Automated Evaluation of GC-MS Analyses
of Chlorinated Phenols

P. JURÁŠEK, R. BREŽNÝ, and M. KOŠÍK

Department of Wood, Pulp, and Paper,
Faculty of Chemical Technology, Slovak Technical University,
CS-812 37 Bratislava

On a model mixture of chlorinated phenols contained in waste waters from the pulp industry some techniques of the automatic interpretation of data of the GC-MS analysis were verified. Efficiency of the programming system based on the pattern recognition technique PRIMA and of the techniques of the library searching of mass spectra was compared. The computer program (chemometric detector), which locates the mass spectra of chlorinated compounds in a few minutes, completes the information about the unknown mixture.

The combination of gas chromatography with a mass selective detector (GC-MS) has become one of the most used and most productive approaches of the analysis of complex mixtures of organic compounds [1—4]. The analysis resulted in the chromatogram and a file of mass spectra of individual components of the mixture. The effective possibility to evaluate the measured data is the chemometric approach with utilization of a computer technique.

The automated spectral interpretation of the low-resolution mass spectrometry can generally be classified in retrieval systems and spectral interpretation systems [5], or, from a view of access to information stored in a database, in direct and indirect database methods [6]. The direct database method comprises library search systems, when the unknown spectrum is compared with each member of the reference spectra of the library with the aim to select the most similar spectra as possible candidates for the unknown spectrum. The indirect techniques use a database only in the development of the method but not in its implementation — they attempt to transmit a large quantity of detailed information into a general form that can be applied independently of the database. The indirect methods can be divided into three categories: artificial intelligence, pattern recognition, and spectral simulation.

The pattern recognition methods involve in general the classification of objects into classes on the basis of their measured properties. In the case of the mass spectra the problem is solved, whether the unknown spectrum belongs to some of the considered structural groups characterized by spectra of the structurally interesting compounds being selected from the spectral library. Hence pattern recognition methods do not try to determine completely the unknown structure, but they only try to recover substructures contained in the unknown compound. The majority of the pattern recognition methods are applicable in the form of computer packages. The most known system is ARTHUR [7]; for personal computers the system PARVUS [8] is designed. The pattern recognition method PRIMA [9] belongs to the methods based on the concept of class distances. In contradistinction to the SIMCA method (the class distance is defined in the principal component space) the class distance is defined in the pattern space. For each class, a separate class distance is calculated on the basis of the centre of gravity and inhomogeneity for this class and then it is used for classification. The application of PRIMA is proposed when the training set or the available computer is small or the data set is incomplete.

The GC-MS data collection is nonselective: all the molecules eluting from the chromatograph are ionized and detected. Because of the tremendous amount of data generated in a continuously scanning GC-MS experiment, a chemometric detector can be used for the extraction of interesting information [10,11]. The chemometric detector is a computer method for the selective detection of a class of compounds in the data matrix.

The aim of this work is to demonstrate possibilities of the library searching, pattern recognition and chemometric detector techniques in evaluating of the GC-MS data of the mixture of chlorinated phenols, which occur in liquors of the pulp industry.

METHODS

Preparing the model mixture of chlorinated phenols sulfuryl chloride (1 cm^3) was added drop-

wise to the solution of phenol, 2-methoxyphenol and 2,6-dimethoxyphenol (100 mg each) in chloroform (10 cm³) during cooling by water (15 °C).The reaction mixture was left standing with exclusion of the air humidity access for 2 h. This mixture was then dry-evaporated in the vacuum evaporator and finally twice — together with a small amount of toluene to remove the hydrogen chloride residues.The evaporation residue was dissolved in 5 cm³ of chloroform and injected (1 mm³) onto the column head of the GC-MS 5890/5970B instrument (Hewlett—Packard).

The instrumental condition of GC-MS analysis: Column HP-1, length 12 m, inner diameter 0.2 mm, film thickness 0.33 µm; carrier gas helium, flow rate 23 cm³ s⁻¹; column temperature 60 °C (3 min hold), then programmed to 250 °C at 6 °C min⁻¹; injector temperature 280 °C; ion-source temperature 280 °C; electron energy 70 eV; mass scan range m/z = 33—450.

For data interpretation the following three chemometric approaches of the automatic evaluation of mass spectra were used.

1. Library Searching

The operating software of the Hewlett—Packard GC-MS system contains two types of mass spectra libraries from the NBS collection: namely NBS.L and PBMNBS.L with corresponding search algorithms [12—14]. The first one uses the correlation index and reduced spectra to 10 peaks; the second one is the PBM system which uses the reduced spectra to 15 to 26 peaks.

