

## References

- [1] Grosjean, D., and Fung, K., *J. Air Pollut. Control Assoc.* **34**, 537 (1984).
- [2] Sonnefeld, W. J., Zoller, W. H., May, W. E., *Anal. Chem.* **55**, 275 (1983).
- [3] Yamasaki, H., Kuwata, K., and Kuge, Y., *Nippon Kagaku Kaishi* **8**, 1324 (1984).
- [4] Arey, J., Zielinska, B., Atkinson, R., and Winer, A. M., *Atmos. Environ.* **21**, 1437 (1987).
- [5] Arey, J., Zielinska, B., Atkinson, R., Winer, A. M., Ramdahl, T., and Pitts, J. N., Jr., *Atmos. Environ.* **20**, 2339 (1986).
- [6] Atkinson, R., Arey, J., Zielinska, B., Winer, A. M., and Pitts, J. N., Jr., The Formation of Nitropolycyclic Hydrocarbons and their Contribution to the Mutagenicity of Ambient Air, In: *Short-Term Bioassays in the Analysis of Complex Environmental Mixtures V*, Sandhu, S. S., DeMarini, D. M., Mass, M. J., Moore, M. M., and Mumford, J. S., Eds., Plenum Press, in press (1987).
- [7] Zielinska, B., Arey, J., Atkinson, R., and McElroy, P. A., Nitration of Acephenanthrylene Under Simulated Atmospheric Conditions in Solution, and the Presence of Nitroacephenanthrylene(s) in Ambient Particles, *Environ. Sci. Technol.*, submitted for publication (1987).
- [8] Zielinska, B., Arey, J., Atkinson, R., and Winer, A. M., The Nitroarenes of Molecular Weight 247 in Ambient Particulate Samples, *J. Chromatogr.*, to be submitted (1987).

## *Pattern Recognition Classification and Identification of Trace Organic Pollutants in Ambient Air from Mass Spectra*

**Donald R. Scott**

U.S. Environmental Protection Agency  
Research Triangle Park, NC 27711

and

**William J. Dunn III and S. L. Emery**

College of Pharmacy  
University of Illinois at Chicago  
Chicago, IL 60612

## Introduction

The most frequently used method of analysis of trace organic pollutants in ambient air is gas chromatography-mass spectrometry (GC-MS). Presently, target compounds in air monitoring samples are identified from gas chromatographic retention

times and a combined forward-reverse search of mass spectral reference spectra. The primary objective of this study is to develop computational pattern recognition procedures for the identification of chemical classes and, if possible, individual compounds from routine GC-MS data files. The additional information on nontarget compounds would supplement compound identifications obtained from the target list. The target set of compounds investigated consisted of 78 substituted benzenes, haloalkanes and haloalkenes.

## Methods

### SIMCA Pattern Recognition

SIMCA pattern recognition was developed by Wold and coworkers [1] for application to classification problems in chemistry. The technique is based on the modeling of chemical classes by disjoint principal component models. Once the class models have been determined, objects are classified by fitting their data to the class models. A standard deviation for each model is calculated from the residuals which represents a class tolerance level around the model in measurement space.

## Data Analysis

In this study an IBM PC-XT microcomputer with 640K memory was used. A modified version of the SIMCA program was used for the majority of the pattern recognition calculations. The low resolution mass spectra of 78 compounds were obtained from the EPA-NIH Mass Spectral Library on an INCOS data system (see table 1). The GC-MS data files were from routine calibration and analysis performed by the Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC.

The data were preprocessed by taking the autocorrelation transform of the mass spectra for mass lags less than 100. In the initial stages of class modeling, the autocorrelation transformed spectra for the 78 training compounds were examined for class separation with two and three dimensional principal components projections of the training data. Three different groups were found: nonhalogenated benzenes; chloroaromatics, chloroalkanes, chloroalkenes; and bromocarbons.

