water and their difference, and by extending the time with further measurements, the machine learning can be improved and hence the number of FA cases reduced. It should be stressed that the earthquakes that occurred during our study were weak ($M_L < 2.5$) and we may expect that both approaches would show more reliable forecasting for higher M_L.

Acknowledgements

The study was funded by the Ministry of Education, Science and Sport of Slovenia. The authors appreciate the technical assistance of the Žabče Divers Club, especially of Mr. Darko Sovdat.

References

Belyaev, A.A., 2001. Specific features of radon earthquake precursors. Geochem. Int. 12, 1245–1250.

Biagi, P.F., Ermini, A., Kingsley, S.P., Khatkevich, Y.M., Gordeev, E.I., 2001. Difficulties with interpreting changes in groundwater gas content as earthquake precursors in Kamchatka, Russia. J. Seismol. 5, 487–497.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont.

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

Di Bello, G., Ragosta, M., Heinicke, J., Koch, U., Lapenna, V., Piscitelli, S., Macchiato, M., 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.

Dobrovolsky, I.P., Zubkov, S.I., Miachkin, V.I., 1979. Estimation of the size of earthquake preparation zones. Pure Appl. Geophys. 117, 1025–1044.

Džeroski, S., 2002. Applications of KDD methods in environmental sciences. In: Kloesgen, W., Zytkow, J. (Eds.), Handbook of Data Mining and Knowledge Discovery. Oxford University Press, Kloesgen.

Džeroski, S., Todorovski, L., Zmazek, B., Vaupotič, J., Kobal, I., 2003a. Modelling soil radon concentration for earthquake prediction. In: Workshop on Mining Scientific and Engineering Datasets, International Conference on Data Mining, Cathedral Hill Hotel, San Francisco, California, May 3, pp. 19–28.

Džeroski, S., Todorovski, L., Zmazek, B., Vaupotič, J., Kobal, I., 2003b. Modelling soil radon concentration for earthquake prediction. In: Grueser, G., Tanaka, Y., Yamamoto, A. (Eds.), Proceedings of the 6th International Conference, Sapporo, Japan, October. Springer, Berlin, pp. 87–99.

Etiope, G., Martinelli, G., 2002. Migration of carrier and trace gases in the geosphere: an overview. Phys. Earth Planet. Int. 129, 185–204.

Fedeli, G., Frondini, F., Italiano, F., Lemmi, M., Martinelli, G., 2001. Radon behaviour in the selected spring of Triponzo during the 1997–1998 Umbria-Marche seismic sequence. In: Hunyadi, I., Csige, I., Hakl, J. (Eds.), Proceedings of the 5th International Conference on Rare Gas Geochemistry, 30 August–3 September, 1999. EP Systema, Debrecen, Hungary, pp. 55–60.

Heinecke, J., Koch, U., Hebert, D., Martinelli, G., 1995. Simoltaneous measurements of radon and CO_2 in water as a possible tool for earthquake prediction. In: Dubois, C. (Ed.), Gas Geochemistry. Univeristy of France, Comté, Science Reviews, pp. 295–303.

Heinecke, J., Koch, U., Martinelli, G., 1997. Radon and CO_2 measurement for earthquake prediction research: status report and model conception about the bad Brambach location (Germany). In: Virk, H.S. (Ed.), Rare Gas Geochemistry. Guru Nanak Dev University, Amritsar, pp. 136–142.

Italiano, F., Martelli, M., Martinelli, G., Paternoster, M., Nuccio, P.M., 2001. Geochemical modelling of earthquake-related anomalies in fluids of Val d'Agri (Southern Italy). Terra Nova 13, 249–257.

King, C.Y., 1978. Radon emanation on San Andreas fault. Nature 271, 516–519.

Kobal, I., 1979. Radioactivity of thermal and mineral springs in Slovenia. Health Phys. 37, 239–242.

Kobal, I., Fedina, Š., 1987. Radiation doses at the Radenci Health Resort. Radiat. Prot. Dosim. 20, 257–259.

Kobal, I., Renier, A., 1987. Radioactivity of the atomic spa at Podčetrtek, Slovenia, Yugoslavia. Health Phys. 53, 307–310.

Kobal, I., Kristan, J., Škofljanec, M., Jerančič, S., 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.

Kobal, I., Vaupotič, J., Mitić, D., Kristan, J., Ančik, M., Jerančič, S., Škofljanec, M., 1990. Natural radioactivity of fresh waters in Slovenia, Yugoslavia. Environ. Int. 16, 141–154.

Kristiansson, K., Malmqvist, L., 1982. Evidence for non-diffusive transport of ^{222}Rn in the ground and new physical model for the transport. Geophysics 47, 1444–1452.

Mjachkin, V.I., Brace, W.E., Sobolev, G.A., Dieterich, J.H., 1975. Two models for earthquake forerunners. Pure Appl. Geophys. 113, 169–181.

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

Ohno, M., Wakita, H., 1996. Coseismic radon changes of the 1995 Hyogoken Nanbu earthquake. J. Phys. Earth 44, 391–395.

Planinić, J., Radolić, V., Čulo, D., 2000. Searching for an earthquake precursor: temporal variations of radon in soil and water. Fizika B (Zagreb) 9, 75–82.

