# STATISTICAL MODELLING OF INDOOR RADON CONCENTRATION USING METEOROLOGICAL PARAMETERS

**R. Dubčáková**

•

VŠB-Technical University of Ostrava, Faculty
of Safety Engineering, Ostrava, Czech Republic

e-mail: renata.dubcakova@vsb.cz

**P. Praks**

•

VŠB-Technical University of Ostrava, Faculty of Electrical
Engineering and Computer Science, Ostrava, Czech Republic

e-mail: pavel.praks@vsb.cz

## ABSTRACT

The radon volume activity in buildings is generally time variable. Its variability is caused by many natural and man-made factors. An example of these factors includes meteorological parameters, soil properties, characteristics of the building construction, properties of water used in the building and also the behavior of inhabitants. These factors can influence each other and also they are related with the exposition of inhabitants. This article reports a continual indoor radon monitoring and a statistical evaluation of a dataset obtained by the 18 days-long measuring in a house located in the Czech Republic.

The contributions of carefully selected meteorological parameters and human influences were also observed. Results of the observation were divided into two parts (inhabited, uninhabited) and analyzed in relation with the indoor radon concentration. The multiplied linear regression was applied to model obtained datasets. Results of time series analyses of the continual indoor radon concentration and the meteorological monitoring are presented and discussed as well.

## INTRODUCTION

One of a natural radionuclide that commonly occurs in rocks is uranium U-238 [[8]]. It is evident that radon rising by radioactive decay from uranium is steadily generated in natural conditions. Considering the human life span, the radon is in principle a stationary source of irradiation.

The radon volume concentration in a building is not stable; it can vary in time because the radon concentration is influenced by many factors e.g. meteorological, structural parameters [[3], [8], [9], [10], [11]]. These are several reasons why it is not straightforward to describe the radon behaviour in houses and exactly determine the indoor radon concentration. The concentration of radon in homes and other buildings varies substantially from one area to another, from one structure to another, and even within the same structure [[2], [3], [8]]. Likewise, in the same building, there is often a substantial variation with time on various temporal scales, i.e., season to season, week to week, and on a daily or hourly basis [[2], [6], [9], [11]].

Understanding the nature and origin of such variability is important as a basis for evaluating the range of the radon concentration problem, interpreting monitored data and for formulating of effective strategies for control.

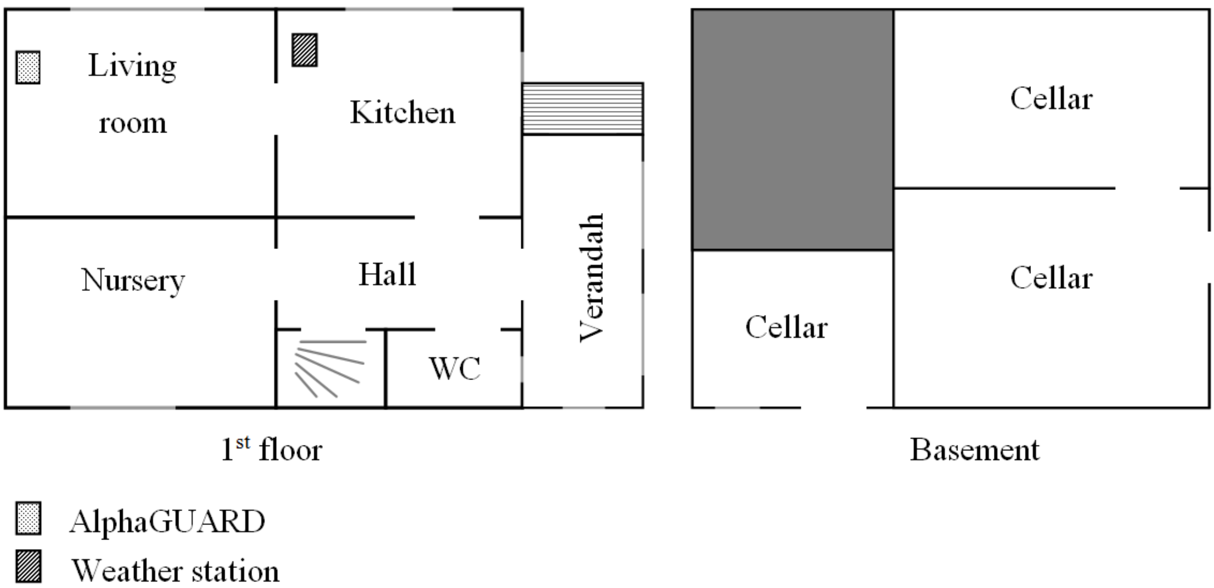
In this contribution, we try to find a relationship between the indoor radon concentration and the current meteorological factors. The aim is also to find factors, which mostly influence the indoor radon concentration and to develop a model for describing the behavior of radon concentration depending on selected factors in the given room. We tried to find how selected meteorological factors influence the indoor radon concentration by using the multiple linear regressions techniques. The most important meteorological factors and parameters of a reduced linear model characterizing the

