



The use of multivariate analysis of the radon variability in the underground laboratory and indoor environment

Jelena Filipović,
Dimitrije Maletić,
Vladimir Udovičić,
Radomir Banjanac,
Dejan Joković,
Mihailo Savić,
Nikola Veselinović

Abstract. The paper presents results of multivariate analysis of variations of radon concentrations in the shallow underground laboratory and a family house, depending on meteorological variables only. All available multivariate classification and regression methods, developed for data analysis in high-energy physics and implemented in the toolkit for multivariate analysis (TMVA) software package in ROOT, are used in the analysis. The result of multivariate regression analysis is a mapped functional behaviour of variations of radon concentration depending on meteorological variables only, which can be used for the evaluation of radon concentration, as well as to help with modelling of variation of radon concentration. The results of analysis of the radon concentration variations in the underground laboratory and real indoor environment, using multivariate methods, demonstrated the potential usefulness of these methods. Multivariate analysis showed that there is a potentially considerable prediction power of variations of indoor radon concentrations based on the knowledge of meteorological variables only. In addition, the online system using the resulting mapped functional behaviour for underground laboratory in the Institute of Physics Belgrade is implemented, and the resulting evaluation of radon concentrations are presented in this paper.

Key words: multivariate analysis • radon variability

Introduction

The research of the dynamics of radon in various environments, especially indoors, is of great importance in terms of protection against ionizing radiation and in designing of measures for its reduction. Research of radioactive emanations (of radon (^{222}Rn) and thoron (^{220}Rn)) are in the domain of radiation physics, but since a few decades ago, subject of radioactive emanation involves many other scientific disciplines, thus giving a multidisciplinary character to this research. Published results and development of many models to describe the behaviour of indoor radon indicate the complexity of this research, especially with models for the prediction of the variability of radon, simply because the variability depends on large number of variables. Large number of factors (such as local geology, permeability of soil, building materials used to build the buildings as well as the habits of people) impact the variation of radon, and therefore, it is important to study their correlation. In this paper, the results of correlative analysis of indoor radon and meteorological variables are presented. Furthermore, the results of multivariate classification and regression analysis is presented. More details of this study can be found in [1].

Indoor radon variation depends significantly on large number of factors, which include the local ge-

J. Filipović, D. Maletić, V. Udovičić✉, R. Banjanac,
D. Joković, M. Savić, N. Veselinović
Institute of Physics Belgrade,
University of Belgrade,
118 Pregrevica Str., 11080 Belgrade, Serbia,
E-mail: udovicic@ipb.ac.rs

Received: 4 January 2016

Accepted: 24 March 2016

ology, soil permeability, building materials, lifestyle characteristics and meteorological variables. In order to analyse the dependence of radon variation on multiple variables, multivariate analysis needs to be used.

The demand for detailed analyses of large amount of data in high-energy physics resulted in wide and intense development and usage of multivariate methods. Many of multivariate methods and algorithms for classification and regression are already integrated into the analysis framework ROOT [2], more specifically, into the toolkit for multivariate analysis (TMVA [3]). Multivariate analysis toolkit is used to create, test and apply all available classifiers and regression multivariate methods implemented in the TMVA in order to find methods that are the most appropriate and yield maximum information on the dependence of indoor radon concentrations on the multitude of meteorological variables. Classification methods are used to find out if it is possible to classify radon concentrations into low and high concentrations, using arbitrary cut value for radon concentrations. Regression methods are used as a next step with a goal to find out which regression method can, if any, on the basis of input meteorological variables only, give an output that would satisfactorily close match the observed variations of radon concentrations. The output of usage of multivariate regression analysis methods is mapped functional behaviour, which can be used to evaluate the measurements of radon concentrations using input meteorological variables only. The prediction of radon concentrations can be an output of mapped function when the prediction of input meteorological variables exists.

Short-term radon measurements in laboratory and real environment

Depending on the integrated measurement time, methods of measurement of radon concentrations in air may be divided into long-term and short-term ones. For the measurements of radon concentration presented in this paper, the SN1029 radon monitor (manufactured by the Sun Nuclear Corporation, NRSB approval-code 31822) has been used as active, short-term measurement device. The device consists of two diffused junction photodiodes as a radon detector and is furnished with sensors for temperature, barometric pressure and relative humidity. The user can set the measurement intervals from 30 min to 24 h. It was set to record simultaneously the radon concentration, temperature, atmospheric pressure and relative humidity.

