United States
Environmental Protection
Agency

Atmospheric Sciences Research
Laboratory
Research Triangle Park NC 27711

Research and Development

EPA-600/S3-84-105 Dec. 1984

EPA

Project Summary

A Comparative Evaluation of GC/MS Data Analysis Processing

E. D. Pellizzari, T. Hartwell, and J. Crowder

Mass spectra obtained by fused silica capillary gas chromatography/mass spectrometry/data system analysis of mixtures of organic chemicals adsorbed on Tenax GC cartridges were subjected to manual and automated interpretative techniques. Synthetic mixtures (85 chemicals representing 15 chemical classes) were prepared to simulate the following design effects: (1) the occurrence of similar and dissimilar overlapping mass spectra from two or more constituents present in unresolved gas chromatographic peaks; (2) the occurrence of similar and dissimilar proportions (concentrations) of unresolved components in gas chromatographic peaks, and (3) the presence of different chemical classes in a mixture. Environmental samples from seven different geographical areas in the continental United States were collected and analyzed.

Using synthetic mixtures, the interpretative methods evaluated for accuracy were (1) manual (skilled interpretor), (2) INCOS data software, (3) Mass Spectra Fourier Transformed/Search software, and (4) a Research Triangle Institute Mass Spectral Search System. A Rindfleisch deconvolution program was also applied to raw data prior to using the automated procedures. Only manual, INCOS, and the Research Triangle Institute system were evaluated with environmental samples. The deconvolution program was also part of this evaluation.

This Project Summary was developed by EPA's Atmospheric Sciences Research Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

The problem of monitoring the environment for the presence of potentially hazardous agents and assessing their capabilities to cause health and ecological effects is enormous. In the United States alone, 30,000 to 50,000 chemical substances are produced each year. There are now well over 4 million chemical substances registered with the American Chemical Society Abstract Service, and approximately 1,000 new chemicals are developed by industry and placed in commercial usage annually. In terms of volume, the production of organic chemicals in the noncommunist world increased from 7 million tons in 1962 to 63 million tons in 1970 and is predicted to increase 250 million tons by 1985.

Analytical techniques have been developed for simultaneous analysis of several hundred vapor-phase organics in an ambient air sample. One example employs high resolution gas chromatography/mass spectrometry/data system (GC/MS/DS), an approach that gives qualitative and quantitative information. However, this technique produces 2,000-3,000 spectra per sample, which is a significant number for qualitative interpretation by an experienced investigator.

A presumably efficient method for interpreting GC/MS/DS data is the use of computerized spectral searching systems. Computerized systems automated to various degrees have not been extensively compared to manual interpretation

and, thus, their accuracy is not well documented. To determine the accuracy of software interpretative techniques, a comparative evaluation of GC/MS/DS data processing techniques was made.

The primary objective was to synthesize chemical mixtures of known composition that would adequately test the accuracy of mass spectral software identification systems developed by commercial sources and at universities. A specific aim of this research was to determine the effects, if any, of overlapping spectra, concentration, and compound classes on the accuracy of the identification by software algorithms for various components in a mixture. By testing algorithms with known authentic chemical compositions, the accuracy of identification routines was assessed. Thus, the initial evaluation employed synthetic mixtures and subsequent evaluation employed ambient air samples.

Three mass spectral identification search routines were tested with raw mass spectral data and with the same spectral data that had been processed through a deconvolution program. The deconvolution program, developed by Rindfleisch, was designed to resolve overlapping mass spectra that are present as coeluting compounds in a GC peak. The three mass spectral identification search routines that were tested were (1) the INCOS data software supplied on a Data General computer (in this case interfaced to a Finnigan 3300 GC/MS, (2) a Fourier transform analysis software (MSFS) routine developed by T. Isenhour coupled with a standard U.S. Environmental Protection Agency mass spectral library, and (3) the Mass Spectral Search Identification Software (RTI/MSSS) routine developed at Research Triangle Institute. Raw mass spectral data were also interpreted by a skilled investigator who had no knowledge of sample composition.

To further test the computer algorithms' ability to identify organic compounds, environmental ambient air samples were collected. Even though synthetic mixtures were used to evaluate the search algorithms, they do not necessarily represent the level of complexity that might be encountered with environmental samples because environmental samples may contain several hundred constituents in various concentration ratios. However, the synthetic mixtures allowed a true method of assessing the accuracy of identification because the composition of the mixtures was exactly known. With environmental samples, the composition is totally unknown and, thus, the comparison of identification was to manual interpretation. The two algorithms that were further examined were the RTI/MSSS and the INCOS system. (The MSFS was not included in this portion of the study since it performed so poorly with synthetic mixtures.)