radon concentration were statistically estimated. Finally, in order to verify the obtained model, we compared measured and estimated data.
This paper is organised as follows. In the first section, we give an overview of the time period of the measurements, the construction of the measured house and a description of used tools for measuring radon concentrations and its location within the house. In the second section, we describe the obtained measured datasets and its modification, statistical analysis, tables and figures of results of the applied analysis and the best alternative model of radon concentrations in the selected room. In the third section, we discuss the selection of the best, statistically reduced models. We also analyse an effect of (non)presence of human to the ability of the model to correctly predict the radon concentration. The summary of interesting results and the future work are discussed in conclusion.

## MATERIALS AND METHODS

### *The description of the measured house*

The measurements were carried in one family house (see Figure 1.). The experimental building is situated in the village Strašín in the southwest region of the Czech Republic near the Šumava Mountains. The house was built in 1930. It is a masonry and partly cellar construction which is situated on a steep land on the edge of the village. The house has two cellars (with separate outdoors), three rooms on the first floor and one attic that was closed during the measurements. All rooms are ventilated by natural means and two of all is heated by stoves.




The ground plan of measured house

The building is connected to the public water pipe-line as required by Czech law, so the water source of radon is eliminated [1]. The measured gamma dose rate in contact with construction materials was very close to the natural background. It indicates that the building materials do not contain any natural radionuclide. So we suppose that the main source of indoor radon is the soil gas. The average value of radon concentration in soil was 20 kBq/m$^3$.
The studied room (living room) is located on the north side of the building and joint with the adjacent heated room (kitchen); see Figure 1.
There was a cinder filling under the wooden floor of the experimental room. The floor was covered by a carpet. The volume of the experimental room was 30 m$^3$.
The house was selected on the basis of screening measurements performed by the Czech National Radiation Protection Institute in Prague, the Czech Republic.

### *The description of the equipment*

We have chosen the continual method of monitoring for the evaluation of radon concentration in context of the selected meteorological factors. The Alphaguard radon monitor has been used for continuous detection of the radon concentration [[5]]. It has an ionization chamber and uses an alpha spectroscopy for the radon detection. The two

common isotopes of radon (i.e., Rn-222 and Rn-220) are identified through their respective energies from the alpha decays. The signal generated from the alpha detection is converted to a digital output. The AlphaGUARD was operated in the diffusion mode at a 60-min cycle.
The actual meteorological conditions were recorded by the wireless weather station Vantage Pro2™ [[4]]. Sensors for detecting outside conditions were located southeast of the building and the indoor station was placed in the adjacent room constantly connected with the experimental room.

## Results and Discussion

Time series of the indoor radon concentration, the indoor climate parameters including temperature, humidity and atmospheric pressure in the measured room were obtained by the radon monitor AlphaGUARD.
The radon monitoring started 3rd October at 5 PM and finished 20th October at 3 PM. The values were recorded every 60 minutes so each time series has 407 values. In order to stabilize the radon monitor and the indoor climate, the first two values (3rd October at 5 PM and 6 PM) were eliminated from analyses. The final analysing time series starts at 3rd October at 7 PM and ends 20th October at 3 PM. The total number of analysed values is 405. We compared these time series with ones obtained by the weather station Vantage Pro2™. The weather station has recorded time series of the following meteorological parameters:

- indoor/outdoor temperature,
- indoor/outdoor relative air humidity,
- indoor/outdoor barometric pressure,
- indoor/outdoor dew-point,
- wind direction and velocity,
- rainfall.