For the purposes of determining the best multivariate methods to use in the analysis, the results are obtained using radon monitor are from measurements in two locations, the Low-Background Laboratory for Nuclear Physics in the Institute of Physics in Belgrade and in a family house.

The underground Low-Background Laboratory for Nuclear Physics is selected for measurement and analysis because routine measurements in this labo-

ratory require low levels of radon concentration with minimum temporal variations. Low-background laboratory is located on the right bank of the river Danube in the Belgrade borough of Zemun, on the grounds of the Institute of Physics. The ground level portion of the laboratory, at 75 m above sea level, is situated at the foot of a vertical loess cliff, about 10 m high. The underground part of the laboratory, useful area of 45 m², is dug into the foot of the cliff. Underground laboratory is surrounded with 30-cm thick concrete wall. The overburden of the underground laboratory is thus about 12 m of loess soil. Significant efforts are being made to contain the low radon concentration within the laboratory. The underground laboratory is completely lined with a hermetically sealed, 1-mm thick aluminium foil. The ventilation system maintains the overpressure of 2 mbar, so as to prevent radon diffusion from the soil. Fresh air entering the laboratory is passed through a two-stage filtering system. The first stage is a mechanical filter for dust removal. The second one is a battery of coarse and fine charcoal active filters. The concentration of radon is kept at an average value of about 10 Bq/m³.

In the Low-Background Laboratory for Nuclear Physics, radon concentrations were measured in period from 2008 to 2011 and continued later on periodically about a couple of months each year. Measurements of meteorological variables used in the analysis were recorded since 2008 and are taken from the meteorological station located 4 km from the laboratory. Measurements of radon concentrations, room temperature, atmospheric pressure and relative humidity inside the laboratory were obtained using radon monitor. The results obtained from the measurements of radon concentrations and their influence on gamma and cosmic ray measurements in the laboratory were published in several articles in international scientific journals [4–6].

The family house selected for the measurements and analysis of variations of radon concentrations is a typical house in Belgrade residential areas, with requirement of existence of cellar. House is built on limestone soil. Radon measurements were carried out in the living room of the family house, which is built of standard materials (brick, concrete, mortar) and isolated with styrofoam. During the period of measurements (spring–summer), the house was naturally ventilated and air conditioning was used in heating mode at the beginning of the measurement period. During the winter period measurements, the electrical heating was used in addition to air conditioning. Measured radon concentrations, room temperature, atmospheric pressure and relative humidity inside the house were obtained using radon monitor. Values of meteorological variables in measurement period were obtained from an automatic meteorological station located 400 m from the house in which the measurement was performed. We used the following meteorological variables: external air temperature, pressure and humidity, solar radiation, wind speed at a height of 10 m above ground, precipitation, evaporation and temperature and humidity of the soil at a depth of 10, 20, 30 and 50 cm.

Correlation and regression analysis of the results

All multivariate methods implemented in the TMVA are used in our search. All multivariate methods in TMVA belong to the family of ‘supervised learning’ algorithms [1]. All methods make use of training events, for which the desired output is known, to determine the mapping function that either describes a decision boundary (classification) or an approximation of the underlying functional behaviour defining the target value (regression). Every MVA methods see the same training and test data. The two best performing multivariate methods for our purposes are boosted decision trees (BDT) and artificial neural networks (ANN).

The determination of correlation coefficients between measured radon concentration and meteorological variables serves as a good tool for identifying the variables with strongest correlation, which are not excluded from the analysis later on. Also, correlation coefficient tables gives a good overview of input data and their intercorrelations. In Fig. 1, the correlation matrix of linear correlation coefficients as an overview of intercorrelations of measured radon concentration and all input meteorological variables are shown for underground laboratory. The input variables in case of analysis of underground laboratory are atmospheric pressure, temperature and humidity in laboratory (P_{rm} , T_{rm} , H_{rm}) and outdoor (P , T , H) and differences in measured values of pressure ($P - P_{rm}$), temperature ($T - T_{rm}$) and humidity ($H - H_{rm}$) in laboratory and outdoor. Input meteorological variables in case of family house are the same as the list of measured meteorological variables from nearby meteorological station, with the addition of differences in temperature ($T - T_{rm}$) and humidity ($H - H_{rm}$) from indoor and outdoor values, where indoor measurements results were obtained using radon monitor.

