Principal component data processing in radon metrology

Bronislaw Machaj,
Piotr Urbanski

Abstract A gauge for the measurement of radon and radon daughters concentration was tested in a radon chamber. Count rate distribution in time at the output of radiation detectors was measured and registered. The count rate distribution in time was then processed employing Principal Component Analysis (PCA) and the Root Mean Square Error (RMSE) of the count rate was investigated. It was found that PCA processing removes great part of count rate random fluctuations originating from radiation statistics, which is resulting in a decrease of the count rate random error and in random error of concentration. The RMSE of radon daughters concentration is about 3 times lower when "raw" results are PCA processed. Such decrease of error, without PCA signal processing, would require 9 times higher air flow through the air filter on which the radon daughters are deposited. In case of the measurements of the radon concentration the drop of the error is 2–3 times higher in case of long counting time.

Key words Principal Component Regression (PCR) • radon daughters measurement • radon measurement

B. Machaj[✉], P. Urbanski
Institute of Nuclear Chemistry and Technology,
Department of Radioisotopes Instruments and Methods,
16 Dorodna Str., 03-195 Warsaw, Poland,
Tel.: +48 22/ 811 06 55, Fax: +48 22/ 811 15 32,
e-mail: bmachaj@orange.ichtj.waw.pl

Received: 2 April 2001, Accepted: 16 July 2001

Introduction

An important parameter of any gauge for measurement of radon or radon decay products concentration in air is a measuring error of the gauge; it is connected with the minimum detectable radon or radon daughters concentration. Both the random error and the minimum detectable concentration are determined by the statistical fluctuations of measured signal. A method of principal component analysis (PCA) applied to the raw results of measurement is able to remove considerable part of random fluctuations of the signal. It can thus be expected that the use of the PCA method in a gauge for measurement of radon and radon daughters to process the raw signal will result in decrease of both the minimum detectable concentration and the measuring error. Measurements of the radon and radon daughters concentration in the radon chamber were carried out. The obtained raw results were then processed employing the PCA method.

Multivariate data processing

Generally, a multivariate regression model can be presented in the form [1, 5, 6]:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (1)$$

where: **Y** is the matrix with n rows and m_y columns representing m_y dependent variables, **X** is the matrix of independent variables with n rows and m_x columns representing n measured spectra in m_x "channels" in case of spectrometric measurements, or n time distributed count rates in m_x time intervals in case of count rates measured in successive time intervals, **B** is the matrix of regression coefficients with n rows and m_y columns, **E** is the matrix with n rows and m_y columns representing the residual error.

Principal component analysis (PCA) is a method of decomposition of the matrix **X** into the sum of outer products of vectors called loadings (**p**) and latent variables or scores (**t**) in the form:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1' + \mathbf{t}_2\mathbf{p}_2' + \ldots + \mathbf{t}_a\mathbf{p}_a' + \mathbf{E}_\mathbf{X} \tag{2}$$

where: a is the number of factors (number of principal components) used ($a<m_x$), $\mathbf{E}_\mathbf{x}$ is the matrix of residuals **X** not explained by the model. The **X** matrix is thus replaced by the sum of limited number of $\mathbf{t}_i\mathbf{p}_i'$ products (usually a few) containing useful information about dependent variables with the rejected $\mathbf{t}_i\mathbf{p}_i'$ with higher indices containing mainly random noise. Employing in Eq. (1) the reduced form of the matrix **X** as given in Eq. (2), one can get the Principal Component Regression (PCR) model.

Random errors of raw and PCA processed data

Random errors of radon daughters concentration

A series of the measurements of radon daughters concentration was carried out with a radon daughters monitor [2, 3]. The count rate distribution of alpha radiation against time from radon daughters deposited on an air filter were registered and are shown in Fig. 1.