The weather station operated during the same time as the AlphaGUARD. The values of the weather station were recorded every 30 minutes, so the total number of recorded values was 813. Because of the different sampling intervals of the weather station from one of the AlphaGUARDs, we have modified time series of meteorological parameters to the one-hour interval by the averaging; we used the arithmetic averages of last two values corresponding to 30 minutes interval values according to the equation 1:

$$x_i = \frac{x_{2i} + x_{2i-1}}{2} \quad \text{, for } i = 1, 2\ldots 406 \qquad (1)$$

We eliminated the first value for the same reasons as for the monitoring of radon. Finally, the adjusted time series of meteorological parameters had 405 values.

### *Monitoring of the human influence*

For reasons of known influence of human activities (such as heating or ventilation), we divided each time series into 2 parts – inhabited and uninhabited. The object was 11 days inhabited and 7 days uninhabited. Although an inhabitant left the house 15th October at 9:10 AM, the second "Uninhabited" part starts on October 15 at 7:00 PM, in order to stabilize indoor climate and regression models after the person has left. For each measured parameter, we finally obtained time series of 291 values for the first "Inhabited" part and time series of 114 values for the second "Uninhabited" part.

### *Statistical analysis of measured data*

In our contribution we try to find a relationship between the indoor radon concentration (dependent variable) and several measured meteorological parameters (independent variables). The statistical analyses of these data sets were done using multiple linear regression techniques [[7]]. The general purpose of multiple regressions is to learn more about relationships between several independent (predictor) variables and the dependent (criterion) variable. The statistical analyses were provided by the Statgraphics Plus 5.1 software.
Because of different effects of radon concentration in the case of inhabited and uninhabited time, we selected various independent parameters, see Tables 1 - 2. Tables shows the results of fitting of various multiple regression models, in order to describe the relationship between indoor radon concentration and 13 predictor variables in case of "Inhabited part" and 7 predictor variables in case of "Uninhabited part". Models were fitted by various combinations of meteorological parameters. The tabulated statistics include the mean squared error (MSE), the adjusted and unadjusted R-Squared values and appropriate parameters.
A linear regression technique uses the estimated MSE to determine the statistical significance of factors under the study. If an MSE is zero, it means that the estimator predicts observations with perfect accuracy.

The *R-squared* coefficient of determination is a statistical measure of how well the regression model approximates the real data-points. The values can vary from 0 to 1. If *R-squared* = 1, the regression curve perfectly fits the observed data [[7]]. The adjusted R-Squared statistic measures the proportion of the variability in the indoor radon concentration which is explained by the model. The explanation of this statistic is almost the same as *R-squared* but it penalizes the statistic as extra variables included in the model. Of course, larger values of adjusted R-Squared correspond to smaller values of the mean squared error.

Models with the largest R-Squared values for "Inhabited Part" of Time Series (left) and explanations of parameters (right)

| MSE | R-Squared | Adjusted R-Squared | Included Variables |
|---|---|---|---|
| 16691.1 | 34.7539 | 33.6092 | CDELM |
| 16955.0 | 33.7222 | 32.5594 | BCDFL |
| 17006.1 | 33.5225 | 32.3562 | CDEKL |
| 17017.9 | 33.4762 | 32.3091 | CDGLM |
| 17027.1 | 33.2066 | 32.2724 | CDEL |
| 17032.6 | 33.4188 | 32.2507 | BDELM |
| 17073.8 | 33.0235 | 32.0868 | CDLM |
| 17118.3 | 32.849 | 31.9098 | DELM |
| 17359.2 | 31.9039 | 30.9515 | CDKL |
| 17361.7 | 31.656 | 30.9416 | CDL |