Principal components models were then derived for each class [2]. None of the class models

## Accuracy in Trace Analysis

Table 1. Compounds included in this pattern recognition study

Compound	Class <sup>a</sup>	Compound	Class <sup>a</sup>
		Dibromomethane	3
		Bromoform	3
<i>p</i> -Xylene	1	1,2-Dichloroethane	2
1,3,5-Trimethylbenzene	1	1,1,1-Trichloroethane	2
Isopropylbenzene	1	1,1,2-Trichloroethane	2
<i>n</i> -butylbenzene	1	1,1,2,2-Tetrachloroethane	2
1-Methyl-4-isopropylbenzene	1	Pentachloroethane	2
<i>o</i> -Dichlorobenzene	2	1,1-Dichloroethane	2
<i>p</i> -Dichlorobenzene	2	Trichloroethene	2
1-Chloro-2-methylbenzene	2	Tetrachloroethene	2
1-Chloro-4-methylbenzene	2	Bromoethane	3
<i>p</i> -Chlorostyrene	2	1,2-Dibromoethane	3
1,1-Dichloroethane	2	1-Chloropropane	2
1,1,1,2-Tetrachloroethane	2	2-Chloropropane	2
1,2,3-Trichloropropane	2	1,2-Dichloropropane	2
3-Chloro-1-propene	2	1,3-Dichloropropane	2
2-Chlorobutane	2	1-Bromo-3-chloropropane	3
1,3-Dichlorobutane	2	1,2-Dibromopropane	3
1,4-Dichlorobutane	2	2,3-Dichlorobutane	2
1,4-Dichloro-2-butene ( <i>cis</i> )	2	Tetrahydrofuran	0
3,4-Dichlorobut-1-ene	2	Benzaldehyde	1
1,4-Dioxane	0	1-Bromo-1-chloroethane	3
1-Chloro-2,3-epoxypropane	2	2,2-Dibromopropane	3
2-Chloroethoxyethene	2	2-Bromopropene	3
Acetophenone	1	2-Bromopropane	3
Benzonitrile	1	3-Bromopropane	3
Benzene	1	1-Bromopropane	3
Toluene	1	1-Chlorobutane	2
<i>o</i> -Xylene	1	1-Bromo-2-chloroethane	3
<i>m</i> -Xylene	1	Bromodichloromethane	3
Ethylbenzene	1	1-Bromobutane	3
Styrene	1	2,2-Dichlorobutane	2
Chlorobenzene	2	Dibromochloromethane	3
Bromobenzene	3	1,1,2-Trichloropropane	2
<i>m</i> -Dichlorobenzene	2	1,3-Dibromopropane	3
1-Chloro-3-methylbenzene	2	1,1,1,2-Tetrachloropropane	2
Chloroform	2	1,2,2,3-Tetrachloropropane	2
Carbon tetrachloride	2	1,3-Dibromobutane	3
Bromochloromethane	3	1,1,2,3-Tetrachloropropane	2
Bromotrichloromethane	3	1,4-Dibromobutane	3

<sup>a</sup> Class 1=nonhaloaromatics; class 2=chlorocarbons including aromatics and aliphatics; class 3=bromo- and bromochlorocarbons; class 0=internal standards and dioxane and tetrahydrofuran.

required more than three principal components. The largest number of autocorrelation coefficients relevant to a specific class was 40 in class 3. Eighty-six percent of the training set compounds were assigned to their correct class as a first choice and were within two standard deviations of the models.

### Hierarchical Classification Scheme [2]

Once a class assignment is made, further identification of each spectrum can be made. This is done

by calculating the distance, using the entire spectrum in autocorrelation space, of each compound to its three nearest neighbors in its assigned class. To obtain a specific structural assignment, the normalized correlation coefficient of the unknown mass spectrum with that of each of its three nearest neighbors is obtained. If the unknown is not identified as one of the three nearest neighbors in the first class assignment, the correlation coefficients of the three nearest neighbors in the second SIMCA class assignment are compared.

## Results

### GC-MS Calibration Samples

Three direct injection calibration samples, which contained all but one of the training set compounds, were analyzed [2]. Of the 77 compounds observed in the calibration runs, 84% were correctly identified. Of the 13 compounds not identified, three were correctly assigned by chemical class.

