SANDIA REPORT
SAND2010-0981
Unlimited Release
Printed February 2010

Final Report: Multicomponent Forensic Signature Development: Interactions with Common Textiles; Mustard Precursors and Simulants

Mark H. Van Benthem, Theodore T. Borek III, Curtis D. Mowry, and Paul G. Kotula

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2010-0981
Unlimited Release
Printed February 2010

Final Report: Multicomponent Forensic Signature Development: Interactions with Common Textiles; Mustard Precursors and Simulants

Mark H. Van Benthem, Theodore T. Borek III,
Curtis D. Mowry, and Paul G. Kotula
Materials Characterization Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-0886

Abstract

2-Chloroethyl phenyl sulfide (CEPS), a surrogate compound of the chemical warfare agent sulfur mustard, was examined using thermal desorption coupled gas chromatography-mass spectrometry (TD/GC-MS) and multivariate analysis. This work describes a novel method of producing multiway data using a stepped thermal desorption. Various multivariate analysis schemes were employed to analyze the data. These methods may be able to discern different sources of CEPS. In addition, CEPS was applied to cotton, nylon, polyester, and silk swatches. These swatches were placed in controlled humidity chambers maintained at 23%, 56%, and 85% relative humidity. At regular intervals, samples were removed from each test swatch, and the samples analyzed using TD/GC-MS. The results were compared across fabric substrate and humidity.

ACKNOWLEDGMENTS

Funding for this project was by the Science and Technology Directorate of the U.S. Department of Homeland Security under contract WFO-018081112 with Sandia National Laboratories. This work was supported by Sandia National Laboratories and the Department of Energy. Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-ACO4-94AL85000.

CONTENTS

1. Introduction 9
   1.1. Mustard Gas 9
   1.2. Mustard Simulants 10
   1.3. Thermal Desorption Gas Chromatography-Mass Spectrometry 10
   1.4. Multivariate Data Analysis 11
      1.4.1. Data Scaling 11
      1.4.2. Principal component analysis (PCA) 12
      1.4.3. Orthogonal Factor Rotation 13
      1.4.4. Multivariate Curve Resolution (MCR) 13
      1.4.5. Trilinear Data Analysis 14
   1.5. Effects of Age and Humidity 15

2. Methods and Experimental 16
   2.1. Sample Preparation 16
      2.1.1. Dilution procedure, sample deposition 16
   2.2. TD/GC-MS 16
      2.2.1. Thermal desorption methods and instrumentation 16
      2.2.2 GC methods and instrumentation 16
   2.3. Data Processing and Analysis 17
   2.4. Fabric Spiking and Aging 17

3. Results 20
   3.1. TD/GC-MS Chemometric Analysis 20
   3.2. Fabric Aging Analysis 29

4. Conclusions 33

5. References 34

Appendix A – Manuscript Pending DHS Release 37

FIGURES

Figure 1. Structural formula for sulfur mustard (HD). 9
Figure 2. Structural formulas for selected mustard simulants. 10
Figure 3: Photo of TD/GC-MS instrumentation used in this work. 17
Figure 4. Swatch Samples for CEPS Aging Study 18
Figure 5. Total Ion Chromatograms of CEPS at Five Thermal Desorption Temperatures. 20
Figure 6. TICs of CEPS at Five TD Temperatures Showing CEPS Peak. 21
Figure 7. Higher Volatility Range TICs of CEPS at Five TD Temperatures. 21
Figure 8. Lower Volatility Range TICs of CEPS at Five TD Temperatures. 22
Figure 9. Eigenvalues of Poisson-scaled Data. 23
Figure 10. PCA-Varimax Solution Obtained for Region 12.7-13.7 Minutes. 23
Figure 11. MCR and PARAFAC Solutions Obtained for Region 12.7-13.7 Minutes. 25
Figure 12. Higher Volatility Range TICs of Neat “New” CEPS at Five TD Temperatures 27
Figure 13. Higher Volatility Range TICs of Neat “Old” CEPS at Five TD Temperatures. 27

Figure 14. Day 0, CEPS on Cotton. 30
Figure 15. Day 2, CEPS on Cotton, 23% RH. 30
Figure 16. Day 14, CEPS on Nylon, 56% RH. 31
Figure 17. Day 28, CEPS on Nylon, 56% RH. 32