Planinić, J., Vuković, B., Radolić, V., Faj, Z., Stanić, D., 2003. Deterministic chaos in radon time variations. In: Bronić, I.K., Miljanić, S., Obelić, B. (Eds.), Proceeding of the 5th Symposium of the Croatian Radiation Protection Association. HDZZ-CRPA, Zagreb, pp. 349–354.

Planinić, J., Vuković, B., Radolić, V., 2004. Radon time variations and deterministic chaos. J. Environ. Radioact. 75, 25–45.

Popit, A., Urbanc, J., Vaupotič, J., Kobal, I., 2002. Radioactivity survey of waters in the Slovenian Karst region. RMZ—Mat. Geoenviron. 49, 487–497.

Pulinets, S.A., Aleseev, V.A., Legenka, A.D., Khegai, V.V., 1997. Radon and metallic aerosols emanation before strong earthquakes and their role in atmosphere and ionosphere modification. Adv. Space Res. 20, 2173–2176.

Scholz, C.H., Sykes, L.R., Agrawal, Y.P., 1973. Earthquake prediction: a physical basis. Science 181, 803–810.

Singh, M., Kumar, M., Jain, R.K., Chatrath, R.P., 1999. Radon in ground water related to seismic events. Radiat. Meas. 30, 465–469.

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

Teng, T.L., 1980. Some recent studies on groundwater radon content as an earthquake precursor. J. Geophys. Res. 85, 3089–3099.

Toutain, J.P., Baubron, J.C., 1999. Gas geochemistry and seismotectonics: a review. Tectonophys. 304, 1–27.

Ui, H., Moriuchi, H., Takemura, Y., Tsuchida, H., Fujii, I., 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.

Ulomov, V.I., Mavashev, B.Z., 1971. Forerunners of the Tashkent earthquake. Izv. Akad. Nauk Uzb. SSR, 188–200.

Vaupotič, J., Kobal, I., 2001. Radon exposure in Slovenian spas. Radiat. Prot. Dosim. 97, 265–270.

Vaupotič, J., 2002. Radon exposure at drinking water supply plants in Slovenia. Health Phys. 83, 901–906.

Virk, H.S., Sharma, A.K., 1997. Microseismicity trends in N-W Himalaya using radon signals. In: Virk, H.S. (Ed.), Rare Gas Geochemistry. Guru Nanak Dev University, Amritsar, pp. 117–135.

Virk, H.S., Singh, B., 1995. Radon recording of the Uttarkashi earthquake. In: Dubois, C. (Ed.), Gas Geochemistry. Univeristy of France, Comté, Science Reviews, pp. 221–229.

Virk, H.S., Walia, V., 2001. Helium/radon precursory signals of Chamoli earthquake, India. Rad. Meas. 34, 379–384.

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

Yasuoka, Y., Shinogi, M., 1997. Anomaly in atmospheric radon concentration: a possible precursor of the 1995 Kobe, Japan, earthquake. Health Phys. 72, 759–761.

Zmazek, B., Vaupotič, J., Živčić, M., Premru, U., Kobal, I., 2000a. Radon monitoring for earthquake prediction in Slovenia. Fizika B (Zagreb) 9, 111–118.

Zmazek, B., Vaupotič, J., Živčić, M., Martinelli, G., Italiano, F., Kobal, I., 2000b. Radon, temperature, electrical conductivity and $^3He/^4He$ measutements in three thermal springs in Slovenia. In: Book of Abstracts: 2nd Dresden Symposium on Radiation Protection New Aspects of Radiation Measurements, Dosimetry and Spectrometry, SARAD and Dresden University of Technology, Dresden, Germany, September 10–14.

Zmazek, B., Vaupotič, J., Bidovec, M., Poljak, M., Živčić, M., Pineau, J. F., Kobal, I., 2000c. Radon monitoring in soil gas tectonic faults in the Krško basin. In: Book of Abstracts: 2nd Dresden Symposium on Radiation Protection New Aspects of Radiation Measurements, Dosimetry and Spectrometry, SARAD and Dresden University of Technology, Dresden, Germany, September 10–14.

Zmazek, B., Vaupotič, J. and Kobal, I., 2002a. Radon, temperature and electric conductivity in slovenian thermal waters as potential earthquake precursors. In: Proceeding of the First Workshop on Natural

Radionuclides in Hydrology and Hydrogeology, Centre Universitaire de Luxembourg, September 4–7.

Zmazek, B., Italiano, F., Živčić, M., Vaupotič, J., Kobal, I., Martinelli, G., 2002b. Geochemical monitoring of thermal waters in Slovenia: relationships to seismic activity. Appl. Radiat. Isot. 57, 919–930.

Zmazek, B., Živčić, M., Vaupotič, J., Bidovec, M., Poljak, M., Kobal, I., 2002c. Soil radon monitoring in the Krško basin, Slovenia. Appl. Radiat. Isot. 56, 649–657.

Zmazek, B., Todorovski, L., Džeroski, S., Vaupotič, J., Kobal, I., 2003. Application of decision trees to the analysis of soil radon data for earthquake prediction. Appl. Radiat. Isot. 58, 697–706.

Zmazek, B., Todorovski, L., Džeroski, S., Vaupotič, J., Kobal, I., 2004. Radon anomalies in soil gas caused by seismic activity. Acta Geotech. Slo. 1, 12–19.