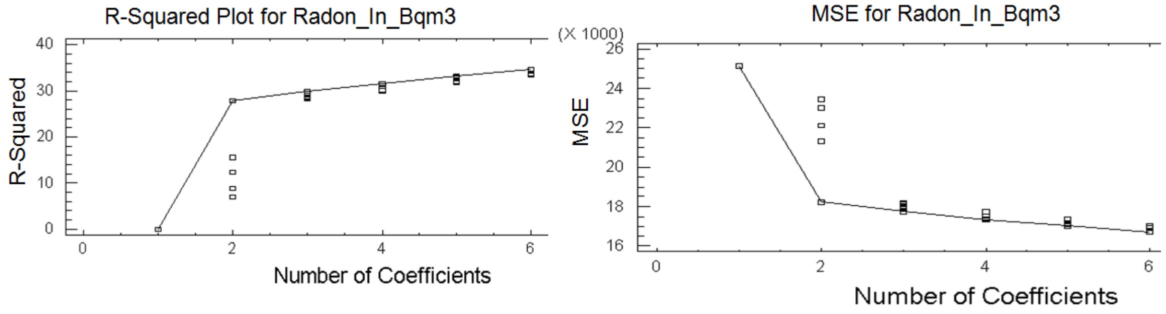
| Signification | Explanation |
|---|---|
| A=DewingPIn_oC | Dewing Point in the Kitchen |
| B=DewingPOut_oC | Dewing Point Outdoor |
| C=Hour | Hours of a Day |
| D=HumIn_proc | Relative Humidity in the Kitchen |
| E=HumIn_procAG | Relative Humidity in the Measured Room |
| F=HumOut_proc | Relative Humidity Outdoor |
| G=P_hPa | Atmospheric Pressure Outdoor |
| H=P_hPa_AG | Atmospheric Pressure in the Measured Room |
| I=Rainfall_mm | Rainfall |
| J=Tin_oC | Temperature in the Kitchen |
| K=Tin_oC_AG | Temperature in the Measured Room |
| L=Tout_oC | Temperature Outdoor |
| M=wind_invms | Wind Velocity |

Models with the largest R-Squared values for "Uninhabited Part" of Time Series (left) and explanations of parameters (right)

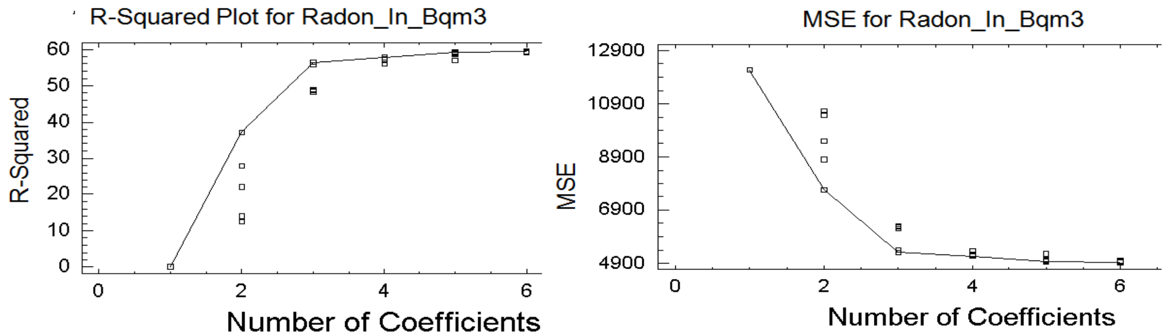
| MSE | R-Squared | Adjusted R-Squared | Included Variables |
|---|---|---|---|
| 4937.2 | 61.2712 | 59.4782 | ABCDG |
| 4950.1 | 61.1701 | 59.3724 | ABCDF |
| 4955.67 | 60.7665 | 59.3267 | ACDE |
| 4966.34 | 61.0427 | 59.2391 | ABCDE |
| 4972.7 | 60.9928 | 59.1869 | ACDFG |
| 4977.77 | 60.5915 | 59.1453 | ACDF |
| 4983.36 | 60.9092 | 59.0994 | ACDEG |
| 5015.33 | 60.2941 | 58.837 | ABCD |
| 5052.4 | 60.0007 | 58.5328 | ACDG |
| 5154.82 | 58.8154 | 57.6922 | ACD |

| Signification | Explanation |
|---|---|
| A=DewingPIn_oC | Dewing Point in the Kitchen |
| B=DewingPOut_oC | Dewing Point Outdoor |
| C=HumIn_procAG | Relative Humidity in the Measured Room |
| D=HumOut_proc | Relative Humidity Outdoor |
| E=P_hPa | Atmospheric Pressure Outdoor |
| F=P_hPa_AG | Atmospheric Pressure in the Measured Room |
| G=Tdiff_oC | Difference of Temperature in the Measured Room and Outdoor |

Figure 2. and Figure 3. show selected statistical models which give the largest adjusted *R-Squared* values for "Inhabited Part" and "Uninhabited Part" of time series, respectively.




