ESTIMATION AND PREDICTION OF THE OUTDOOR $^{222}\text{Rn}$ AND $^{220}\text{Rn}$ PROGENY CONCENTRATIONS USING METEOROLOGICAL VARIABLES

E. SIMION1,2, I. MIHALCEA2, V. CUCULEANU3,4, F. SIMION1,3

1National Environmental Protection Agency, Radioactivity Laboratory, Splaiul Independentei 294, RO-060031, Bucuresti, Romania
2Faculty of Chemistry, University of Bucuresti, 4-12 Regina Elisabeta Av., 030017-Bucuresti, Romania
3Faculty of Physics, University of Bucuresti, Magurele, RO-077125, Bucuresti, Romania
4Academy of Romanian Scientists, 54 Splaiul Independentei, RO-050094, Bucharest, Romania

Received November 15, 2012

The present paper presents the study of the multiple linear regression model for the estimation and prediction of the time series of radon and thoron progeny concentrations in atmosphere. Radon and thoron progeny data measured in Environmental Radioactivity Monitoring Station Botoşani, part of the National Environmental Radioactivity Survey Network are modeled, at different time scales, by making use of the multiple linear regression with meteorological parameters as independent variables. The collinearity and multicollinearity of independent variables have been analyzed. The estimations were checked by using the regression statistics: multiple correlation coefficient, coefficient of determination, F-test values, level of significance (p-value) and residuals. The predictions performances have been analyzed by means of the residuals, coefficient of determination and the relative error specific to each time interval from prediction period.

Key words: radon, thoron, progeny concentration, multiple linear regression, statistics.

1. INTRODUCTION

Inert radioactive gases radon ($^{222}\text{Rn}$) and thoron ($^{220}\text{Rn}$) are produced as decay products of $^{226}\text{Ra}$ and $^{224}\text{Ra}$ from the decay series of $^{238}\text{U}$ and $^{234}\text{Th}$ that are present in all terrestrial materials.

Due to exhalation process the atmospheric concentrations of $^{222}\text{Rn}$, $^{220}\text{Rn}$ and their daughters enter the atmosphere where they are affected by the meteorological parameters specific to the planetary boundary layer: air temperature (Ta), atmospheric pressure (Pa), wind speed (V), precipitations (Pr) and relative humidity (Urel) [1, 2, 3].


In order to perform estimations and predictions for $^{222}\text{Rn}$ and $^{222}\text{Rn}$ progeny concentrations, the meteorological parameters have been used as independent variables in the multiple linear regression model [4].

The following meteorological variables were used as independent variables of the regression model: air temperature, relative humidity, atmospheric pressure, precipitation and wind speed. In order to obtain reliable results, the collinearity and multicollinearity of predictors have to be analyzed. Study of the collinearity is based on the analysis of correlation matrix of the predictor variables. In case of multicollinearity, two or more predictor variables are highly linearly correlated.

The regression model performances for the estimation of the time series values are quantified by the following regression statistics: multiple correlation coefficient ($R$), coefficient of determination ($R^2$), p-value (significance level), F (test), sum squared residuals (SSR). In case of prediction, the model performance was quantified by means of correlation coefficient $R$, p-value and SSR [7, 8].

The processed data regarding the radon and thoron progeny concentrations were measured in the year 1996 by the Environmental Radioactivity Monitoring Station Botoșani (ERMS), Romania. The meteorological data were measured in the same area. The choice of this year is justified by the fact that does not exist missing data for both radon and thoron progeny concentrations and corresponding meteorological data.

2. DATA AND METHOD

2.1. MEASURING METHOD

The measuring method of radon and thoron progeny activity concentrations is based on the aerosol aspiration on glass fiber filters with high filtration efficiency (96-99%), using high volume aerosol samplers with aspiration head located at 2 meters above the ground. The filter activity has been measured using a low background total beta counter and a $^{90}\text{Sr}/\text{Y}$ reference standard for determination of equipment detection efficiency [5, 6].

The counter background was between 2.5 and 6 counts per minute. The sampling time interval was 5 hours. The filters have been measured immediately after sampling (3 min. after), for 1000 seconds and after 20 hours, for 3000 seconds [14].

