Environmental Impact Assessment Regarding Indoor Air Quality Using Statistical Methods

ELENA BUCUR, ANDREI VASILE, LUOANA FLORENTINA PASCU, CAROL BLAZIU LEHR, GABRIELA GEANINA VASILE*
National Research and Development Institute for Industrial Ecology - ECOIND, 71-73 Drumul Podul Dambovitei Str., 060652, Bucharest, Romania

This paper brings into attention of the indoor air specialists new information regarding the assessment of the potential cumulated impact of the air chemical compounds and microclimate factors on materials; for exemplification it was selected a wooden church made by oak and the monitoring values for temperature, humidity and the concentration of four chemical compounds with a destructive potential on organic materials generally: NO_2, SO_2, O_3 and $PM_{2.5}$.

Keywords: *binary logistic regression, cumulative impact, indoor air*

The interest for statistical methods designed for data series interpretation has increased over the time, correlated with the need of processing the scientific information and observations; at the moment there is almost no area where statistical methods do not find practical applicability. Regardless of the application, the statistic operates with various distributed data bases, interdependent or not, as is the case of the data series obtained in the long-term air quality monitoring of both indoor and outdoor.

Scientific literature abounds in definitions for the indoor air quality (IAQ) relative to maximum admissible concentrations; in the acceptance of ASHRAE (American Society of Heating, Refrigeration and Air-Conditioning Engineers), for example, *acceptable air quality* that implies air pollutants concentrations below the maximum permissible limit values for health protection, only if at least 80 % of indoor space occupants are satisfied with the quality of indoor air [1].

We have seen an increased interest in indoor air quality, over the last few years, other than workplace atmospheres; this is due to: 1) awareness of the effects of building sealing measures and the excessive use of cleaning, maintenance and ambient products that have led, among other things, to an increase in concentrations and a greater diversification of chemical, physical or biological pollutants in the air with effect on the human health [1-9] and 2) the growth of indoors time spent; studies conducted in recent years it have shown that we spend 80-90% of the total time indoors. Generally, air quality is appreciated by reference to the limit values set by the World Health Organization (WHO) or other national/international regulations for each parameter. In terms of limits, a much discussed topic is whether only the reference to the limit value is enough and ensures a correct interpretation of the possible effect of the air pollutants on the materials in contact with the environment.

We are trying to provide a possible answer to this question in this paper, analyzing how the regression can bring more information in the air quality assessment process. Statistical regression is increasingly used in recent years in various fields of activity: Morrison et al. [10] established the causality relationship between the incidence of respiratory diseases and the air quality, using the correlations found between the soil metal concentration and the incidence of respiratory diseases, Danciulescu et al. [11] identified the correlation between the concentration of gaseous compounds in ambient air and the noise level in urban areas, and Yoo et al. [12] analyzed the influence of precipitation intensity on the quality of air at the soil level in South Korea; linear correlation and regression were used by Kliuncininkas et al. [13] to identify the major sources of air pollution inside Kaunas, Lithuania with particulate matter and PAHs, while Bucur et al. [14] applied the binary logistic regression to identify and hierarchize the pollutants present in a museum depending on the potential impact on the exhibits.

Experimental part

In order to exemplify how statistical techniques can complete the image of an indoor air quality assessment study, we used a series of gas pollutant monitoring data and microclimate parameters obtained in a study conducted in a wooden church during the period 2014-2015; information about the measurement methods are presented in the previous article [15].

Materials and methods

In order to have a better quality of the data, we wanted to improve the characterization of the monitoring data series by verifying the normality of the distribution; for this purpose, the values of skewness and kurtosis have been determined. For a normal distribution, the value of the asymmetry coefficient (skewness), which varies in the range [-1; 1], should be tilted to 0 and the value of the kurtosis coefficient to 3 [16].

Secondly, we have proposed a more in-depth analysis of the potential effect of the church environment on the wood it is made of and the wooden objects (wood and furniture icons) located inside, analyzing the data series using statistical regression.