The Comparison of R-Squared (left) and Mean Squared Error (right) with the Number of Coefficients for "Inhabited Part" of Time Series




The Comparison of R-Squared and Mean Squared Error (MSE) with the Number of Coefficients for "Uninhabited Part" of Time Series

Depending on the size of the *R-Squared* and MSE (Tables 1., 2. and Figures 2., 3.) we have chosen the favourite models, which has the highest *R-Squared* and the lowest MSE.

## *The best model for the "inhabited part" of time series*

The best model for the "inhabited part" of time series contains 5 variables: hours of a day (*Hour*), relative humidity in the kitchen (*HumIn_proc*), relative humidity in the kitchen (*HumIn_procAG*), temperature outdoor (Tout_oC) and wind velocity (wind_invms). Table 3. shows the estimates of variables, its errors of the estimate and corresponding P-values. Since the P-value is less than 0.05, there is an indication of possible serial correlations. In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0.0098, belonging to the wind velocity (*wind_invms*). Since the P-value is less than 0.01, the highest order term is statistically significant at the 99% confidence level and no simplification of the model is statistically recommended.

The Best Alternative Model of the "Inhabited Part" of Time Series.

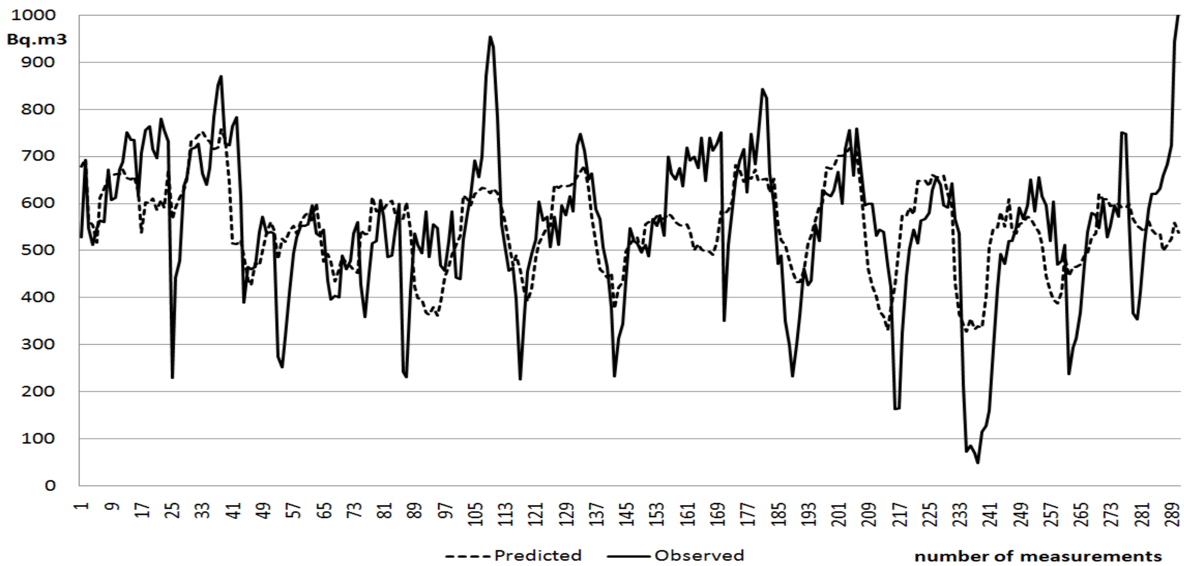
| Parameter | Estimate | Error | P-Value |
|---|---|---|---|
| CONSTANT | 610.157 | 228.463 | 0.0080 |
| Hour | -3.39318 | 1.17631 | 0.0042 |
| HumIn_proc | 15.9026 | 3.28874 | 0.0000 |
| HumIn_procAG | -12.1664 | 4.42536 | 0.0064 |
| Tout_oC | -18.858 | 1.96029 | 0.0000 |
| wind_invms | 77.8566 | 29.9478 | 0.0098 |

The R-Squared statistic (Table 1.) indicates that the model of the "inhabited part" explains 34.75% of the variability of the indoor radon concentration. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 33.60%. The standard error of the estimate (Table 1.) shows the standard deviation of the residuals to be 129.194.