Multivariate methods within the package TMVA in ROOT can search for best multivariate approximation of functional behaviour for the classification function of radon concentration depending on meteorological variables. In the analysis, several mul-

Linear correlation coefficients in %										
Radon	17	4	25	14	5	1	13	5	-14	100
H-H _{mm}	-81	13	-73	-94	10	79	-94	-8	100	-14
P-P _{mm}	9	-15	13	1	15			100	-8	5
T-T _{mm}	80	-14	77	99	-14	-68	100		-94	13
H	-43	3	-18	-65	3	100	-68		79	1
P	-4	95	-12	-13	100	3	-14	15	10	5
T	86	-13	80	100	-13	-65	99	1	-94	14
H _{mm}	84	-17	100	80	-12	-18	77	13	-73	25
P _{mm}	-7	100	-17	-13	95	3	-14	-15	13	4
T _{mm}	100	-7	84	86	-4	-43	80	9	-81	17
T _{mm} P _{mm} H _{mm} T P H T-T _{mm} P-P _{mm} H-H _{mm} Radon										

Fig. 1. Correlation matrix with linear correlation coefficients as an overview of radon and meteorological variables intercorrelations in case of the Low-Background Laboratory for Nuclear Physics.

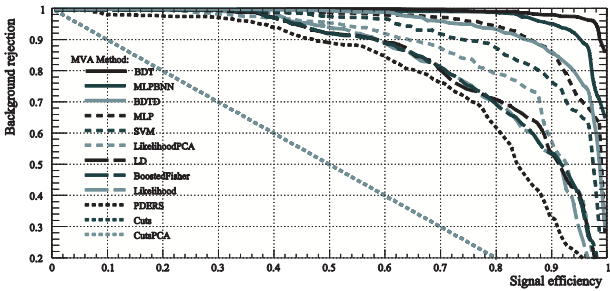


Fig. 2. ROC curve for all multivariate methods in case of house measurements.

tivariate methods were tested, and best performed method was BDT. This can be seen by presenting the receiver operating characteristics (ROC) curve for all tested multivariate methods in case of house measurements (Fig. 2). The BDT method has the highest value of integrated ROC function.

BDT has proven to be the most effective method for the classification of radon concentrations in case of data obtained from the house as well as those obtained from measurements in the Low-Background Laboratory for Nuclear Physics.

The next step in the analysis is the regression analysis, which is the way of finding a mapped function behaviour of dependence of radon concentrations and meteorological input variables. The regression analysis was done using the TMVA packages, already used in classification analysis, and for the same set of measured radon concentration and meteorological variables in underground laboratory and a family house in Serbia. Multivariate method BDT was found to be the best suited for regression analysis also, as was the case in classification analysis.

The data of measured radon concentration in house and BDT evaluated values, using only the values of meteorological variables, without the knowledge of measured values (i.e. in the testing set of multivariate analysis), is presented for comparison in Fig. 3.

One of the possible application of having resulting mapped function, given by multivariate regression analysis, is to have prediction of radon concentration values (evaluated) based on meteorological variables alone. The online application of the regression multivariate analysis can be imple-

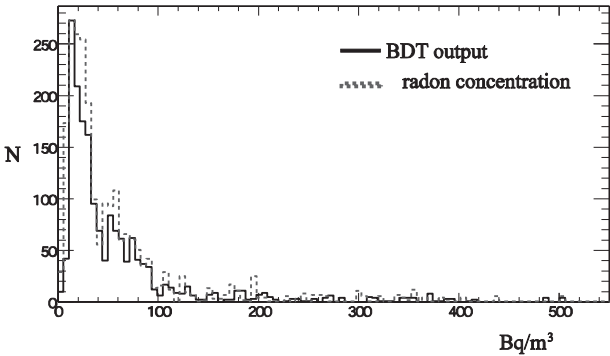


Fig. 3. BDT evaluated (predicted) values of radon concentrations based on meteorological variables using regression analysis within TMVA packages in house (left) and measured values (right).

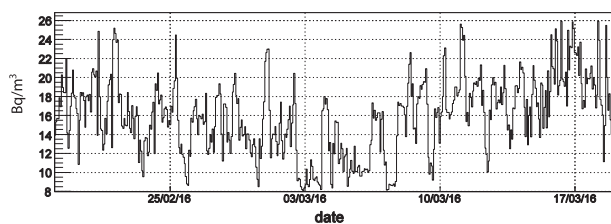


Fig. 4. BDT evaluated (predicted) values of radon concentration, based on meteorological variables alone of underground laboratory posted online and updated daily.

mented, as the one posted online for evaluation (and prediction) based on meteorological variables alone (Fig. 4).