Using the firm software the special libraries named AROM.L (it contains spectra reduced to 10 peaks) and PBMAROM.L (spectra reduced to 15—26 peaks) were created. Each library contains 172 of mass spectra of chlorinated phenols and benzenes; 39 of them were obtained from synthesized compounds, the other ones were taken over from the NBS.L and PBMNBS.L libraries.

For the analysis of GC-MS data of the model

mixture of chlorinated phenols four ways of library searching were used:

a) searching using simple correlation index with 10 peaks in the NBS.L library (38 917 spectra),

b) searching using simple correlation index with 10 peaks in the special library AROM.L (172 selected spectra),

c) searching using the PBM algorithm in the PBMNBS.L library (43 005 spectra),

d) searching using the PBM algorithm in the PBMAROM.L library (172 selected spectra).

2. Pattern Recognition (PRIMA)

Eight structural classes of chlorinated phenols occurring in bleaching waters of the pulp industry were defined (Table 1). For each class a special sublibrary of the mass spectra selected from the libraries AROM.L and PBMAROM.L was created. For each structural class four or five characteristic fragments in the mass spectra (Table 1) were determined. Thus the characteristic fragments of all classes created a collection of 36 properties. The property values were the peak heights of these fragments. In this way from spectra in the sublibraries a data matrix (training data set) was created. From the data of the GC-MS analysis of chlorinated phenols the data matrix (a recognition data set), the rows of which corresponded to scans and columns to 36 characteristic properties, was then created in the same way.

Both, the training set and the recognition set were then represented by input data for the own pattern recognition process (PRIMA). In the training, the classes were characterized by the centre of gravity (the vector formed from the property averages) and by the inhomogeneity (dispersions of the properties). These two parameters were calculated from the submatrix of the training data matrix corresponding to the given class. The recognition consisted of two steps. First, the distances of the point corresponding to the unknown spectrum from the centre of gravities of the classes were calculated

Table 1. Structural Classes and Corresponding Characteristic Fragments of Chlorinated Phenols Analyzed by Pattern Recognition Methods

Structural class	Substituent	Characteristic fragments
I	x-chloro, 2-methoxy	51, 115, 143, 158, 160
II	x,y-dichloro, 2-methoxy	177, 179, 192, 194
III	x,y,z-trichloro, 2-methoxy	183, 211, 215, 226, 225
IV	x,y,z,u-tetrachloro, 2-methoxy	245, 247, 249, 262
V	2-methoxy	28, 81, 95, 109, 124
VI	x-chloro, 2,6-dimethoxy	127, 130, 173, 188, 190
VII	x,y-dichloro, 2,6-dimethoxy	161, 207, 222, 224
VIII	x,y,z-trichloro, 2,6-dimethoxy	241, 256, 258, 260

Table 2. Part of the Output Results of the Distances between Unknown Spectra and Structural Classes

Spectrum	Classification to class		Distances to the centre of gravities of classes						
			II	III	IV	V	VI	VII	VIII
540	VII	30.37	27.97	34.72	37.76	32.71	36.31	8.79	38.37
541	VII	29.58	27.20	33.97	36.82	31.60	35.31	7.95	37.44
664	VIII	31.13	33.66	36.25	35.40	33.38	36.57	36.58	3.34
665	VIII	30.76	33.58	36.41	35.28	33.01	36.24	36.34	2.67
666	VIII	30.74	33.49	36.67	35.15	32.84	36.20	36.29	2.62

(Table 2). In the second step, the class to which the unknown objects belong must be decided.

Empirically, it was determined that the classification was assigned if: the distance to the class was smaller than the class distance threshold — it was equal to one half of the average of the distances between the centres of gravities of the classes determined in the training set; the distance of the unknown spectrum to all other classes was 2.5-times greater (or more) than to the assigned class.

The program system enabling the indicated procedure to create pattern recognition involves three independent programs:

PRLIB — it enables to create the sublibraries by importing the mass spectra from commercial libraries, from own measurements, from literature, *etc.*, to determine the characteristic fragments, it creates a training set;

MSDATA — it creates a recognition set from the measured data;

EXPRIMA — it executes the pattern recognition method PRIMA [9], classification.

3. Chemometric Detector

Another method used for the GC-MS data evaluation utilized chemometric detector. The computer program is able to search a GC-MS data set for isotope patterns (cluster of peaks in the mass spectrum) indicative of any number (1—10) of chlorine atoms.

