

DOI: <http://doi.org/10.21698/simi.2018.ab28>

STATISTICAL METHODS IN AIR POLLUTION CONTROL

Elena Bucur, Andrei Vasile, Luoana Florentina Pascu, Carol Blaziu Lehr, Gabriela Vasile

National Research and Development Institute for Industrial Ecology - ECOIND, 71-73 Drumul Podu Dambovitei, district 6, 060652, Bucharest, poluare.aer@incdecoind.ro, Romania

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

Introduction

This paper brings to the attention of the specialists from the indoor air quality domain, new information's regarding the assessment of the possible cumulative impact of the microclimate factors and air pollutants, on the materials; for example was chosen an oak wooden church, and also the results of the temperature, humidity and concentration of four chemical compounds with destructive potential on organic materials monitoring, as NO₂, SO₂, O₃ and particulate matter PM_{2.5}. Generally, air quality is appreciated by referencing to the World Health Organization (WHO) limit values, or to the other National/International Regulations for each parameter. But the question is, if only the reference to the limit value is enough and ensures a correct interpretation of the possible cumulative effect of the air pollutants on the materials, in contact with the environment.

We are trying to provide an answer and also a solution in this paper, to the mentioned question and how the regression can bring more information in the indoor air quality assessment process.

Materials and methods

In this paper, the binary logistic regression was used for assess the cumulative potential effect and the probability (P) calculation that the wood from which the church was made, is affected by the indoor environment, microclimate (temperature and humidity) and the concentration of the chemical pollutants from the indoor air from the wooden church. In the binary logistic regression, the dependent variable Y is categorical, with only two categories: YES and NO, numerically expressed by 1 and 0. For the logistic regression, were considered as predictor variables NO₂, SO₂, O₃ and PM_{2.5} concentrations, temperature and humidity, and as a dependent categorical variable, the effect of the environment on wood.

The selection of the criterion for the values of the dependent variable was referenced to the value limit recommended by ASHRAE for general collections (20µg/m³ for NO₂; 5.7 µg/m³ for SO₂; 10µg/m³ for O₃; 10µg/m³ for PM_{2.5}; 25°C for temperature) and 55% for humidity. Under these conditions, the dependent variable was assigned with the following values:

1, possible effect, if at least one of the indicator values exceeds the recommended limit value;

0, no effect, if the values of all indicators are below of the recommended limit values.

In our case, if we note with p the probability that the wood was affected by the environment and with $(1-p)$ the probability that the wood was not affected, the ratio $\frac{p}{1-p}$ is the ratio of probabilities, and the general regression equation model will be:

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

where p , of the form $P(y=1|x_1, x_2, \dots, x_k)$ is the probability that the wood was affected if the predictor variables x_1, x_2, \dots, x_k are taken into consideration and can be calculated with the relation (2) :

$$P(y = 1|x_1, x_2 \dots \dots x_k) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}} \quad (2)$$

Statistical tests were performed using the SPSS 20.0 programme (Statistical Package for Social Sciences 20.0) and the Forward LR method was used for the logistic regression.

Results and conclusions

From the data obtained, related with the indoor air, we found that the the monitored parameters are situated on the medium values below the ASHRE recomaded values for general collections, less ozone values whose are over the recommended limit with approx. 1.5%

Based on the obtained results by applying the binary logistic regression, using the Forward LR monitoring method, in the data series, only three of the parameters were identified as statistically significant: NO₂, RH and O₃, the other three being eliminated from the model. Starting from the *logit* coefficient values corresponding to statistical significance variables, the equation of the model is the form of the (3) equation and the probability of the cumulative effect of the significant environmental parameters can be calculated with the relation (4):

$$\ln(\text{odds}_{ratio}) = 6,978x_{O_3} + 3,625x_{RH} + 0,457x_{NO_2} - 272,614 \quad (3)$$

$$P = \frac{e^{6,978x_{O_3} + 3,625x_{RH} + 0,457x_{NO_2} - 272,614}}{1 + e^{6,978x_{O_3} + 3,625x_{RH} + 0,457x_{NO_2} - 272,614}} \quad (4)$$

Considering the equation form, and also the most significant environmental parameters, the value, P, of probability, can also be assimilated as a potential cumulative impact of the environment on the materials.

In conclusion, we can appreciate the importance of the statistical interpretation of the monitoring data in the air quality assessment process by: i) characterizing the data series and the distribution, ii) identifying the pollutants with semnificative effect, iii) quantifying their effect on materials by calculating the probability, p , and last but not least iv) assessing the possible cumulative impact of the environment on the interest materials.