Limitation of multivariate methods

As the multivariate methods used in the analysis are 'supervised learning' algorithms, the performance of the main result of multivariate analysis, the resulting mapped functional behaviour, depends on learning process. Limitation of multivariate analysis in the analysis of radon dependence on meteorological variables are coming from small number of measurements used in learning process, unlike the great number of measurements in high-energy physics experiments. As the next logical step in multivariate analysis presented in this paper should be inclusion of variables such as local geology, permeability of soil, building materials used to build the buildings as well as the habits of people, the requirement for efficient multivariate analysis is to have many measurements in many different houses. Many measurements would help to get good mapped functional behaviour, as opposed to possible existence of theoretical modelling that is independent on number of measurements. In this sense, if the number of measurements is not great, multivariate analysis can be used only as hell to indicate which variables are more important to be used in theoretical modelling, for comparison of mapped and modelled functions, and modelled function test. Another important limitation of multivariate analysis is that no 'straightforward' interpretation of mapped functional behaviour is possible, or simply, the mapped function is a 'black box'. This comes from the fact that the error minimization in learning algorithms, while mapping the functional behaviour, is an important part in learning process.

Conclusions

The paper presents the results of multivariate analysis of variations of radon concentrations in the shallow underground laboratory and a family house, depending on meteorological variables only. This test of multivariate methods, implemented in the

TMVA software package, applied to the analysis of the radon concentration variations connection with meteorological variables in underground laboratory (with ventilation system turned on and off) and typical house in Serbia, demonstrated the potential usefulness of these methods. It appears that the method can be used for the prediction of the radon concentrations, on the basis of predicted meteorological variables. The next step in multivariate analysis presented in this paper should be inclusion of variables such as local geology, permeability of soil, building materials used to build the buildings as well as the habits of people. The requirement for efficient multivariate analysis is to have many measurements in many different houses, which makes multivariate method very useful only when having many measurement, for instance, during radon mapping campaigns. Many measurements would help to get good mapped functional behaviour, as opposed to possible existence of theoretical modelling that is independent on number of measurements. Generally, multivariate analysis can be used to help indicate which variables are more important to be used in theoretical modelling, furthermore, for comparison of mapped and modelled functions, and modelled function test.

Another usage of the results of classification multivariate analysis presented in this paper is the implementation of online warning system for possible increased radon concentration in family houses based on meteorological variables only.

References

1. Maletić, D., Udovičić, V., Banjanac, R., Joković, D., Dragić, A., Veselinović, N., & Filipović, J. (2014). Comparison of multivariate classification and regression methods for indoor radon measurements. *Nucl. Technol. Radiat. Prot.*, 29, 17–23.
2. Hoecker, A., Speckmayer, P., Stelzer, J., Therhaag, J., Von Toerne, E., & Voss, H. (2007). TMVA – Toolkit for Multivariate Data Analysis. *PoS ACAT 040*. arXiv:physics/070303.
3. Brun, R., & Rademakers, F. (1997). ROOT – An Object Oriented Data Analysis Framework. *Nucl. Instrum. Methods Phys. Res. Sect. A-Accel. Spectrom. Dect. Assoc. Equip.* 389(1/2), 81–86.
4. Udovičić, V., Grabež, B., Dragić, A., Banjanac, R., Joković, D., Panić, B., Joksimović, D., Puzović, J., & Aničin, I. (2009). Radon problem in an underground low-level laboratory. *Radiat. Meas.*, 44, 1009–1012.
5. Udovičić, V., Aničin, I., Joković, D., Dragić, A., Banjanac, R., Grabež, B., & Veselinović, N. (2011). Radon time-series analysis in the Underground Low-Level Laboratory in Belgrade, Serbia. *Radiat. Prot. Dosim.*, 145(2/3), 155–158.
6. Banjanac, R., Udovičić, V., Dragić, A., Joković, D., Maletić, D., Veselinović, N., & Grabež, B. (2013). Daily variations of gamma-ray background and radon concentration. *Rom. J. Phys.*, 58(Suppl.), S14–S21.