ODOURS IDENTIFICATION IN THE ENVIRONMENT BY MEANS OF MATHEMATICAL METHODS

Jindriska MATERNOVA, Jacques NICOLAS, Philippe ANDRÉ
Fondation Universitaire Luxembourgeoise, Avenue de Longwy, 185, 6700 Arlon (BELGIUM)
tél. +32-63-21 58 57, fax +32-63-21 58 00, e-mail nicolas@ful.ac.be

Serge VIGNERON
Société Belge de Filtration, Voie Minckelers 1, 1348 Louvain-la-Neuve (BELGIUM)

We test and compare some mathematical discrimination or pattern recognition methods in order to evaluate their ability for odour pollution source identification. Tested applications concern three categories of steel industry plants, tested methods are multiple linear regression, factor and cluster analysis, discriminant function analysis and neural networks. The concentrations of three compounds families constitute a pattern recognised by discriminant analysis and neural network as characteristic for each of the sources.

1. Introduction

Air pollution problems stem from many sources, different in their nature and producing the odorous emissions generally different from the point of view of chemical composition. These emissions lead to the complaints of the population and can have undesirable consequences for the human health, even if the concentrations of chemical compounds do not exceed the level of toxicity.

The actual state of development of analytical chemistry, especially of gas chromatography, allows to sample the emissions *in situ* and to analyse very complex gaseous mixtures. Also the technique requires an adequate equipment, time and above all, the know-how. The chemical composition of odours from a great number of different sources was studied and is today well known (see for example *Thistlethwayte et al.* [1], *F.N.D.A.E.* [2]). Nevertheless, to assess the origin of odour pollution unanimously remains still a difficult task: the different industry activities are usually concentrated on one site, their emissions contain some compounds in common, the odour can be transported in the atmosphere on quite long distances from the source.

The aim of the present paper is to test ability of some discrimination and pattern recognition mathematical techniques to identify the origin of odour pollution on the basis of chemical analysis. We expected from successful method to provide a good discrimination between the possible sources on the basis of concentrations of a quite reduced number of chemical compounds, common to all of them. Fulfilment of this requirement would simplify the chemical analysis and would make possible the detection of the pollution origin even in the case of emissions having similar composition.

We have tested different mathematical techniques on the emissions from steel industry : namely coking plants, rolling mills and coil-coating plants. This specific application was selected because sufficient data were available to carry out the processing procedures. Nevertheless, the final goal of our work is to implement such techniques in an artificial nose suited to more general environmental odours.

2. Experimental part

Sampling and chemical analysis

Gaseous emissions from about hundred places in coking plants, cold rolling mills and coil-coating plants were sampled in plastic sampling bags and analysed by a gas chromatograph coupled with a mass spectrometer. The mass spectrometer was assisted by computerised treatment system, allowing compounds identification through a data bank. On the basis of the chromatographic analysis the total concentrations (in g / Nm^3) of compounds belonging to one chemical family, and total content of all chemical compounds in the mixture (C_TOTAL) were calculated. Fifteen families were taken in consideration: H_2S, NH_3, organic compounds of sulphur (S_ORG), nitrogen (N_ORG), and chlorine (Cl_ORG), alkanes, alkenes, alcohols, aldehydes, ketones, organic acids, esters of organic acids, benzene and its alkyl derivatives (BTX), polyaromatics (HAP) and phenols.

Gases containing organic acids were bubbled through a solution of a base and the total acids concentration was evaluated by titration. The concentration of acids found by the gas chromatography and by the titration was compared. The comparison allowed us to be sure that the acids are not subject to chemical modifications in the sampling bags during the transport.

The chemical analysis of each gaseous mixture was completed by olfactometric measurements. The gas was successively diluted in a dynamic olfactometer by pure air till the limit of odour perception was achieved and the dilution f at the olfactory threshold was calculated:

$$f = (Da + Dgo) / Dgo \qquad (1)$$

Da flowrate of the pure air, and Dgo, flow rate of the odorous emission, are used to prepare the mixture diluted to the threshold of the perception. The details on the sampling, and chemical and olfactometric analysis can be found in *Vigneron and Hermia* [3].

Mathematical methods

The statistical multivariate methods (factor, cluster and discriminant function analysis, multiple linear regression) and neural network with backpropagation algorithm and log-sigmoid neurons were applied for the mathematical analysis of the chemical and olfactometric data. The method ANOVA was used to compare the dilutions at the olfactory threshold f and the total concentrations of chemical compounds in emissions.