All programs were created in the HP-Pascal 3.0 language and are applicable in the HP 59970C Chemstation software.

RESULTS AND DISCUSSION

The model mixture of the chlorinated phenols occurring in waste waters from the pulp bleaching was prepared by chlorination of phenol, 2-methoxyphenol, and 2,6-dimethoxyphenol using sulfuryl chloride. The GC-MS analysis showed the presence of about 20 compounds (Fig. 1).

The aim of our work was to demonstrate the effectiveness of methods for interpreting the mass spectra of the mixture components, analyzed by GC-MS, by using computer methods.

The library searching as the most used method was utilized with four types of the mass spectra libraries. In the areas of *1a* and *1c* commercial libraries NBS.L and PBMNBS.L were employed and in the cases *1b* and *1d* the specialized sublibraries AROM.L and PBMAROM.L (simultaneously containing various compounds with chlorinated and phenol functional groups) were used.
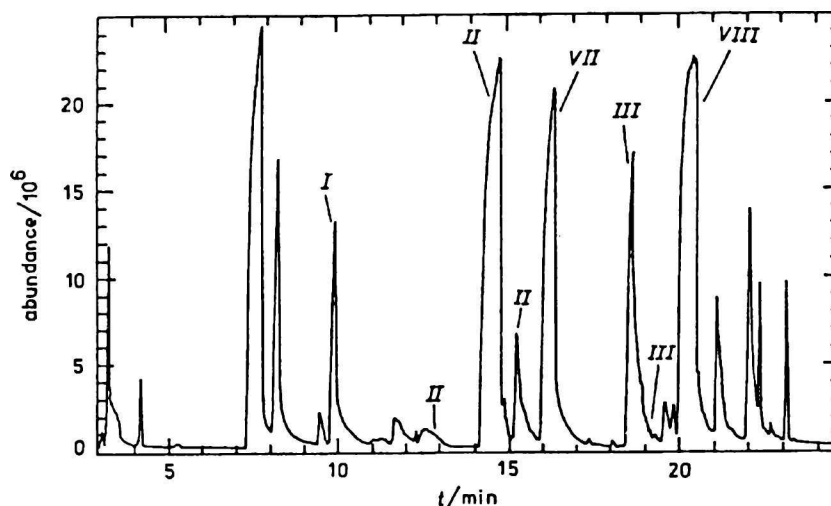
Fig. 1. Part of the chromatogram of the model mixture of chlorinated phenols and results of classification by using pattern recognition method.

Table 3. Results of the Analysis of Model Mixture Obtained by Chemometric Methods

Chromatographic peak	Substituent	Library searching				Pattern recognition	Class	Cluster detector (Cl atoms)
		1a	1b	1c	1d			
1	x-chloro	+	+	+	+	N		1
2	Nonidentified compound	N	N	N	N	N		N
3	2,4-dichloro	+	+	+	+	N		2
4	x,y-dichloro	+	+	+	+	N		2
5	2,4-dichloro, 6-methyl	+	+	+	+	N		2
6	x-chloro, 2-methoxy	−	+	N	+	+	I	1
7	2,4,5-trichloro	+	+	+	+	N		3
8	4,5-dichloro, 2-methoxy	+	+	+	+	+	II	N
9	x,y-dichloro, 2-methoxy	+	+	+	+	+	II	2
10	x,y-dichloro, 2-methoxy	+	+	+	+	+	II	2
11	x,y-dichloro, 2,6-dimethoxy	N	+	N	+	+	VII	2
12	x,y,z-trichloro, 2-dimethoxy	N	+	N	N	+	III	3
13	x,y,z-trichloro, 2-dimethoxy	N	+	N	N	+	III	N
14	Nonidentified compound	N	N	N	N	N		3
15	Nonidentified compound	N	N	N	N	N		N
16	x,y,z-trichloro, 2,6-dimethoxy	N	+	N	+	+	VIII	3
17	Nonidentified compound	N	N	N	N	N		3
18	Nonidentified compound	N	N	N	N	N		3
19	Nonidentified compound	N	N	N	N	N		N
20	Nonidentified compound	N	N	N	N	N		N
Time of the analysis/min		20	0.5	12	3	13		

+ Correct determination, − false determination (the apparent similarity of spectra), N nonidentified or explicitly incorrect results.