The output of the analysis is a multiple linear regression model, which describes the relationship between the indoor radon concentration (*Radon_In_Bqm3*) and five independent variables. The expression of the fitted model is:

*Radon_In_Bqm3=610.157-3.39318*Hour+15.9026*HumIn_proc-12.1664*HumIn_procAG-18.858*Tout_oC+77.8566*wind_invms* (2)

Figure 4. shows how the fitted model follows the shape of the observed values of indoor radon concentration.




The Comparison of the Predicted and Measured Values of "Inhabited Part" of Time Series.

## The best model for the "uninhabited part" of time series

The best model for the "uninhabited part" of the time series was estimated under the value of the R-Squared and the mean squared error as before. By analysing Table 2. and Figure 3., the best model for the "uninhabited part" of time series contains these five variables: dewing point in the kitchen (*DewingPIn_oC*), dewing point outdoor (*DewingPOut_oC*), relative humidity in measured room (*HumIn_procAG*), relative humidity outdoor (*HumOut_proc*), difference of temperature in the measured room and outdoor (*Tdiff_oC*), see Table 4.

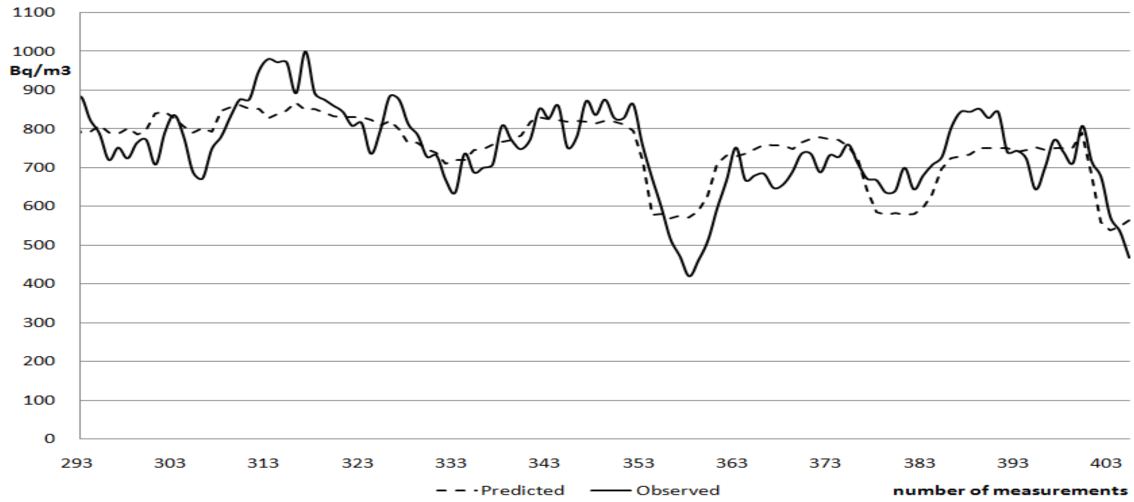
The Best Alternative Model of the "Uninhabited Part" of Time Series.

| Parameter | Estimate | Error | P-Value |
|---|---|---|---|
| CONSTANT | -1036.0 | 459.595 | 0.0262 |
| DewingPIn_oC | 60.416 | 20.9351 | 0.0047 |
| DewingPOut_oC | -36.7471 | 19.5222 | 0.0625 |
| HumIn_procAG | 15.7088 | 5.60164 | 0.0060 |
| HumOut_proc | 10.0782 | 3.62535 | 0.0064 |
| Tdiff_oC | -30.3273 | 18.3721 | 0.1017 |

The R-Squared statistic (Table 2.) indicates that the model as fitted explains 61.27% of the variability of the indoor radon concentration. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 59.48%.
The expression of the fitted model for the "uninhabited part" is:

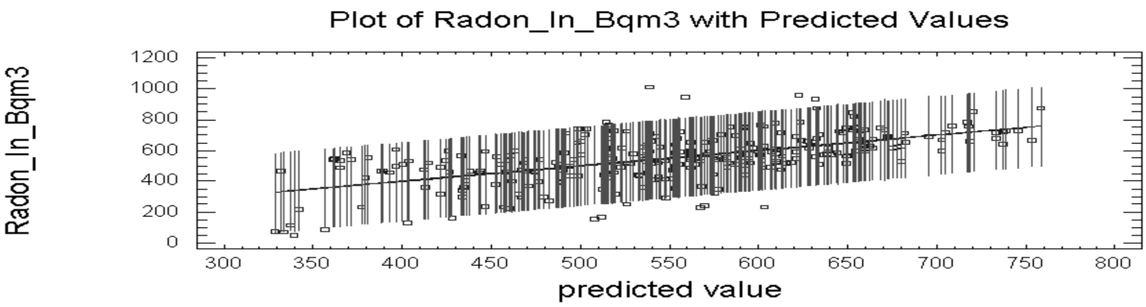
*Radon_In_Bqm3 = -1036.0 + 60.416*DewingPIn_oC - 36.7471*DewingPOut_oC+ 15.7088*HumIn_procAG + 10.0782*HumOut_proc - 30.3273*Tdiff_oC* (3)

Figure 5. shows how the fitted model follows the shape of the observed values of indoor radon concentration.

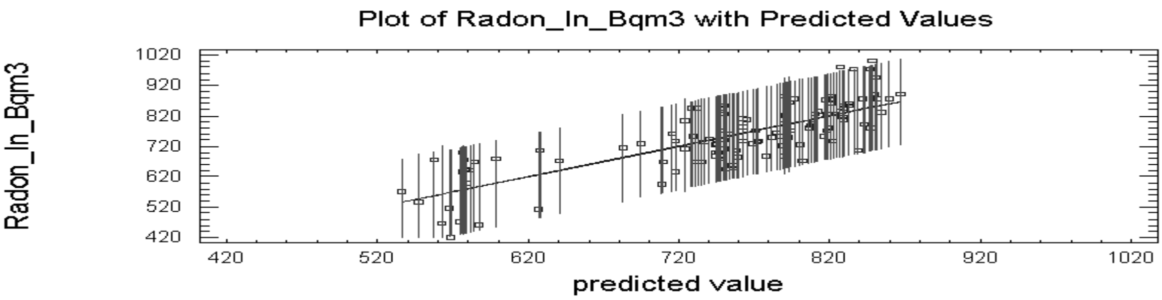



The Comparison of the Predicted and Measured Values of "Uninhabited Part" of Time Series

Figure 6. and Figure 7. show limits for predicted values at 95 % confidence level. Confidence intervals show how precisely the model estimates the observed values given the amount of available measured data and the measured noise. Figures also show statistically unusual observations (outliers). It can be caused by the errors of measuring equipments, the statistical fluctuation of the radioactive decay, stochastic characteristics of the building structure, the duration of measurements and by limitations of the used multiple linear regression models.
In the case of „inhabited part" of time series (see Figure 6.) we can observe more unusual observations. It approves the presence of human and his hardly predictable activities (heating, ventilation, …), which significantly influence the indoor radon concentration.




Limits for predicted values at 95 % confidence level – "inhabited part" of time series




Limits for predicted values at 95 % confidence level – "uninhabited part" of time series

## Conclusion

In this paper, we measured and statistically analysed a relationship between the indoor radon concentration and various meteorological parameters. Because of human activities, which significantly influence the indoor radon concentration, we separated the dataset into 2 parts: "inhabited" and "uninhabited". The statistical analyses of these datasets were realized by the multiple linear regression techniques. The "uninhabited" part of the measured data can be quite reliably predicted by the multiple linear models (R-Squared ~ 61%). On the other hand, the R-Squared of the "inhabited" part is