The first two measurements provide filter activities necessary for the determination of the radon and thoron progeny concentrations and the last measurement indicates the presence of artificial radioactivity in the atmosphere.

In 1996, ERMS Botoșani had a 4 aspiration program: 02:00–07:00 (aspiration A1), 08:00–13:00 (aspiration A2), 14:00–19:00 (aspiration A3) and 20:00–01:00 (aspiration A4). The measured data were monthly reported to the National Reference Radioactivity Laboratory from National Environmental Protection Agency-Bucharest.
2.2. RADON AND THORON PROGENY CONCENTRATIONS

The data used for this study are averaged daily values obtained from 4 aspirations, for both $^{222}$Rn and $^{220}$Rn progeny. The time series of daily data were used to develop the multiple linear regression model for both estimation and prediction of the progeny concentrations.

![Graphs showing variations in precipitations, atmospheric pressure, relative humidity, air temperature, wind speed, and progeny concentrations over months for the year 1996.](image)

Fig. 1 – $^{222}$Rn and $^{220}$Rn progeny concentrations and meteorological variables in 1996.

In the Figure 1, the monthly average values of $^{222}$Rn and $^{220}$Rn progeny concentrations and meteorological variables for the year 1996 are presented.
As can be seen in Figure 1, both $^{222}\text{Rn}$ and $^{220}\text{Rn}$ progeny are related with meteorological data. In case of $^{222}\text{Rn}$ progeny maximum concentrations were in winter time - January due to less pronounced quantities of precipitations. Also, when were large amounts of precipitations in September, concentration of $^{222}\text{Rn}$ and $^{220}\text{Rn}$ progeny has decreased [12, 4, 15].

The air temperature is the parameter with the greatest influence in the determination of progeny concentrations.

In case of thoron progeny, due to the small half-live (56s), there were no significant changes in concentrations in winter, the rest of concentrations being similar to that of radon.

2.3. ESTIMATION AND PREDICTION MODEL

In order to find the linear relationship between a dependent variable and several independent (or predictor) variables by using the multiple linear regression model it was studied the performances of estimations and predictions of radon and thoron progeny concentrations considering meteorological data [10].

Multiple linear regression model can be used for estimations and predictions, with statistical significance, only if the independent variables are not in a relation of collinearity (correlation coefficients close to 0.99 and larger) and if two or more predictor variables are not in a relation of multicollinearity - Variance Inflation Factor (VIF) must be <10 [9, 13, 11].

Collinearity and multicollinearity analysis was studied on meteorological data for 4 month of the year, each month being representative for each season, data that was used to estimate and predict radon and thoron progeny concentrations.

<table>
<thead>
<tr>
<th>Month</th>
<th>January</th>
<th>July</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr$</td>
<td>$Pr$</td>
<td>$Pr$</td>
</tr>
<tr>
<td>Urel</td>
<td>0.343</td>
<td>0.172</td>
</tr>
<tr>
<td>$Pa$</td>
<td>0.575</td>
<td>0.297</td>
</tr>
<tr>
<td>$Ta$</td>
<td>0.205</td>
<td>0.068</td>
</tr>
<tr>
<td>$V$</td>
<td>0.096</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Correlation matrices for January, April, July and October

Table 1
The correlation coefficients (R) and corresponding level of significance for the independent variables from Table 1, reveal a high correlation between relative humidity and air temperature (0.312 ÷ 0.492) with high level of significance (0.005 ÷ 0.086) and between precipitations and atmospheric pressure (0.230 ÷ 0.575) with p-value between (0.001 ÷ 0.213). From the results exposed in Table 1 can be seen that the independent variables can not generate a relation of collinearity.

The multicollinearity was determined by calculating the VIF for the data sets used in the regression analysis for estimations and predictions of 222Rn and 220Rn progeny concentrations. According to the values of VIF presented in Table 2, for all months that was investigated showed that multicollinearity is not present for none of the considered predictors.
The estimation of progeny concentrations is made by using the analytical expression of the multiple linear regression with the corresponding meteorological variables from the time period on which the regression is defined.