Employing MATLAB software, the matrix of raw spectra, shown in Fig. 1, was replaced by a sum of two principal components according to Eq. (2). The results of such transformation of the raw spectra are shown in Fig. 2. As it can be easily seen from Fig. 1 and Fig. 2 the spectra after transformation are more “smooth” and a considerable part of statistic fluctuations is removed. Thus, the PCA data processing acts as a filter of random fluctuations [7]. To assess random errors of raw and PCA processed count rate time distribution, the Root Mean Square Error (RMSE) was computed from the relation:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}\left(x_i - \hat{x}\right)^2}{m}} \tag{3}$$

x_i – measured (raw) or PCA processed count rate, $\hat{x}$ – simulated count rate, m – number of measuring points

The computations of RMSE of PCA processed count rate with respect to simulated count rate (without fluctuations), Fig. 3, showed that RMSE of PCA processed count rate is approximately 3 times lower comparing to the RMSE of the “raw” count rate computed in a similar way. It can thus be expected that the random error of radon daughters measurement can be decreased if PCA data processing is applied to the raw (measured) count rate. To compare the measuring errors when the data are PCA processed with the errors without PCA processing the following procedure was applied.

Considering the matrix **Y** containing radon daughters concentration and the corresponding matrix **X** containing raw count rate time distribution here employed, the data were regressed to get regression coefficients **b** and the estimated radon daughters concentration $\mathbf{Y}_{\mathrm{est}}$. The PCA processing was carried out at $a=2$ (two principal components) that sufficiently well removes the random error (99.85% of **X** block and 98.62% of **Y** block variance is captured by the model). Then, the simulated count rates from Fig. 3 were randomized 100 times (according to Poisson distribution) and such rates were then processed in two ways:

1. By employing equations relating radon daughters and alpha potential energy concentration to the raw count rates of the type [3]:

$$\begin{aligned} A &= a_1N_1 + a_2N_2 + a_3N_3 \\ B &= b_1N_1 + b_2N_2 + b_3N_3 \\ C &= c_1N_1 + c_2N_2 + c_3N_3 \\ E &= e_1N_1 + e_2N_2 + e_3N_3 \end{aligned} \tag{4}$$

where: A, B, C, E – concentrations of ^{218}Po, ^{214}Pb, ^{214}Bi (^{214}Po) and potential alpha energy, a_1,a_2,...e_3 – coefficients, N_1,N_2,N_3 – raw count numbers at three time intervals. The RMSE error of obtained A, B, C, E concentrations was then computed according to Eq. (3) where x_i – concentration

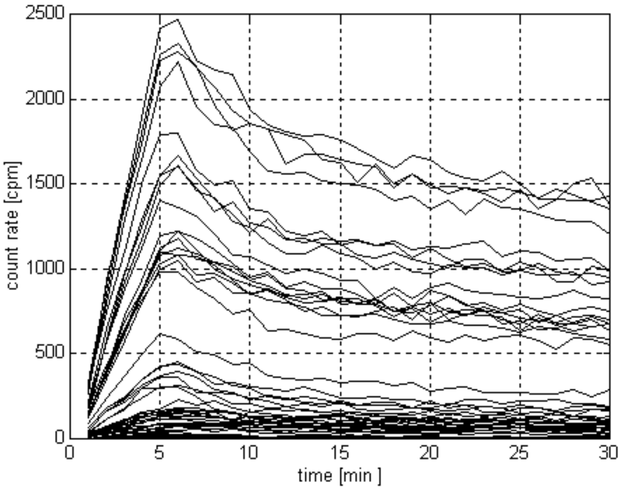

Fig. 1. Count rate distribution in time measured by means of a radon daughters monitor.

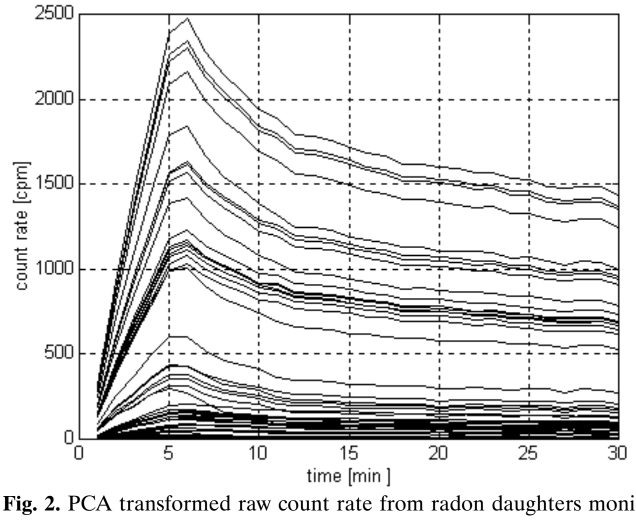

Fig. 2. PCA transformed raw count rate from radon daughters monitor.