only 34.7 % and the model is characterized by the mean squared error, which is approximately 16691.1/ 4937.2 ~ 3.40 times larger than the mean squared error of the model belonging to the "uninhabited" part.
Surprisingly, the adjusted models of the indoor radon concentration in the selected room were in both cases determined i.a. by the humidity and in the "uninhabited" case by the dew point. The known influence of temperature differences was not statistically proved for the "inhabited" case. It could be caused by human activities, especially by the frequent ventilation (also during nights) or heating by the stove: The stove could exhaust the indoor air by radon.
The large value of MSE of the "inhabited" part suggests unpredictable human activities. The human activity parameter was included indirectly by the form of timing (hours of a day). The description of uncertainties of the human activities, which significantly influence the radon concentration (awake, sleeping, …), will be the subject of the future research.
Analyses of radon concentrations are now connected to one selected room of the partly basement family house. The datasets of monitoring of the rest of the building will be available and the analogous analyses for remaining rooms are also planned.
In the future work, it would also be interesting to deeply analyse relationships among meteorological parameters, in order to improve and/or to simplify the developed statistical models of radon concentration in the room. We are just testing nonlinear models using the software Eureqa [12].

## Acknowledgments

This work was supported by the budget from the Czech Republic Government project no. 1M06047 – CQR.CZ and the internal grant of Faculty of Safety Engineering, VŠB-Technical university of Ostrava no. 040/2101/BI0409031.
We would like to thank the members of Expert group for measurements of the National Radiation Protection Institute in Prague, the Czech Republic (SÚRO); for their technical support and all-round assistance. P. Praks would like to thank Michaël Hurtgen (Université Libre de Bruxelles, Brussels, Belgium) for his helpful comments.

## References

[1] Act No. 18/1997 On Peaceful Utilisation of Nuclear Energy and Ionising Radiation (the Atomic Act) and on Amendments and Alterations to Some Acts. The Parliament of the Czech Republic.
[2] Bossew, P. & Lettner, H. 2007. Investigations on indoor radon in Austria, Part 1: Seasonality of indoor radon concentration. *Journal of Environmental Radioactivity* 98: 329 – 345.
[3] Cothern, C. R. & Smith, Jr. J.E. (ed) 1988. *Environmental Radon*. New York: Springer.
[4] Davis-Weather: Vantage Pro2™ Weather Stations [online]. Davis Instruments, Hayward, California. Available from: www.davisnet.com/weather/products/vantage2.asp [Accessed 9th June 2009]
[5] Genitron Instruments - Products: Radon Monitoring with AlphaGUARD [online]. Saphymo GmbH, Frankfurt a.M, Germany. Available from: http://www.genitron.de/download/ag_fb_gb_01_a4.pdf [Accessed 8th June 2009]
[6] Karpinska, M., Mnich, Z., Kapała, J., Antonowicz, K. & Przestalski, M. 2005. Time changebility in radon concentration in one-family dwelling houses in the northeastern region of Poland. *Radiation Protection Dosimetry* 113: 300 – 307.
[7] Montgomery, D. C. & Runger, G.C. 2003. *Applied Statistics and Probability for Engineers*. New York: John Wiley & Sons.
[8] Mueller Associates, Inc., SYSCON Corporation, Brookhaven National Laboratory 1988. *Handbook of Radon In Buildings: Detection, Safety, and Control*. New York: Hemisphere Publishing Corporation.
[9] Nazaroff, W.W., Feustel, H., Nero, A.V., Revzan, K.L., Grimsrud, D.T., Essling, M.T. & Toohey R-E 1984. *Radon transport into a single-family house with a basement* [online]. Paper LBL-16572, Lawrence Berkeley National Laboratory, University of Carolina. Available from: http://repositories.cdlib.org/cgi/viewcontent.cgi?article=3915&context=lbnl [Accessed 23rd January 2009].
[10] Nazaroff, W.W. & Nero, A.V. 1988. *Radon and its decay products in indoor air*. New York: John Wiley & Sons Inc.
[11] Nero, A.V., Gadgil, A.J., Nazaroff, W.W. & Revzan, K.L. 1990. *Indoor Radon and Decay Products: Concentrations, Causes, and Control Strategies* [online]. Technical Report, Paper LBL—27798, Lawrence Berkeley National Laboratory, University of Carolina, Washington, D.C. Available from: http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=6318312 [Accessed 1st March 2009].
Schmidt, M. & Lipson, H. 2009. Distilling Free-Form Natural Laws from Experimental Data. *Science* 324, no. 5923: 81 – 85.