TABLES

Table 1. Aging Study Thermal Desorption Parameters 18
Table 2. Aging Study GC-MS Analysis Parameters 19
Table 3. NIST Mass-spectral library matches for PCA-Varimax Model Factors* 24
Table 4. NIST Mass-spectral library matches for MCR Model Factors* 25
Table 5. Neat "New" CEPS library matches for PCA Model Factors 28
Table 6. Neat "Old" CEPS library matches for PCA Model Factors 28
Table 7. Neat "New" and "Old" CEPS library matches for MCR Model Factors 29
Table 8. Peak Area of 2-(phenyl thio)ethyl acetate, normalized to Day 2 31

NOMENCLATURE

ALS	alternating least squares
amu	atomic mass unit
ATD	automated thermal desorption
CWA	chemical warfare agent
CEES	chloroethyl ethyl sulfide
CEMS	2-chloroethyl methyl sulfide
CEPS	2-chloroethyl phenyl sulfide
DOE	Department of Energy
DHS	Department of Homeland Security
HD	sulfur mustard
HSI-CM	hyperspectral imaging-confocal microscopy
MCR	multivariate curve resolution
MS	methyl salicylate
MSD	mass spectrometer detector
NIPALS	nonlinear iterative partial least square
PARAFAC	parallel factors analysis
PCA	principal component analysis
RH	relative humidity
SNL	Sandia National Laboratories
SVD	singular value decomposition
TD/GC-MS	thermal desorption coupled gas chromatography-mass spectroscopy
TIC	total ion chromatogram
WWI	World War One
WWII	World War Two

1. INTRODUCTION

Chemical warfare agents (CWAs) are a significant concern as a terrorist actor's weapon of mass destruction and fear. Most CWAs are easy to synthesize and can be done so at fairly low cost, an ideal combination for small terror cells bent on instilling fear in a target population. As a result, they present a significant concern to policy makers and public safety officials. These weapons also pose a challenge for law enforcement agencies charged with post-attack investigation. Unlike commercially-produced toxic compounds, CWAs are likely to be produced, rather than purchased, by the end user. Thus, there is likely not a paper trail tying the end-product material directly to the criminal. Therefore, the forensics investigator must rely on linking the chemical precursors used in the synthesis to a suspect.

Identification of the sources of chemical reagents used in the "home" production of a chemical agent can be an important aspect in the forensic process. Complicating this problem, many CWAs have multiple synthetic pathways utilizing different chemical precursors. Given these multifarious challenges, a key in the forensic process is identifying and quantifying minor components and contaminants in the final synthetic product which link it to the precursors and/or synthetic route. One goal of this project sought to evaluate methods of identifying minor components in a simulant of one CWA, sulfur mustard.

Another consideration of this problem regards aging of CWAs. Since civilian structures may be quarantined after deployment of CWAs, many days may elapse before the collection of samples for forensic purposes could occur. Presumably the structure would be secured, assessed, decontaminated, and, finally, re-entered. In this regard, efforts here sought two objectives; what are the significant aging effects of a sulfur mustard surrogate based on substrate and relative humidity conditions, and how does one determine forensic information using relatively low-cost analytical instrumentation that is available to most metropolitan police crime laboratories so a local evaluation of evidence may be performed.

1.1. Mustard Gas

Mustard gas or sulfur mustard (HD), 1,1´-Thiobis[2-chloroethane], is an oily liquid with a weak, slightly sweet odor reminiscent of oil of mustard or garlic.[1-3] Figure 1 contains the structural formula for sulfur mustard. It is an alkylating agent and a strong vesicant that predominantly affects the eyes, skin and respiratory tract of its victims.[4] The first presumed report of sulfur mustard synthesis was in 1822, as the byproduct of reactions of sulfur chloride with ethylene.[5] This, and subsequent syntheses in 1860,[1, 6] were performed by chemists without a complete understanding of the hazards associated with their products. Sulfur mustard was finally isolated and purified in 1886 by Meyer.[7-8]

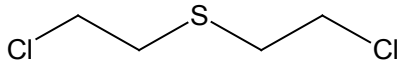

Figure 1. Structural formula for sulfur mustard (HD).

Sulfur mustard has been used intentionally and unintentionally in various military conflicts over the past 100 years.[5] It was used extensively in the First World War (WWI), by the Axis powers at various times preceding the Second World War, and by Iraq during the Iran-Iraq War and

against the Kurds prior to the 1991 Gulf War. During WWII, the mustard gas was released as a consequence of a German air raid on Allied ships moored in the port of Bari, Italy.[9-10]

As a weapon of terrorism, there are no reported uses of mustard gas, but its ease of fabrication and acute and chronic effects,[4, 8] make it a potent concern for policy makers.[11-12]

1.2. Mustard Simulants

Useful simulants or surrogates for mustard have been investigated for environmental studies and protective equipment testing.[13-14] These include 2-chloroethyl methyl sulfide (CEMS), 2-chloroethyl phenyl sulfide (CEPS), chloroethyl ethyl sulfide (CEES), and methyl salicylate (MS). For this work, CEPS was selected because of our familiarity with this compound, its structural similarity to HD, and its lower toxicity.