Table 1. Average RMSE of radon daughters concentration from raw and PCA processed count rate spectra with simulated random fluctuations.

Measurement number	RMSE(A) (Bq/m^3)	RMSE(B) (Bq/m^3)	RMSE(C) (Bq/m^3)	RMSE(E) (μJ/m^3)
1–25 simulated raw	168	124	104	0.229
1–25 PCA processed	89.7	15.8	28.7	0.062
1–49 simulated raw	458	348.0	280	0.665
1–49 PCA processed	211	36.5	68.8	0.139

1–25 measurement number: A = 0–873 Bq/m^3, B = 0–797 Bq/m^3, C = 0–1046 Bq/m^3, E = 0.11–3.75 μJ/m^3;
1–49 measurement number: A = 0–11 730 Bq/m^3, B = 0–9180 Bq/m^3, C = 0–10 460 Bq/m^3, E = 0.11–53.7 μJ/m^3.
A – ^{218}Po, B – ^{214}Pb, C – ^{214}Bi, E – alpha potential energy.

A, B, C, or E from Eq. (4), $\hat{x}$ – estimated concentration A, B, C or E, m=100 numbers of simulations. The RMSE is given in Table 1 as “simulated raw”.

2. Radon daughters concentration was computed from the relation:

$$\mathbf{Y}_{\mathrm{PRED}} = \mathbf{X}_{\mathrm{r}}\mathbf{b'} \tag{5}$$

where $\mathbf{Y}_{\mathrm{PRED}}$ – matrix of predicted radon daughters A, B, C and alpha potential energy E concentration, $\mathbf{X}_{\mathrm{r}}$ – matrix of randomized count rate spectra, **b** – regression coefficients obtained earlier. The RMSE of radon daughters and alpha potential energy concentrations obtained in such a way was computed in respect to the estimated concentration A, B, C, E. The RMSE is given in Table 1 as “PCA processed”.

An average RMSE for the two groups of count rate time distribution are given in Table 1, corresponding to low (approx. 0–1000 Bq/m^3) and high (approx. 0–10 000 Bq/m^3) radon daughters concentration. It can be seen from Table 1 that random error due to the statistic fluctuations has considerably decreased when the PCA processing is employed. The random error of radon daughters and alpha potential energy concentration is on the average more than three times lower when PCA processing is applied to the raw data, when compared to the errors when “raw” count rates are used for computation of the concentrations.

Table 2. Average RMSE of radon concentration from raw and PCA processed count rate spectra with simulated random fluctuations.

Measurement number	RMSE (1–180 min) (Bq/m^3)	RMSE (161–180 min) (Bq/m^3)
1–10 simulated raw	88.9	146
1–10 PCA processed	41.0	138
1–21 simulated raw	216	304
1–21 PCA processed	79.4	309

1–10 measurement number: radon concentration = 405–8790 Bq/m^3;
11–21 measurement number: radon concentration = 9590–45 350 Bq/m^3

Random error of the radon concentration

Another set of measurements of radon concentration in air was carried out in the radon chamber, employing the Lucas cell as the detector of alpha radiation originating from radon and radon decay products. A series of count rate from the Lucas cell against time was measured. The Lucas cell of the size ϕ54×74 mm was used in these investigations [4]. Radon concentration, y, was computed from the relation:

$$y = \frac{N}{60}\frac{1000}{v}\frac{1}{3\varepsilon k} \qquad [\mathrm{Bq/m^3}] \tag{6}$$

where N (p/min) – mean count rate measured in the period 161–180 min after the radon had been introduced into the Lucas cell, ν=0.17 (l) – Lucas cell volume, ε(p/dis) – alpha detection efficiency of the Lucas cell, k – coefficient taking into consideration the decrease of radon activity at the time of pulse counting and due to the incomplete radiation equilibrium in the Lucas cell. The measured (raw) count rate time distribution are shown in Fig. 4.