### GC-MS Field Samples

The GC-MS data files for three field samples were also analyzed [3]. The identities of the target compounds were determined by using both GC retention times and a combination of forward and reverse spectral matching techniques with stringent matching parameters. The identification of other compounds not on the target list was based on a Finnigan search technique. The application of the pattern recognition scheme to the transformed data for the target compounds resulted in 88% correct classification. The compound identification results were 85% accurate.

There were 75 different nontarget compounds identified in 120 occurrences in the three samples. The classification results agreed very well for the two class 1 and class 2 spectra. However, a very large number of alkanes and alkenes were incorrectly classified as chlorocompounds. Further details of this study are given in references [2], [3], and [4].

Although the research described in this article has been funded by the U.S. Environmental Protection Agency under Cooperative Agreement CR-811617 with the University of Illinois at Chicago, it has not been subjected to Agency review. The mention of commercial products does not constitute endorsement or recommendation for use.

## References

- [1] Wold, S., Albano, C., Dunn, W. J. III, Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., and Sjoström, M., in B. R. Kowalski (Editor), *Chemometrics: Mathematics and Statistics in Chemistry*, Reidel, Boston (1984), pp. 17-96.
- [2] Dunn, W. J. III, Koehler, M. G., Emery, S. L., and Scott, D. R., *Application of Pattern Recognition to Mass Spectral Data of Toxic Organic Compounds in Ambient Air*, *Chemometrics Intell. Lab. Sys.*, in press, Sept. (1987).
- [3] Scott, D. R., Dunn, W. J. III, and Emery, S. L., *Classification and Identification of Hazardous Organic Compounds in Ambient Air by Pattern Recognition of Mass Spectral Data*, *Environ. Sci. Tech.*, in press, Sept. (1987).
- [4] Scott, D. R., *Anal. Chem.* 58, 881 (1986).

## *Annular Denuders and Filter Packs Designed to Measure Ambient Levels of Acidic and Basic Air Pollutants*

Robert K. Stevens

U.S. Environmental Protection Agency  
Research Triangle Park, NC 27711

Measurements of acidic (e.g., SO<sub>2</sub>, HNO<sub>3</sub>) and basic gases (e.g., NH<sub>3</sub>) that coexist with fine particles (<2.5 μm) are used in models to assist in determining the origin and age of aerosols. Bias associated with each measurement method used to obtain this air quality data can degrade the real correlation between species. In addition, the sensitivity of most instrumental methods to measure SO<sub>2</sub>, NH<sub>3</sub>, HNO<sub>3</sub> or HNO<sub>2</sub> is limited to measurements at concentration levels above 2-5 ppb. Measurement of HNO<sub>3</sub> has also proven to be more difficult due to losses of HNO<sub>3</sub> in the sampling inlets used in some of the measurement procedures. Difficulties in differentiating atmospheric HNO<sub>3</sub> from HNO<sub>3</sub> produced by the dissociation of NH<sub>4</sub>NO<sub>3</sub> during sampling, further complicates monitoring procedures.

Possanzini et al. [1] described the development of an annular denuder system which removes reactive gases (e.g., HNO<sub>3</sub>, SO<sub>2</sub>) from air samples an order of magnitude more efficiently per unit length and at lower Reynolds number than open tubular denuder designs. Recently we developed an improved version of the annular denuder that incorporates several important features to minimize losses of key species during sampling and reduce the possibility of leaks in the components that join the various parts of the system.

To demonstrate the applicability of this improved design of the annular denuder, a series of field studies were conducted in Research Triangle Park, NC during the fall of 1986 and winter of 1987. An annular denuder system (fig. 1) is composed of four components: an inlet (to remove large particles), coated denuders (to collect the acidic and basic gases), filter pack (to collect fine particles and HNO<sub>3</sub> that may evaporate from the filter), and a flow controller-pump assembly. In this study a new glass Teflon-coated impactor inlet was designed with a short tube extending below the impactor to prevent large particles and rain droplets from entering the impactor. With the exception of the impactor surface, the entire inlet is