In the paper, the binary logistic regression was used to assess the cumulative potential effect and to calculate the probability *(P)* that the wood from which the church is made to be affected by the indoor environment, respectively the microclimate (temperature and humidity) and the concentration of the chemical pollutants present in the indoor air. In the binary logistic regression, the dependent variable *Y* is categorical, with only two categories: *YES* and *NOT* expressed numerically by 1 and 0; the value 1 is usually attributed to the possible variant (producing the event) and 0 to the opposite variant (the event does not occur). Thus, if the event studied is *the damage to the wood from which the church is built*, based on a predetermined classification

* email: gabriela.vasile@incdecoind.ro; Phone: (+40)720048195

criterion, we will assign the number 1 to the values that can affect the wood and 0 to the values that do not show this effect.

In our case, if we mark with p the probability that the wood will be affected by the environment and with *(1-p)* the probability that the wood will not, the ratio p / 1-p represent the probability ratio *(odds_ratio)* and it is calculated with the formula (1):

$$odds\_ratio = \frac{p}{1-p} \quad (1)$$

Unlike the linear regression in which the value of the dependent variable is obtained on the basis of predictive variables, using logistic regression we obtain the value of a transformation of the dependent variable, marked with *logit (p)*, relation (2), respectively the logarithm of the *chance ratio:*

$$logit(p) = \ln\frac{p}{1-p} \quad (2)$$

In the case when we have k predictor variables, the equation of the general model of regression will be [17]:

$$logit(p) = \ln\frac{p}{1-p} = \beta_0 + \beta_1 x_1 \dots \dots + \beta_k x_k \quad (3)$$

when p is $P(y=1|x_1,x_2,\dots,x_k)$ and represents the probability that the wood can be affected if the predictor variables $x_1,x_2,\dots,x_k$ are taken into consideration and can be calculated [17] with the formula (4):

$$P(y=1|x_1,x_2 \dots \dots x_k) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\dots\dots+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\dots\dots+\beta_k x_k}} \quad (4)$$

In order to calculate the probability, it is necessary to: i) determine the coefficients β_i, ii) establish the parameters that can significantly influence this process and iii) simplify the relationship by eliminating the parameters that have insignificant influence.

Statistical tests were performed using the SPSS 20.0 program (Statistical Package for Social Sciences 20.0) and the Forward LR method applied for the logistic regression, which allows for a single run to identify the prediction variables with statistical significance and the logit coefficients βi.

Results and discussions

Characterization of data series

The results of monitoring process, respectively the indicators of the central tendency, the dispersion and the extreme values are summarized in table 1; in the same table are also presented the results of the statistical tests for verifying the normality of the data series distribution (skewness and kurtosis) and the limit values recommended by ASHRAE [18] for the indoor measurements, for: NO_2, SO_2, CO, CO_2, O_3, $PM_{2.5}$, temperature and humidity.

Analysing the data obtained, we observe that for indoor air, the concentrations of the monitored indicators are, on average, below the recommended ASHRAE values, with the exception of ozone whose average exceeds the recommended limit by approximately 1.5%. Maximum values exceeding the recommended limits for NO_2, SO_2, CO, CO_2 and $PM_{2.5}$ occur on days when church services are running and return to normal shortly after the end of the job. From the point of view of the microclimate parameters, during the analysed period, the wood from which the church is built ensures the values that do not pose a special risk for the long-term preservation of the interior objects and the building itself if we take in consideration the recommended interval for humidity (RH <75%). However, studies conducted in recent years have shown a particular sensitivity of wood to microclimate conditions, especially humidity; it is recommended, for preventive preservation, to maintain wood objects at a maximum of 55% humidity.

We can therefore consider the humidity as a parameter that can affect the wood from which the church is made. The average of humidity exceeds the recommended value by approximately 3.6%.

The values of statistical tests that characterize the distribution normality (skewness and kurtosis) have shown deviations from the normal distribution for all data series, which should be taken into account for the regression analysis.

The use of binary logistical regression in assessing the cumulative effect of the environment

For the logistic regression, the following parameters were considered as predictors' variables: concentrations of NO_2, SO_2, O_3 and $PM_{2.5}$, temperature and humidity, and as a dependent (categorical) variable *the effect of the environment on wood*.