It was found that the raw count rate time distribution of Lucas cell shown in Fig. 4 can be replaced by the first principal component model (99.45% of **X** block variance and 99.93% of **Y** block variance is captured by such PCR model) according to Eq. (2). The count rate distribution of the Lucas cell represented by the first principal component is shown in Fig. 5. Also in the case of the radon concentration measurement the fluctuations of the PCA processed count rate distribution are lower than of the “raw” (measured)

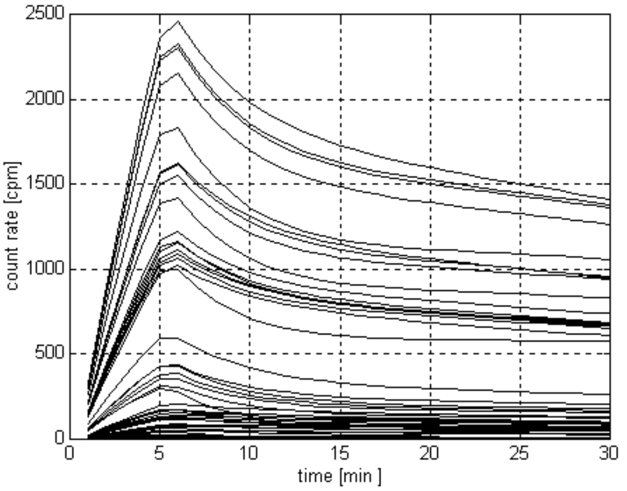

Fig. 3. Simulated count rate from radon daughters monitor.

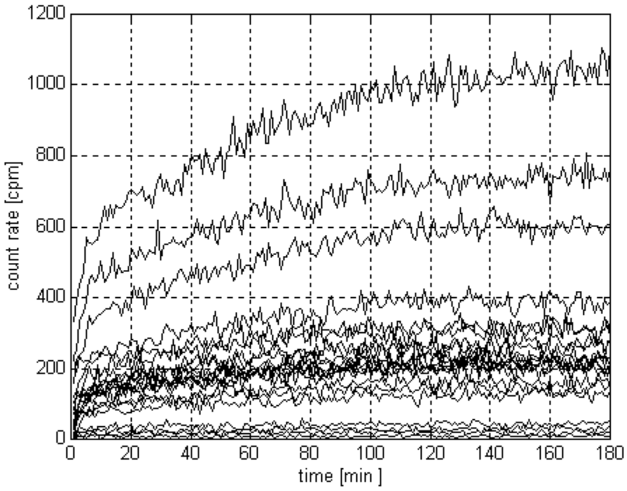

Fig. 4. Count rate distribution in time measured (raw) by means of the Lucas cell.

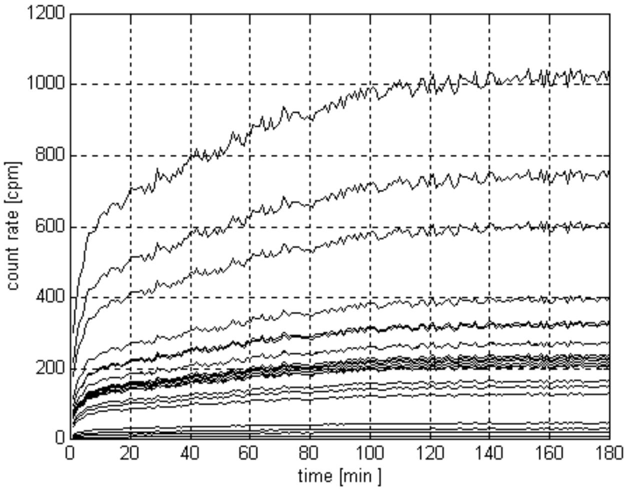

Fig. 5. PCA transformed count rate spectra from the Lucas cell.

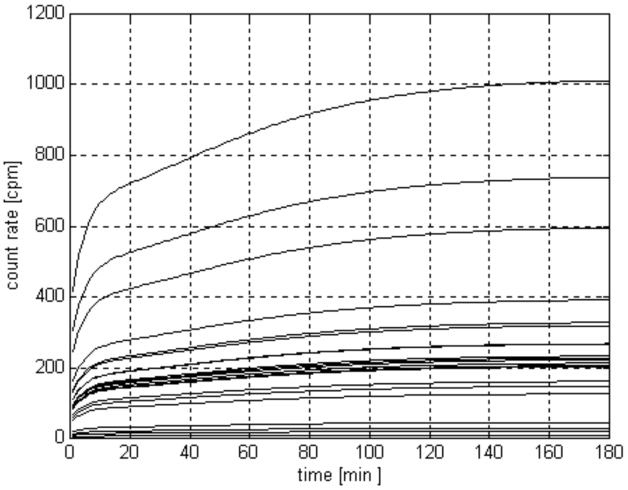

Fig. 6. Simulated count rate from the Lucas cell.

spectra. In both cases the RMSE of count rate was computed with respect to simulated (without fluctuations) count rates according to Eq. (3), Fig. 6. To estimate the expected decrease of random error of the radon concentration, the following computations were carried out.