The results of the library searching are shown in Table 3. The analyses *1a* and *1c* were successful only for compounds encompassed in the libraries and for good chromatographically separated peaks. For the other peaks the results were nondefined or explicitly incorrect, which was obvious from the visual comparison of the unknown spectrum with the offered library spectrum. The visual comparison is in the case *1c* more simple and reliable because the PBM-reduced library contains spectra having 15 to 26 peaks. Better results were achieved by using the special libraries. In the cases *1b* and *1d*, the identification was more successful than in the cases *1a* and *1c*, because special libraries contain also spectra of chlorinated phenols being absent in the commercial libraries NBS.L and PBMNBS.L. Some peaks were not identified, because of lack of similar structure reference spectra and also because of bad chromatographic separation (mixture spectra).

The other approach to the automatic spectra interpretation is pattern recognition. The program system of application of the pattern recognition PRIMA method was elaborated. In contrast to the library searching methods the technique was used for filtering the GC-MS data of the model mixture in order to classify the mixture components into eight structural classes (Table 1). In one case (Table 3, peak *13*) the peak was correctly assigned to the class *III*, although it was the mixture spectrum. The compound was identified by the library search only with a small probability.

The information obtained was completed even by results of a chemometric detector. Table 3 shows complete results of the detection. For all identified compounds the number of chlorine atoms was confirmed, except peaks *8* and *13*, because the isotope peak patterns in the mass spectrum were deformed (the low intensity of the signal). For peaks *14*, *17,* and *18* a number of chlorine atoms was proposed. This information could help in the subsequent human interpretation.

CONCLUSION

In general we can state that the library search method provided correct results in the case of good chromatographic separation and detection, and when the analyzed spectra occur in the spectra library.

We have shown that the pattern recognition method enables one to create the program means which can substitute the extensive program systems for the library searching. The advantages of this approach are following:

1. Every scan of the chromatogram is evaluated so that each small peak can be detected (the library searching is automatically applicable only to the spectrum from the top of the peak or

from the average spectrum, because of the tremendous data processing);

2. The mixture component may be correctly identified also in cases of the worse chromatographic separation, in as much as the pattern for the certain class may be recognized also in the mixture spectrum;

3. The program system is universal and it can be modified for various structural classes by changing the characteristic properties describing the common pattern of a training set.

The success of this approach depends on the selection of the properties derived from the spectra that form the pattern.

The chemometric detector has confirmed the results of previous techniques or it gives one an additional information useful in the human interpretation of mass spectra. The technique can be used in a variety of chemical problems to quickly locate the mass spectra of chlorine-, bromine- or sulfur-containing compounds (execution time 0.2 s per scan). In dependence on the signal threshold parameter every scan in the chromatogram may be treated and therefore also the small peaks can be detected. The program does not need large data storage capacities for the spectral library (it requires only the presence of the char-acteristic isotope peak cluster in the unknown mass spectrum).

REFERENCES

1. Lindström, K. and Nordin, J., *J. Chromatogr. 128*, 13 (1976).
2. Bjorseth, A., Lunde, G., and Gjos, N., *Acta Chem. Scand., B 31*, 797 (1977).
3. Knuutinen, J., Tarhanen, J., and Lahtiperä, M., *Chromatographia 15*, 9 (1982).
4. Knuutinen, J., *J. Chromatogr. 248*, 289 (1982).
5. Lebedev, K. S., Tormyshev, V. M., Derendyaev, B. G., and Köptyug, V. A., *Anal. Chim. Acta 133*, 517 (1981).
6. Small, G. W., *Anal. Chem. 59*, 535A (1987).
7. Koskinen, J. R. and Kowalski, B. R., *J. Chem. Inf. Comput. Sci. 15*, 119 (1975).
8. Forina, M. and Leardi, R., *Trends Anal. Chem. 7*, 53 (1988).
9. Juricskay, I. and Veress, G. E., *Anal. Chim. Acta 171*, 61 (1985).
10. Varmuza, K., *Trends Anal. Chem. 7*, 50 (1988).
11. LaBrosse, J. L. and Anderegg, R. J., *J. Chromatogr. 314*, 93 (1984).
12. Manuals for GC-MS system fa. Hewlett—Packard, Chemstation 59970 version 3.1.
13. Pesyna, G. M., Venkataraghavan, R., Dayringer, H. E., and McLafferty, F. W., *Anal. Chem. 48*, 1362 (1976).
14. Atwater, B. L., Stauffer, D. B., McLafferty, F. W., and Peterson, D. W., *Anal. Chem. 57*, 899 (1985).