The criterion for separating the values of the dependent variable was the reference to the limit values recommended by ASHRAE [18] for general collections (20μg/m^3 for NO_2; 5.7μg/m^3 for SO_2; 10μg/m^3 for O_3; 10μg/m^3 for $PM_{2.5}$; 25^{0}C for temperature and 55% for humidity). Under these conditions, the following values were assigned to the dependent variable:

- 1, *possible effect,* if at least one of the indicators values exceeds the recommended limit value;
- 0, *no effect*, if the values of all indicators are below the recommended limit values;

The equation of the mathematical model for the estimation of the cumulative effect is rendered by the eq. (3) and the probability of effect on the wood from which

Table 1

RESULTS OF AIR QUALITY MONITORING INSIDE THE WOODEN CHURCH AND VALUES OF THE INDICATORS FOR DATA SERIES CHARACTERIZATION

	I- NO_2 μg/m^3	I-SO_2 μg/m^3	I-O_3 μg/m^3	I-CO mg/m^3	I-CO_2 ppm	I- $PM_{2.5}$ μg/m^3	I-temp ^{0}C	I-RH %
No. of measurements	143	143	143	143	143	13**	143	143
Average	**14.05**	**5.33**	**10.14**	**0.289**	**462.2**	**9.01**	**23.27**	**56.89**
Median	9.049	4.95	9.93	0.251	435.4	8.6	23.11	56.83
Standard deviation	14.26	1.61	0.586	0.155	110.7	2.06	1.43	2.359
Minimum	3.32	2.89	9.38	0.16	396.5	6.7	20.72	50.94
Maximum	75.20	14.86	11.67	1.32	1304.4	13.2	26.7	62.6
Skewness	2.69	2.68	0.857	4.46	5.16	1.16	0.473	0.086
Kurtosis	6.72	11.93	-0.336	22.81	30.31	0.620	-0.290	-0.272
ASHRAE*	4-20	1-5.7	1-10			1-10	<25	25-75

*ASHRAE Handbook – Chapter 21, 2007 for general collections, C control class for microclimate parameters;** Daily average;

the church is made shall be calculated using the relationship (4), after determining and replacing in equations the values of coefficients β_i corresponding to the parameters with statistical significance. For the determination of the coefficients β_i, the series of data obtained from monitoring were processed using binary logistic regression, LR Forward method. From a total of 143 data combinations obtained, 129 were used to calibrate the mathematical model, the rest of 14 combinations representing the quantity required for validation of the model (10 % data combinations). The binary logistical regression analysis is carried out in two steps. In the initial step (Block 0), a prediction is performed only based on the constant (the base model), the *logit coefficients* corresponding to the variables being considered null; these coefficients are taken into account in the next step (*Block 1*). This type of model, built only with the constant, provides useful information in the next step for testing logistic models developed using predictor variables (Omnibus Test, Hosmer and Lemeshow Test). Table 2 summarizes the results for Block 1; the analysis was finalized after the introduction of three variables with statistical significance in the analysis, only 3 steps being necessary.

Classification table shows the correspondence between the entered values of the dependent variable and the predicted values, a quality measure of the fitting obtained from the percentage value of the correct classification, Overall Percentage, of 97.7%. This value can be assimilated with the accuracy of the prediction.

The table *Variables in the Equation* provides information about the logit coefficients corresponding to the variables used to make the prediction at this stage: the value of the coefficient for each predictor variable *(B)* for each analysis step, the value of the *Wald Test* for the appreciation of statistical significance for coefficients (Wald column) and the level of significance, which in SPSS is denoted with Sig; unlike the Enter method in which the variables with statistical significance are identified based on the value of Sig. which must be <0.05 for a confidence level of 95%, for the Forward LR method the program operates also the comparisons, and the table *Variables in the Equation* shows only prediction variables with statistical significance.

In our case NO_2, RH and O_3. A *logit coefficient* is *statistically significant* if a variation equal to the unit of the corresponding predictor variable results in a significant change in the value of the dependent variable; if a variable has no statistical significance, it can be removed from the model without any particular implications on the accuracy of the prediction.

The *Model Summary Table* shows the values of the variability indicators, *Pseudo-R^2* Cox & Snell and Nagelkerke coefficients. Because it is universally acknowledged that the Cox & Snell coefficient underestimates the real value, we used the Nagelkerke coefficient to analyse the contribution of the six predictor variables to the variability of the dependent variable. As shown in the table, the six predictive variables (NO_2, SO_2, O_3, $PM_{2.5}$, temperature and humidity) explain to a great extent (92.9%) the effect of the environment on materials in contact with the indoor air.

In order to estimate the prediction efficiency of the model we applied a couple of tests: Omnibus and Hosmer and Lemeshow, whose results presented in table 3 indicate an improvement in prediction power by introducing NO_2 and O_3 parameters compared to the first stage based only on RH.