Multiple linear regression model assumes a linear relationship between the progeny concentration as dependent variable and the meteorological data as predictor variables [4]:

\[ y(t) = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \ldots + \beta_n x_n(t) \]  \hspace{1cm} (1)

where: \( y(t) \) is the dependent variable associated to predictor variables \( \{x_i(t)\}_{i=1,2,...,n} \)

\( \beta_0 \) is the regression constant and \( \beta_1, \ldots, \beta_n \) are the regression coefficients.

The performances of regression model are quantified by the following statistics:

- **R** – Multiple correlation coefficient. \( R \) is measuring the degree of correlation between \( ^{222}\text{Rn} \) or \( ^{222}\text{Rn} \) progeny concentrations and the ensemble of meteorological data.

- **R^2** – Squared multiple correlation coefficient. \( R^2 \) is quantifying the proportion of the variation of dependent variable that is explained by the independent variables.

- **F-Test** – This is a global test of significance for the ensemble of coefficients. Large values of the F-test provide evidence against the null hypothesis.

- **p-value** – The p-value characterizes the significance level. A p-value \(< 0.05\) indicates that the null hypothesis may be rejected.

- **SSR** – Sum Squared Residual quantifies the deviation between the progeny concentrations estimated or predicted by the multiple regression and the measured values of the respective concentrations.

The prediction is performed using analytical expression of regression obtained for a previous time period by keeping the regression coefficients constant and by introducing the meteorological data from the prediction interval.

3. **ESTIMATION AND PREDICTION OF \(^{222}\text{Rn} \) AND \(^{220}\text{Rn} \) PROGENY**

The estimation of progeny concentrations is made by using the analytical expression of the multiple linear regression with the corresponding meteorological variables from the time period on which the regression is defined.

Multiple linear regression model assumes a linear relationship between the progeny concentration as dependent variable and the meteorological data as predictor variables.
From the Figure 2 containing representative months for each season, it can be clearly seen, that the estimation values generated by the multiple regression model, are very close to the measured values. This means that the considered predictor variables may describe well the radon and thoron concentration dynamics in the respective months.

Making use of the dedicated software [12] and Excel, the regression coefficients are determined by the least square method. In the Table 3 the
regression statistics for the estimation of the $^{222}$Rn and $^{220}$Rn progeny concentrations for a representative month of each season, are presented.

Table 3
Regression statistics for $^{222}$Rn and $^{220}$Rn progeny estimation

<table>
<thead>
<tr>
<th>Month</th>
<th>Regression statistics for $^{222}$Rn progeny</th>
<th>Regression statistics for $^{220}$Rn progeny</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multiple R</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Jan '96</td>
<td>0.730</td>
<td>0.534</td>
</tr>
<tr>
<td>Apr '96</td>
<td>0.559</td>
<td>0.313</td>
</tr>
<tr>
<td>Jul '96</td>
<td>0.809</td>
<td>0.655</td>
</tr>
<tr>
<td>Oct '96</td>
<td>0.434</td>
<td>0.188</td>
</tr>
</tbody>
</table>

One may notice that the multiple correlation coefficients has the greatest values in July with R values equal to 0.809 with very high level of significance ($p<0.01$) for both radon and thoron progeny concentrations. This proves that the corresponding predictor variables are describing quite well the variability of radon progeny in this month.

The prediction is achieved by using the analytical expression of regression obtained for a previous time period, with predictor variables from the time interval on which the prediction is done [4]. The predictions of the progeny concentrations $[y]_{N+1}$ on the time interval $N+1$, were performed by using the following analytical expression:

$$[y]_{N+1} = \beta_0 + \beta_1 [x_1]_{N+1} + \beta_2 [x_2]_{N+1} + \ldots + \beta_n [x_n]_{N+1}$$

where: $[x_1]_{N+1} , [x_2]_{N+1} \ldots [x_n]_{N+1}$, are the values of independent variables in the time interval $N+1$.