A three layer architecture (Figure 1) was used for pattern recognition with the neural network. In the learning phase, the input vectors from the training set were presented to the network. Simultaneously "1" was shown on the corresponding output and "0" elsewhere. After the training, the input vectors were presented once more to the system for recognition. The "leaving one out method" (*Everit* [4]) was also tested. By this method, the network is trained with the learning set minus one observation. This observation is then used to verify the capacity of the system to classify an observation of unknown origin. Thewhole process is repeated, each time omitting another observation.

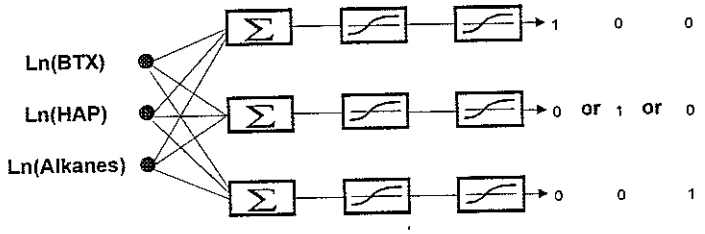

Fig. 1 - Scheme of the architecture of neural network.

3. Results and discussion

The emissions from the three plants under study are very complex; between 60 and 85 compounds were detected following the source. In coking plants, H_2S, NH_3, organic compounds of nitrogen and sulphur together with alkanes and aromatic hydrocarbons (BTX, HAP and phenols) were found. In coil-coating units, neither H_2S, NH_3, nor organic compounds of nitrogen and sulphur were observed, but the emissions were rich in oxygen derivatives (alcohols, aldehydes, ketones,...). Alkanes, BTX, and HPA were present as well. Rolling mills emissions exhibit the same chemical composition, and contain in addition phenols and in some cases Cl_ORG.

The statistical distribution of concentrations of all chemical families, as well as that of the dilutions at the threshold **f** was asymmetric with more observations in the range of small values (see by way of example figure 2). Normal probability plots of the logarithms have shown that all data have the log normal distribution. For this reason, logarithms of concentrations and of dilutions at olfactory threshold were used in all mathematical procedures. This fact is not reminded subsequently in the text.

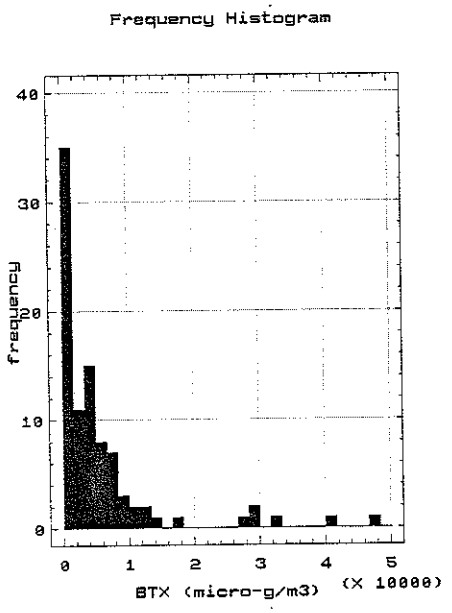

Fig. 2 - Sample frequency histogram for the BTX concentration.

The comparisons of the means of chemical compounds constant, C_TOTAL, and of the dilution at the threshold **f** have indicated that the emissions originating from the coil-coating units are significantly more rich on substances than the emissions from the others two sources, but they have the same odour level. This fact is probably due to the differences in odorous properties of the present compounds and confirms that not all of them contribute to the odour of the mixture.

The stepwise forward multiple linear regression has shown that the dilution at the olfactory threshold **f** can be calculated as a linear combination of the concentrations of some chemical compounds (Table 1). Although up to six chemical families were considered in the calculations, the equations using only three of them are already able to predict quite well the values of **f** (correlation coefficient of equation $r^2 > 0.9$). These families are unfortunately different following the source. Phenols and organic

compounds of sulphur and nitrogen enter the game in the case of coking plants, whereas BTX, alcohols and ketones (BTX, alcohols and acids) are important in coil-coating (rolling mills) units.