Procedure

Compounds representing alkanes, alcohols, aldehydes, esters, ethers, ketones, nitriles, nitrogen-containing compounds, aromatics, halogenated alkanes, halogenated aromatics, aromatic aldehydes, sulfur-containing compounds, acids, and phenols were selected because they have been identified in ambient air throughout the continental United States and are potentially toxic. Also, these chemicals in many cases are only partially resolved as GC peaks and in a few cases have similar mass spectra.

Thus, this set of chemicals has the desired design characteristics of providing overlapping spectra in unresolved GC peaks and compound class effects on the quality of chromatographic peaks, both of which affect the quality of mass spectra used to evaluate automated interpretative methods.

The criteria used for selecting sites were (1) the chemical classes potentially emitted because of the particular industrial activities, (2) if the potential emission rates were of sufficient magnitude to provide significant measurable levels in the ambient air, (3) the assessibility of sites to locating sampling systems, (4) the presence of dense populations near the chemical industry, and (5) the potential for a unique background or interferences to the qualitative analysis by GC/MS/DS.

The individual chemicals were chromatographed on a 25-m bonded phase fused silica capillary column. Using retention time data, the chemicals were arranged in order of resolution and grouped according to similar retention times. From these data, mixtures of appropriate chemicals were loaded onto Tenax GC cartridges for analysis by GC/MS. The mixtures were designed to evaluate the effect of (1) similar and dissimilar overlapping mass spectra from partially resolved GC peaks, (2) similar and highly disproportionate concentrations of unresolved chemicals, and (3) different chemical classes present with different chromatographic peak quality.

Nineteen Tenax GC cartridges were loaded with chemicals in various combinations (L:50-200 ng; M:200-650 ng; H:650-1500 ng; HS:1500-5000 ng). Cartridges were analyzed using a Finnigan 3300 GC/MS equipped with an INCOS data acquisition system. The same fused silica capillary column and GC conditions employed for retention time determinations were used for GC/MS analysis. All GC/MS data acquired by the INCOS system were processed on hard disk and archived on magnetic tapes.

The GC/MS analysis was conducted within 24 h of preparing synthetic mixtures; the entire set of cartridges required two and one-half days to analyze.

The sampling and analysis methods employed in this study for environmental samples were developed by other researchers. Three samples were collected in triplicate. Two locations were used at each site; one of these locations was used to collect a high and a low volume sample (25 and 5 liter, respectively) over a period of approximately 25 min. Locations were always selected using meteorological considerations so that each location was generally downwind from the industrial facility.

The identification algorithms used on the raw data and resulting data from the Rindfleisch cleanup algorithm for identification of chemicals in the synthetic mixtures were (1) manual interpretation (composition was unknown to interpreter), (2) the RTI/MSSS, (3) the MSFS, and (4) the INCOS system.

The use of the INCOS software for identification of chemicals employed the chromatographic peak-top enhancement algorithm, a form of spectral deconvolution. The operator decided the location of the peak-top by inspecting the ion chromatograms and mass spectra in concert with the peak-top enhancement routine. Thus, the operator, not the computer, selected the mass spectrum to be submitted to the INCOS search algorithm. Except for the manual method that compared mass spectra to the Eight Peak Index and Wiley Library and did not employ knowledge of retention times, each of the algorithms was compared with and without the Rindfleisch deconvolution system.

Mass spectra obtained from the 12 environmental samples were submitted to (1) manual interpretation, (2) INCOS software, and (3) the RTI/MSSS. Statistical analysis was performed for mass spectral data of synthetic and environmental samples to test the comparability of the different identification routines of compound identification and to identify which factors (e.g., Rindfleisch deconvolution), if any, affect their comparability.

Results and Discussion

The GC resolution was inadequate for the synthetic mixtures with many constituents partially resolved from one another, a desirable condition for this study. The DB-1 fused silica capillary column yielded symmetrical peaks with minimal tailing for most chemicals except the strong acids and bases.

Except for manual interpretation by skilled researchers, the computer algorithms provided a ranking value with each analyte's identity in a sample. The five highest-ranking values were examined to determine whether the correct identity was present as one of these choices. The data (and rankings) were sorted to gain insight into the accuracy of identification of these chemicals by automated and manual methods.

Table 1 presents the overall percent correct identities by the method of identification (includes data from all mixtures). These results clearly indicate that the highest percent accuracy was obtained with manual interpretation and the INCOS algorithm. The poorest results were obtained with the Fourier transform mass spectral search algorithm, which achieved only 49% correct identifications. On an overall basis, the application of the Rindfleisch deconvolution algorithm to raw data prior to subjecting the mass spectra to the various identification routines appeared to have very little effect on increasing the percent of correctly identified chemicals.

