ANALYSIS OF POLYCYCLIC AROMATIC HYDROCARBONS (PAHs) IN SOIL BY SOFT INDEPENDENT MODELING OF CLASS ANALOGY (SIMCA)

Snežana Sremac[1], Nataša Đurišić[2], Biljana Škrbić[2], Antonije Onjia[1]

[1]The Vinča Institute of Nuclear Sciences, P.O.Box 522, 11001 Belgrade, Serbia
[2]Faculty of Technology, Bulevar Cara Lazara 1, 21000 Novi Sad, Serbia

1. Introduction

Increasing importance is being given to the presence of polycyclic aromatic hydrocarbons (PAHs) micropollutants in the environment, particularly since some are known to be carcinogenic. PAHs released into the environment arise mainly from anthropogenic sources. Due to strong adsorption of PAHs by soil particles, suitable methods for the measurement of PAH content in soil and subsequent data interpretation are extremely important. Several gas chromatography or high-performance liquid chromatography methods can be successfuly used for the determination of indiviudal PAH concentrations in soil samples, while the data interpretation is still underdeveloped. As the number of measured PAHs and samples increase, the examination of data becomes more difficult. To overcome such difficulties, numerous pattern recognition techniques have been introduced in environmental chemistry.

SIMCA (soft independent modelling of class analogy)[1] is a well known supervised pattern recognition method, which describes each class separately in a principal components (PC) space. The number of PCs used for each class is selected in such a way that they account for most of the variance within each class. New sample considered is assigned to one or more classes according to its Euclidean distance towards the constructed PC space. It can be member of one, more than one or none of the classes.

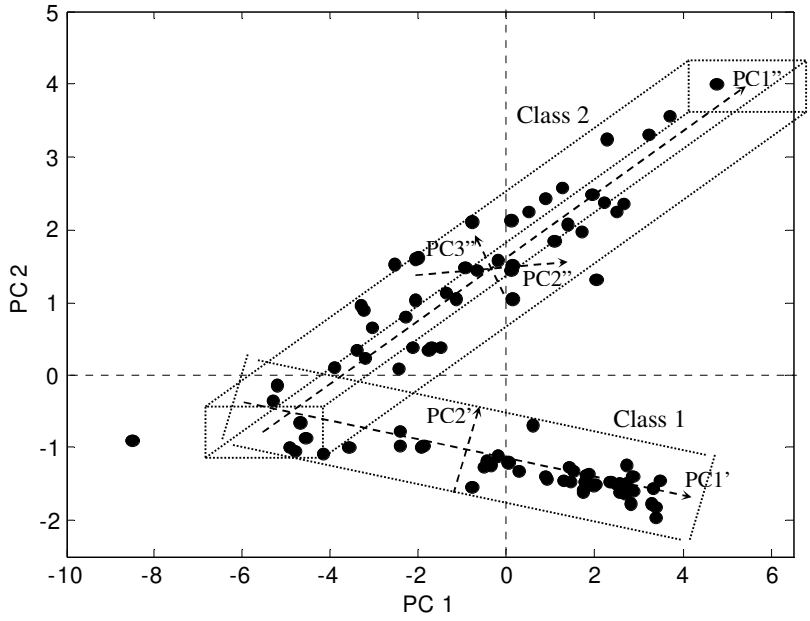
The SIMCA method in this work has been applied to the interpretation of the PAHs content in soil in different sites and from different times.

2. Experimental

The soil samples were collected at six locations near the Novi Sad oil rafinery (Serbia). Eleven PAHs (phenanthrene, anthracene, fluoranthene, pyrene, benzo(a)anthracene, chrysene, benzo(b+k)fluoranthene, benzo(a)pyrene, dibenzo(a,h)anthracene, benzo(g,h,i)perylene, and indeno(1,2,3,c,d)pyrene) were determined. PAHs were extracted in a Soxhlet apparatus and measured by a gas chromatograph with a flame ionization detector. This experimental procedure was given in details elsewhere [2]. The SIMCA data analysis in this work was made by means of PLS_Toolbox 2.1 software package.

3. Results and discussion

The measured data (concentrations of PAHs were in the range between 0.75 and 86.2 μg/g dry soil) were appended to the data found in literature [3 - 6], so that the concentration matrix of eleven PAHs (variables) in a total of ninety-four soil samples (observations) was used for multivariate modeling. Each soil sample was considered as an assembly of eleven variables forming a data vector. These vectors belonging to the same group were then termed a ‘class’. In the PCA space, the samples are scattered in two directions (Fig. 1). This distribution indicates that the data are suitable to be treated independetly. Therefore, SIMCA modeling should be used instead of traditional PCA. According to Fig. 1, the data are splitted into two classes: the data from ref. [5] and [6] in class 1 and the data from ref. [3] and [2] in class 2,



2 Fig. 1. Score plot of soil samples

while the data from ref. [4] are randomly distributed in both classes. In addition, some samples belong to both (or neither) classes simultaneously, hence there is an overlap region.

Each class was independently modelled using PCA in the way that each class could be described by a different number of PCs. Class 1 falls roughly on a plane, the axes of which are the first two PCs of this class, while the members of class fall mainly into a box, which is defined by the first three PCs of class 2. The variance explained by these PCs is shown in Table 1.

Table 1.Classification results using SIMCA

		No. PCs	Variance %	Error %
Recognition	Class1	2	90.5	0.00
	Class2	3	89.6	4.26
Prediction	Class1	2	86.8	6.82
	Class2	3	92.6	2.27

Using the assigned classification, the model was built and tested: (a) with all samples in both the training and testing set (recognition); (b) with approximately a half randomly chosen samples in the training set and the remaining samples in the testing set (prediction). In all cases the leave-one-out method of cross-validation was applied. The results of SIMCA analysis are shown in Table 1. In the SIMCA model, 100 and 95.7 % of the soil samples were correctly classified, and 93.18 and 97.7 % of the soil samples were correctly predicted.

4. Conclusion

SIMCA pattern recognition method applied to a chemical dataset provides information useful in obtaining a satisfactory classification of the soil samples according to their PAHs content. In this study, a good recognition and prediction success rate of the soil samples were obtained.

5. Reference

[1] S. Wold, Patt. Recog., 8 (1976) 127.
[2] B. Škrbić, N. Miljević, J. Environ. Sci. Heal. A 37 (2002) 1029.
[3] M. Malawska, B. Wilkomirski, 4th International Symposium and Exhibition on Environment in Central and Eastern Europe, Proceedings, 34 (CD), Warsaw, 1998.
[4] O. Franzle, J. Krinitz, W. Schmotz, T. Delschen, J. Leisner-Saaber, Report No. UBA-FB 95-048, Geographisches Institut der Universitat Kiel, 1995.
[5] K. Jones, J. Stratford, K. Waterhouse, N. Vogt, Environ. Sci. Technol. 23 (1989) 540.
[6] V. Linkeš, J. Kobza, M. Švec, P. Ilka, P. Pavlenda, G. Barančikova, L.Matuškova, Environment Monitoring of Slovak Republic for the Period 1992/96, Bratislava, 1997.