Starting from the values of the logit coefficients, *B*, corresponding to the variables with statistical significance (NO_2, RH and O_3), the equation of the model is the form of the equation (5) and the cumulative probability of the significant environmental parameters can be calculated with the relation (6):

Table 2

RESULTS OF BINARY LOGISTIC REGRESSION - FORWARD LR (BLOCK 1)

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	55.749[a]	0.595	0.807
2	31.179[b]	0.665	0.902
3	23.448[c]	0.684	0.929

a. Estimation terminated at iteration number 8; parameter estimates changed by less than .001.
b. Estimation terminated at iteration number 9; parameter estimates changed by less than .001.
c. Estimation terminated at iteration number 11; parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted: Effects on wood .00	Predicted: Effects on wood 1.00	Predicted: Percentage Correct
Step 1	Effects on wood	0.00	47	3	94.0
		1.00	5	74	93.7
	Overall Percentage				93.8
Step 2	Effects on wood	0.00	49	1	98.0
		1.00	2	77	97.5
	Overall Percentage				97.7
Step 3	Effects on wood	0.00	49	1	98.0
		1.00	2	77	97.5
	Overall Percentage				97.7

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	RH	2.355	0.476	24.489	1	0.000	10.539
	Constant	-129.227	26.084	24.545	1	0.000	0.000
Step 2[b]	RH	2.573	0.599	18.450	1	0.000	13.104
	O_3	4.356	1.485	8.609	1	0.003	77.969
	Constant	-184.333	41.136	20.079	1	0.000	0.000
Step 3[c]	NO_2	0.457	0.261	3.076	1	0.079	1.580
	RH	3.625	1.117	10.534	1	0.001	37.526
	O_3	6.978	2.684	6.760	1	0.009	1072.507
	Constant	-272.614	86.691	9.889	1	0.002	0.000

a. Variable(s) entered on step 1: RH;
b. On step 2: O_3;
c. Variable(s) entered on step 3: NO_2.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	116.507	1	0.000
	Block	116.507	1	0.000
	Model	116.507	1	0.000
Step 2	Step	24.570	1	0.000
	Block	141.077	2	0.000
	Model	141.077	2	0.000
Step 3	Step	7.731	1	0.005
	Block	148.808	3	0.000
	Model	148.808	3	0.000

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	26.688	8	0.001
2	122.112	8	0.000
3	3.264	8	0.917

Table 3
RESULTS OF THE MODEL PREDICTION POWER PREDICTION (BLOCK 1)

Table 4
RESULTS OF THE CALCULATION OF THE PROBABILITIES OF EFFECT AND THE IMPACT ON WOOD

NO_2, μg/m^3	SO_2, μg/m^3	Temp, ^{0}C	RH, %	O_3, μg/m^3	$PM_{2.5}$, μg/m^3	P, %	Impact
4.60	4.55	24.32	54.73	9.96	8.80	53.2	moderate
10.61	5.06	23.62	54.86	9.86	6.70	50.1	moderate
9.02	4.67	23.70	54.89	9.81	6.70	46.22	moderate
10.43	4.52	24.01	55.01	9.87	8.66	60.7	moderate

$$\ln(odds_{ratio}) = 6.978x_{O3} + 3.625x_{RH} + 0.457x_{NO2} - 272.614 \quad (5)$$

$$P = \frac{e^{6,978x_{O3}+3,625x_{RH}+0,457x_{NO2}-272,614}}{1+e^{6,978x_{O3}+3,625x_{RH}+0,457x_{NO2}-272,614}} \quad (6)$$

The probability value can also be calculated using the SPSS 20.0 program; this feature was applied to validate the mathematical model by running the 14 unused data combinations in the model calibration process. Following the validation process, all results were correct. We can appreciate in these conditions that the predictive mathematical model developed by binary logistic regression analysis of monitoring data can be used with confidence for calculating the likelihood of effect on the wood from which the church and wooden objects are made in this space. Starting from the values calculated for probabilities, *P*, Bucur et al [15] assimilated this value with the cumulative impact of the environment on exhibits in a museum; for easy interpretation they proposed an impact rating in five classes: i) very low impact for $0 < p < 0.2$; ii) low impact on $0.2 < p < 0.4$; iii) moderate impact on $0.4 < p < 0.6$; iv) strong impact for $0.6 < p < 0.8$ and v) very strong impact for $p > 0.8$.