Table 2 presents a summary of these design effects on the accuracy of identification. The results in Table 2 demonstrate that manual interpretation and the INCOS algorithm appeared to give the highest percent correct identification; the worst percent correct identification was obtained with MSFS. For a few chemical classes, the application of Rindfleisch deconvolution to raw data appeared to improve accuracy of identification. However, in most cases, there was no significant improvement. As anticipated, higher percent identification accuracy was obtained when the mass spectra were dissimilar for constituents in unresolved chromatographic peaks. An exception to this trend was the results obtained with the MSFS. Finally, for those chromatographic peaks representing triplets, quartets, quintets, and sextets, the percent accuracy was greater when the constituents were at approximately equivalent concentrations than when they appeared in combinations of high-low concentrations.

Table 1. Overall Percent Correct Identification by Method of Identification

Method	0	1/-1	2/-2	3/-3	4/-4	5/-5	Missing
Manual	29.3	70.7	--	--	--	--	--
RTI/MSSS							
-D[a]	20.9	59.0	6.0	2.4	1.1	0.5	0
+D[b]	26.9	60 9	7 5	2.8	1.5	0.5	0
MSFS							
-D	42.4	49.5	3.3	2.3	0.9	1.5	0
+D	41.9	48.4	5.0	2.8	0.7	1.3	0
INCOS							
-D	21.9	75.2	1.3	0.8	0.3	0.5	0
+D	25.4	58.2	2.0	0.9	0.2	1.1	13.1

[a]Without deconvolution.
[b]With deconvolution.

Table 2. Summary of Design Effect on Percent Accuracy of Identification

		RTI/MSSS		MSFS		INCOS	
Design Effect	Manual	-D[a]	+D[b]	-D	+D	-D	+D
Similar Spectra/Medium Level/ Doublet (N=18)[c]	61	55	44	61	39	67	67
Similar Spectra/Hi-Low Level/ Doublet (N=36)	61	58	55	55	50	61	55
Dissimilar Spectra/Medium Level Doublet (N=38)	78	66	74	47	60	89	84
Dissimilar Spectra/Hi-Low Level Doublet (N=76)	68	57	67	50	50	84	76
Medium Level/Triplet (N=50)	78	72	66	54	54	72	66
Hi-Low Level/Triplet (N=100)	74	62	64	48	46	80	84
Medium Level/Quartet (N=57)	70	60	65	46	49	75	72
Hi-Low Level/Quartet (N=114)	61	46	53	40	40	66	59
Medium Level/Quintets (N=62)	77	61	68	48	54	77	68
Hi-Low Level/Quintets (N=122)	73	59	53	47	44	78	60
Medium Level/Sextets (N=64)	72	66	67	56	48	76	-[d]
Hi-Low Level/Sextets (N=128)	72	58	58	54	48	72	65

[a]-D = without deconvolution
[b]+D = with deconvolution.
[c]N = number of observations.
[d]Missing data.

The percent correct and incorrect identities by chemical group and method of identification were examined. In alls, the identification routines gave the highest accuracy for halogenated alkanes. On the other hand, the worst results for all cases were observed with aldehydes. For alkanes, alcohols, aldehydes, esters, ethers, nitriles, nitrogen compounds, halogenated alkanes, halogenated aromatics, aromatic aldehydes, and acids, the highest percent correct identities were obtained with the INCOS algorithm. For ketones, aromatics, sulfur compounds, and phenols, manual identification was superior. In no cases did the MSFS give the best results for any chemical class studied.

The percent correctly identified chemicals by level and method of identification was compared. Each chemical is more often correctly identified (regardless of the method of identification) when the concentration of the chemical in a mixture is high. There were some exceptions, however, and these were *n*-dodecane, *n*-propyl acetate, di-*n*-butyl ether, 2-methylbenzofuran, pyridine, *m*-ethyltoluene, trimethylbenzene, naphthalene, 2-methylthiophene, and acetic acid. In these cases, when the levels were either high or very high, the identification routines

failed more frequently to correctly identify the chemicals. As previously indicated, these results also indicate that the MSFS had extreme difficulty in properly identifying any of the alkanes.

An overall comparison was also made of the percent correct and incorrect identifications between manual and computer algorithm methods. The combination of manual identification and INCOS algorithm on raw data (nondeconvoluted) yielded the highest agreement (63.6%), i.e., both methods correctly identified the same chemical. The lowest percentage agreement observed was for manual and deconvoluted data using the MSFS (43.0%).