From the Figure 3 where the prediction values generated by the multiple regression model are compared with measured values showed better results obtained for the first part of the representative months it can be clearly seen because of the changing of meteorological conditions from the prediction period considering the fact that the equation constants are calculated considering the meteorological data from the previous month. This means that the considered model for estimation and prediction may describe well the radon and thoron progeny concentrations dynamics in the respective months.
As for criteria for prediction performances, we have used the SSR and correlation coefficient (R) between predicted and observed progeny concentrations, the corresponding level of significance (p-value) and the relative error specific to each predicted concentration value.
Table 4
Regression statistics for 222Rn and 220Rn progeny prediction

<table>
<thead>
<tr>
<th>Month</th>
<th>222Rn progeny</th>
<th>220Rn progeny</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prediction characteristics</td>
<td>Prediction characteristics</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>p-value</td>
</tr>
<tr>
<td>February</td>
<td>0.562</td>
<td>9.2e-4</td>
</tr>
<tr>
<td>May</td>
<td>0.527</td>
<td>2.1e-3</td>
</tr>
<tr>
<td>August</td>
<td>0.443</td>
<td>1.2e-2</td>
</tr>
<tr>
<td>November</td>
<td>0.176</td>
<td>3.4e-1</td>
</tr>
</tbody>
</table>

The performances of predictions presented in Table 4 indicates that both radon and thoron progeny concentrations are being well simulated by the multiple linear regression model except the case of radon in November (R = 0.176), the explanation consisting in a high variation of meteorological conditions for the month of the prediction period – November and from the time period used to determine the equation for prediction – October.

4. CONCLUSIONS

There have been studied the statistical and physical aspects of the modeling of the 222Rn and 220Rn progeny concentrations using multiple linear regression with meteorological variables as predictors.

Because the multiple regressions was used for estimations of January, April, July and October in Figure 2, and the same months were used to predict the progeny concentrations in the next months, in Figure 3 are given the time distributions of the predicted air concentrations for February, May, August and November.

The regression statistics from Table 3 shows that for the radon and thoron progeny concentration estimation, the greatest values of multiple R with high level of significance, have been obtained in July due to the fact that radon having the lifetime comparable with the ventilation time of entire planetary boundary layer is very sensitive to the air temperature which induce the vertical transport of the air masses.

The analysis of the progeny concentration prediction has shown that limitations of the prediction procedure, is more pronounced in case of the 220Rn progeny concentrations prediction, because thoron progeny concentrations are much lower than those of radon progeny. Thoron progeny concentrations are more
sensitive to the variations of the dependent and independent variables used in the estimation and prediction procedures.

The prediction performances described in Table 4 showed that $^{220}\text{Rn}$ progeny concentrations are much more sensitive to the variability of the independent variables in comparison with $^{222}\text{Rn}$ progeny concentrations due to a much lower air concentrations.

This paper sustain the results and conclusions presented in a previous one [4], "Modeling the $^{222}\text{Rn}$ and $^{220}\text{Rn}$ progeny concentrations in atmosphere using multiple linear regression with meteorological variables as predictors", which used data from another location of România, Bacău.

Acknowledgments. Our colleagues from Environmental Radioactivity Monitoring Station Botoșani are acknowledged for their effort in sample collection.

REFERENCES

1. C.R. Hosler, Meteorological effects on atmospheric concentrations of radon ($\text{Rn}^{222}$), $\text{RaB}(\text{Pb}^{214})$, and $\text{RaC/Bl}^{214}$ near the ground, Monthly Weather Review, 94(2), 89–99 (1966).
4. Simion F., Cuculeanu V., Simion E., Geicu A., Modeling the $^{222}\text{Rn}$ and $^{220}\text{Rn}$ progeny concentrations in atmosphere using multiple linear regression with meteorological variables as predictors, Romanian Reports in Physics, nr. 2 (2013).
5. C.A. Baciu, Radon and thoron progeny concentration variability in relation to meteorological conditions at Bucharest (Romania), J. Environ. Radioact. 83(2), 171–189 (2005)
6. V. Cuculeanu, F. Simion, E. Simion, A. Geicu, Dynamics, deterministic nature and correlations of outdoor $^{222}\text{Rn}$ and $^{220}\text{Rn}$ progeny concentrations measured at Bacău, Romania. J. Environ. Radioact., ELSEVIER, nr. 102, 703–712 (2011)