Employing the matrix **Y** containing radon concentration for different samples of radon laden air and the corresponding matrix **X** containing raw count rate distribution at the time interval 1–180 min, the both variables were then regressed in order to get regression coefficients **b** and estimated radon concentration $\mathbf{Y}_{est}$. The data were processed employing count rate distribution **X** in the time range 1:180 min, that gave the regression coefficients $\mathbf{b}_1$ and the corresponding estimated radon concentration $\mathbf{Y}_{est1}$, then in the range of time 161:180 min that gave the regression coefficients $\mathbf{b}_2$ and estimated radon concentration $\mathbf{Y}_{est2}$. Next, the simulated count rate spectra shown in Fig. 6 were randomized 100 times and such spectra were then processed in two ways:

1. Radon concentration was computed according to Eq. (6) from randomized count rate for the period 161–180 min (coefficient $k=0.976$), and in the period 1–180 min ($k=0.846$). The RMSE was computed from the achieved, in such a manner, radon concentration with respect to the values given in matrix $\mathbf{Y}_{est1}$ and $\mathbf{Y}_{est2}$. The results of computations are given in Table 2 as "simulated raw".

2. Radon concentration $\mathbf{Y}_{PRED}$ was computed from the relation (5) from the matrix **X**, which contained randomized count rate at the time interval 1:180 min and 161:180 min. Then, RMSE and average RMSE for two groups of the radon concentration was computed and is given in Table 2 as "PCA processed".

Comparison of the results of computations shows the reduction of the RMSE by a factor approx. 3 for the counting time 1:180 min and indicates no improvement for counting time 161–180 min.

Conclusions

The Principal Component Regression based on Principal Components Analyses (PCA) applied to the raw count rates from the alpha radiation detector forming a part of radon daughters monitor, is improving considerably accuracy of the monitor. The random error of measurement due to statistical fluctuations of count rates decreases approximately three times. The minimum detectable radon daughters concentration is also decreased by the same degree. In the case of the Lucas cell as radon detector, the improvement of the accuracy is approximately 3 times higher when the PCA method is used for processing raw count rate time distribution obtained in counting period of time 1–180 min since radon sample is introduced into Lucas cell.

The multivariate data processing is based on the set of count rates measured in a fixed time period. This fact limits the application of such processing to the cases where the calibration of radon or radon daughters can be performed using such spectra. In the case of a continuous measurement of the radon concentration such processing cannot be used.

The price that has to be paid for the reduction of the random error employing Principal Component Data Processing is that the data from the measuring head have to be measured every minute and slightly more sophisticated processing has to be applied to the data from a measuring head. As majority of the present gauges are equipped with microprocessor systems this is no serious problem.

References

1. Geladi P, Kowalski BR (1986) Partial least squares regression: a tutorial. Anal Chim Acta 185:1–17
2. Gierdalski J, Bartak J, Urbanski P (1993) New generation of the mining radiometers for determination of radon and its decay products in the air of underground mines. Nukleonika 38;4:27–32
3. Machaj B (1999) Modification of the RGR monitor of radon daughters concentration in air. Nukleonika 44;3:479–490
4. Machaj B, Urbanski P (1999) Continuos measurement of radon concentration in the air with Lucas cell by periodic sampling. Nukleonika 44;4:579–594
5. Martens H, Naes T (1991) Multivariate calibration. Wiley & Sons, Chichester
6. Rencher AC (1996) Multivariate statistical inference and application. John Wiley & Sons, New York
7. Wold S, Anti H, Lindgren F, Ohman J (1998) Orthogonal signal correction of near-infrared spectra. Chemometr Intell Lab Syst 44:175–185