A comparison was also made between computer identification algorithms to determine the percent that the methods gave either correct or incorrect results. The highest agreement (55.3%) was between INCOS (nondeconvoluted data) and RTI/MSSS (deconvoluted data). Similarly, the worst agreement was with MSFS. Finally, a comparison between raw data (nondeconvoluted) and deconvoluted data using the INCOS algorithm was also performed. Surprisingly, the agreement when both correctly identified the analytes was only 61.6%.

Because the correct answer on the identification of the chemicals in the environmental ambient air samples could not be known, the only comparison that could be made was relative to manual interpretation. The greatest percent agreement occurred between manual and deconvoluted data interpreted by the INCOS system. However, this agreement was only 53.9%. The percent agreement by chemical class was also determined; however, the number of observations in some cases was rather small. In general, the INCOS automated method agreed more often by chemical class with manual than RTI/MSSS.

Conclusions and Recommendations

An evaluation of the effect of similar versus dissimilar mass spectra occurring in an unresolved chromatographic peak revealed that a higher accuracy of identification occurred when the mass spectra were dissimilar. Deconvolution applied to raw data did not appear to improve the accuracy of identification for the automated procedures, whether or not the spectra were similar.

As the level of chemical increased (50 ng to 5,000 ng/cartridge), the accuracy of identification also increased until the mass spectra became saturated. The accuracy decreased after this point. The saturation effects (>1,000 ng/cartridge) were observed to decrease accuracy of identification for *n*-dodecane, *n*-propyl acetate, di-*n*-butyl ether, 2-methylbenzofuran, pyridine, *m*-ethyltoluene, 1,3,5-trimethylbenzene, naphthalene, 2-methylthiophene, and acetic acid.

A comparison was made of the ability of each interpretative procedure to accurately identify the chemicals in each mixture. The overall best percent correct identification was achieved on data without deconvolution (-D) using the INCOS software (75.2%). The remaining procedures ranked as follows: manual (70.7%) > RTI/MSSS (with deconvolution (+D), 60.9%) > INCOS (+D, 58.2%) > RTI/MSSS (-D, 59.0%) > MSFS (-D, 49.5%) > MSFS (+D, 48.4%). In general, the use of a deconvolution algorithm did not always appear to aid in correctly identifying chemicals in these synthetic mixtures.

The automated interpretative methods were compared to manual interpretation for the percentage of correctly identified chemicals in a mixture. The highest percent agreement (-D, 63.6%) was between INCOS and manual interpretation; the lowest percent agreement was between manual interpretation and MSFS (+D, 43.0%). When comparing only automated procedures, the best agreement (55.3%) was between INCOS (-D) and RTI/MSSS (+D).

Because environmental air samples represented truly unknown mixtures, the correct answers were unknown and, thus, all comparative statistics on the answers (identities) were made relative to manual interpretation. The following observations were apparent:

1. A higher percent agreement occurred between manual and RTI/MSSS or INCOS methods with the higher-volume (thus, higher chemical levels) samples than with low-volume samples.
2. The application of deconvolution to the raw mass spectra caused the percent agreement to decrease for many samples.
3. The highest overall agreement (53.9%) across all samples was between manual interpretation and INCOS interpretation (+D).
4. INCOS more often agreed with the answers obtained manually than with the RTI/MSSS, when comparing by chemical classes.
5. Aldehydes and halogenated aromatics were always ranked in the top six best agreements regardless of automated method used.
6. Sulfur compounds always ranked among the worst three in percent agreements regardless of the automated method.

For the chemical classes studied, this program established that accuracy can best be achieved by applying either the INCOS or manual interpretative procedures. The accuracy is still unacceptably low. Several recommendations were offered:

1. Evaluate additional algorithms, e.g., the STIRRS (Cornell University).
2. Include the use of relative retention data for chemicals analyzed under the same GC/MS conditions as environmental samples and evaluate the change in accuracy, if any, by their inclusion.
3. Expand the list of chemical classes to include polynuclear aromatics, pesticides, etc.
4. Attempt to optimize the Rindfleisch deconvolution algorithm.
5. After accomplishing items 1-4 above, consider the need for further research on development of automated algorithms or optimization of those appearing to be most promising.

E. D. Pellizzari, T. Hartwell, and J. Crowder are with the Research Triangle Institute, Research Triangle Park, NC 27709.
***Leonard Stockburger** is the EPA Project Officer (see below).*
The complete report, entitled "A Comparative Evaluation of GC/MS Data Analysis Processing," (Order No. PB 85-125 664; Cost: $16.00, subject to change) will be available only from:
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650
The EPA Project Officer can be contacted at:
Atmospheric Sciences Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

BULK RATE
POSTAGE & FEES PAID
EPA
PERMIT No. G-35

Official Business
Penalty for Private Use $300

☆ U.S. GOVERNMENT PRINTING OFFICE: 1985-559-111/10752