Applying this procedure to our study we selected four pollution situations in which all the parameter values are below the recommended limit values, we calculated with the relationship (6) the value of the effect probabilities on the wood from which the church is made and we included the context of pollution in the class of the corresponding impact (Table 4).

It can be noticed that although all the values of the parameters are below the recommended limit values, the impact may be moderate or even strong, which requires measures to reduce the level of pollution and the values of the microclimate parameters. Thus, a clean air supply that dilutes 5% of indoor air will cause a strong drop effect and an impact shift from strong to very low (P = 0.1%) for the fourth pollutant context shown in table 4.

Conclusions

Based on the results we can appreciate the importance of the statistical interpretation of the monitoring data has in the process of assessing the air quality and environmental impact inside the building on the materials from which the building or objects of interest are built. By applying the binary logistic regression to the monitoring data series, it is demonstrated that: i) we can identify the pollutants with significant effect, 2) we can quantify their effect on the materials by calculating the probability, *P*, and last but not least 3) appreciated a possible cumulative impact of the environment on materials of interest.

Acknowledgements: This study was accomplished through the Nucleu Research Project, developed with the MCI support, Agreement no. 38N/2018, project no. PN – 18 05 02 02; the authors would like to thank all those who contributed to this study.

References

1.ASHRAE, Museums libraries and archives, 2007, Chapter 21.
2.ROMAGNOLI, P., BALDUCCI, C., PERILLI, M., GHERARDI, M., GORDIANI, A., GARIAZZO, C., GATTO, M.P., CECINATO, A., Atmos. Environ., 92, 2014, p. 51.
3.DESTAILLANTS, H., MADDALENA, R.L., SINGER, B.C., HODGSON, A.T., MCKONE,T.E., Atmos. Environ., 42, no. 39, 2008, p. 1371.
4.HOREMANS, B., VAN GRIEKEN, R., Atmos. Environ. 44, 2010, p. 1497.
5.LEE, C-W., HSU, D-J., 2007, Atmos. Environ., 41, 2007, p. 6598.
6.DIODIU, R., GALAON, T., BUCUR, E., STEFAN, D.S., PASCU, L.F., Rev. Chim. (Bucharest), **67**, no. 12, 2016, p. 2396.
7.DIODIU, R., GALAON, T, BUCUR, E, PASCU, L.F., Rev. Chim. (Bucharest), **67**, no. 8, 2016, p.1466;
8.BANCIU, A.R., NITA-LAZAR, M., NICULESCU, L.D., 20th International Symposium the Environment and the Industry, 2017, p.248
9.BRATU, M., TANASE, G., DANCIULESCU, V., NICOLESCU, I., PETRESCU, M., 20th International Symposium the Environment and the Industry, 2017, p.182.
10.MORRISON, S., FORDYCE, F.M., MARIAN SCOTT, E., Environ Geochem Helth., 36, 2014, p. 319.
11.DANCIULESCU, V., BUCUR, E., PASCU, L.F., VASILE, A., BRATU, M., J. Environ. Prot. Ecol., 16, no. 3, 2015, p.815.
12.YOO, J.M., LEE, Y.R., KIM, D., JEONG, M.J., STOCKWELL, W.R., KUNDU, P.K., OH, S.M., SHIN, D.B., LEE., S.J., Atmos. Environ., 82, 2014, p. 226.
13.KLIUCININKAS, L., KRUGLY, E., STASIULAITIENE, I. RADZIUNIENE, I., PRASAUSKAS, T., JONUSAS, A., KAUNELIENE, V., MARTUZEVICIUS. D., Atmos. Environ., 97, 2014, p. 83.
14.BUCUR E., DANET, A.F., LEHR, C.B., LEHR, E., NITA-LAZAR, M., J. Air Waste Manag. Assoc., 67, no. 4, 2017, p. 391.
15.BUCUR E., VASILE, A., DIODIU, R., CATRANGIU, A., PETRESCU, M., J.Environ. Prot. Ecol., 16, no. 1, 2015, p.7.
16.JABA, E., GRAMA, A., Statistical analysis with SPSS under Windows, Polirom Publishing House, Iasi, 2004, in Romanian.
17.POPA, M., Multivariate statistics applied in psychology, Polirom Publishing House, Iasi, 2010, in Romanian.

Manuscript received: 15.07.2018