COKING PLANTS		COIL-COATING		ROLLING MILLS	
VARIABLES	r^2	VARIABLES	r^2	VARIABLES	r^2
PHENOLS	0.724	BTX	0.863	BTX	0.499
N_ORG	0.795	ESTERS	0.892	ACIDS	0.782
S_ORG	0.932	ALCOHOLS	0.91	ALCOHOLS	0.914
BTX	0.96	KETONES	0.924	ALKANES	0.976
H_2S	0.972			KETONES	0.9997
ALKANES	0.979				

Table 1 - Results of the forward stepwise linear regression. Threshold for Snedecor's F=1.

Figure 3 shows the results of the factor analysis, i.e., the representation of observation points in the plane of the first two factors. The fifteen variables are represented by their names, but for clearness reasons, the 99 observations are summarised by their centre of gravity for the three types of plant. We observe some variables clusters, corresponding to the three emissions sources under study. For example BTX - ketones - alcohols for the coil-coating units, N_ORG - NH_3 - HAP, or H_2S - S_ORG - phenols for the coking plants. Factor analysis, applied to the whole set of variables, classes rather well the observations and distinguishes the emissions sources.

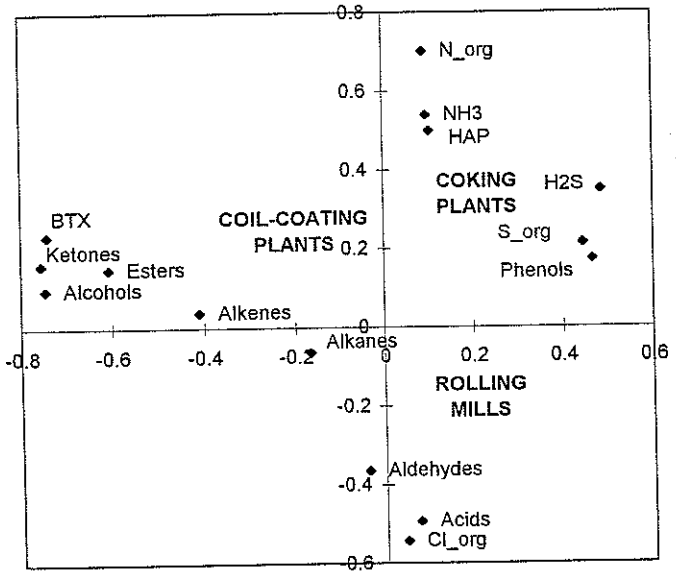

Fig. 3 - Plot of the first two factors weights for the fifteen chemical families observed in the coking plants, coil-coating units and rolling mills plants.

For the practical application, it is sufficient to choose one among the compounds that are typical for each of the three plants. For example BTX, ketones or esters for coil-coating plants, NH_3, HAP, H_2S or S_ORG for coking plants and acids, aldehydes or alkanes for rolling mills.

Three chemical families were detected in almost all samples of odorous emissions. We have studied: alkanes, benzene and its alkyl derivatives, and polyaromatic hydrocarbons more typical respectively of rolling-mills, coil-coating and coking plants. We have used the concentrations of this compounds in the cluster analysis and in the discriminant function analysis. Figure 4 shows the results of the first one, projected in the HAP - BTX plane. The clusters, found by the procedure, are identified by the lines surrounding each of them. Obviously, cluster analysis is unable to distinguish the three emissions sources. The points belonging to the rolling mills and to the coking plants are mixed together and the points of the coil-coating units form two clusters.

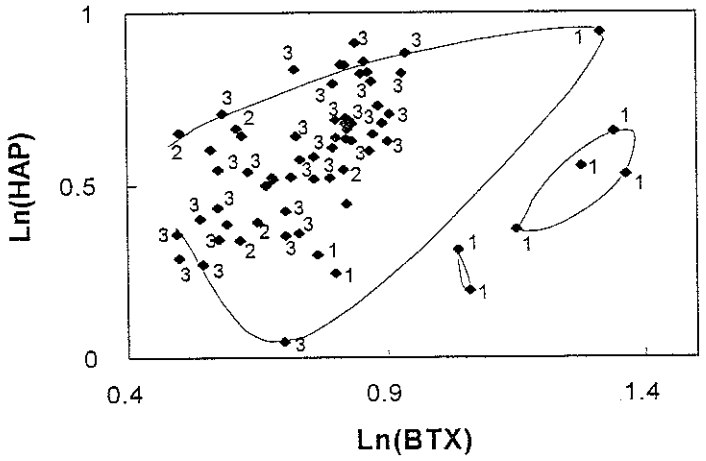

Fig. 4 - Cluster analysis plot in the plane of the variables HAP and BTX. 1 - coil-coating units, 2 - rolling mills, 3 - coking plants.

The stepwise procedure was used in the discriminant function analysis. The software has considered all the three variables, i.e., BTX, HAP and alkanes concentrations and has calculated two discriminant functions and three classification functions. These functions are linear combinations of the variables.

The discriminant functions are calculated always in order to allow the best discrimination between the different groups, the classification functions are used to allocate the experimental observations inside the groups. Figure 5 shows the graphic representation of the groups discrimination. For each observation, the value of the first discriminant function is plotted against the value of the second. We can see that the points for the three plants form three well separated clouds.

Also the classifications functions allow a good identification of the emissions source. Only one observation from coil-coating, and one from rolling mills emissions was wrongly classified. Seven observations (among 52) from coking plants were classified in a wrong group. In summary, 87 % of observations were correctly identified. It is quite encouraging, but one should look at those classifications as a diagnostic tool for identifying areas of strength and weaknesses in the current classification functions. These classifications are not *a priori predictions* but rather *post hoc classifications* and consequently they overestimate the quality of the model.

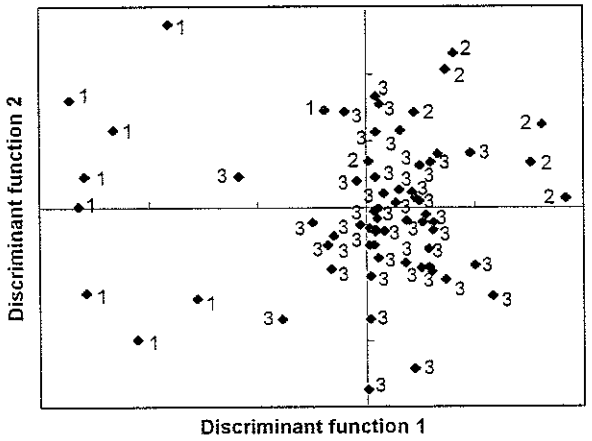

Fig. 5 - Discriminant function analysis plot in the plane of the discriminant functions. 1 - coil-coating, 2 - rolling mills, 3 - coking plants.

Neural network approach provides still better results. With the training set, 67 from 68 observations, thus 98.5 %, were correctly identified. Like in the case of discriminant function analysis, this estimate of the method strength is a little too optimistic. The "leaving one out method" gives us a more realistic one. In our case, the method was able to classify correctly 95.5 % of all observations. The learning phase of the neural network is slower in comparison with that of the discriminant function analysis, but further application is very quick and easy.

The mathematical procedures we have used in this work, can be classified into two groups. The factor and cluster analysis are unsupervised techniques. They are free to "build up" a model on the basis of the experimental data. This model can be used to allocate new observations, nevertheless, the aim of the procedures is chiefly to bring out pattern of similarities among observations. The neural network and discriminant function analysis belong to the supervised procedures. They request a teaching group of observations gathered from some items, whose membership in a specific group is already known. New items are then classified by determining how typical is their individual pattern of variables for one of the groups. Our results show that these techniques suit better to the identification of the odour emissions origin.

4. Conclusions

A suitable combination of the concentrations of compounds belonging to three odorous chemical families and a pattern recognition technique based either on the discriminant function analysis or on neural network has proved to be an efficient tool for the recognition of the source of odorous gaseous effluents. For the data set considered in this paper, neural network with backpropagation algorithm exhibits better misclassification rate. Thus, such procedures can be applied as data processing tools for artificial nose based on non-specific gas sensors. The factor analysis can be used as well, but the method needs a larger set of variables.

References

[1] Thistlethwayte D.K.B., Hardwick B., Goleb E.E., *Odor nuisances from waste disposal works and techniques for deodorization. Chimie Industrie - Génie chimique* ,106 (11), (1973) 795.

[2] F.N.D.A.E., *Lutte contre les odeurs des stations d'épuration.* Documentation technique n 13 - Ministère de l'Agriculture et de la Pêche (France), 1993

[3] S. Vigneron, J. Hermia, *Les nuisances olfactives dans des ateliers de la sidérurgie à froid: étude des méthodes de prévention et des techniques d'abbatement.* Rapport final. Société Belge de Filtration, Louvain-la-Neuve, 1991

[4] B.S. EVERIT, *Statistical Methods for Medical Investigations* - Second edition. Edward Arnold edition, London (UK), 1994