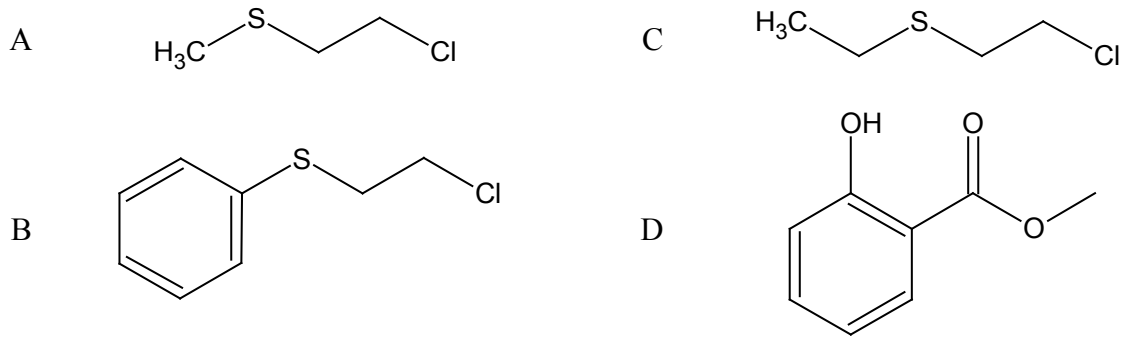

Figure 2. Structural formulas for selected mustard simulants.
The species depicted are CEMS (A), CEPS (B), CEES (C), and MS (D).
(See text for abbreviations.)

1.3. Thermal Desorption Gas Chromatography-Mass Spectrometry

Thermal desorption coupled with gas chromatography-mass spectrometry (TD/GC-MS) is a flexible experimental technique that permits the direct analysis of volatile organic species from liquid or solid substrates. TD/GC-MS also facilitates the indirect analysis of organic species when collected on a suitable adsorbent.

Thermal desorption is a sample preparation process that heats a sample to a specified temperature under an inert gas chromatography carrier gas. The volatile species desorbed from the sample substrate are collected and concentrated on an adsorbent-packed trap, which is typically held at a sub-ambient temperature during desorption from the original substrate. After desorption from the sample is complete, the trap is heated to volatilize the adsorbed species and introduce them to the gas chromatograph-mass spectrometer for separation and identification. Thermal desorption permits the direct analysis of volatile species from samples with little sample preparation; since no extraction solvents are used, dilution, solubility issues, and loss of analytes that may be part of a solvent extraction method are avoided.

Typically, TD/GC-MS is performed using a single, predetermined sample desorption temperature. This allows the chemist to extract materials that may be of interest, while not

damaging the collection medium or generating decomposition products. Another issue is to select a temperature that will not decompose the target analytes. A disadvantage to this technique is that all of the target materials are driven out of the sample in aggregate. An alternative strategy, which depends on the variable volatility of different species, is to step the desorption temperature to generate a series of desorptions.

The technique of stepping through a series of temperatures works as follows: (1) Perform an initial desorption at a temperature slightly above ambient temperature that is easy to maintain by the temperature controller. (2) Concentrate these higher-volatility species in the trap and then introduce them into the GC-MS. (3) After collection of the full-mass spectrum-chromatogram, step the temperature up to the next desired level and return to step one. (4) Repeat until the maximum desired desorption temperature is achieved. This process allows one to progressively extract the sample, removing the target species from the matrix as a function of volatility and/or adsorption affinity. In addition, by collecting a full mass spectrum for each elution time increment one can generate three-way data. Since three-way data that follow a trilinear model has some special properties with regard to multivariate analysis, this technique can be very advantageous.

The methods and instrumentation used in this study are widely available, unspecialized, and relative low cost. This is by design so that the methods developed can easily be replicated by the widest possible audience responsible for analysis or performing forensic investigations. As such, the GC-MS conditions are not optimized toward particular targets or species, but in fact are unrestricted in order to detect the widest possible range of marker compounds. For example, the scan range was set from 50 to 380 atomic mass units (amu) and scanned at each time increment. In the case of a forensic investigation targeted methods may be used to look in an optimized way for specific species, such as selected ion monitoring methods and smaller scan ranges or temperature ramps. For unknowns, critical information can be overlooked or not even detected.

As more knowledge is gained as to markers of particular CWAs and their synthesis methods, the chemical detection methods can be improved in the usual ways for lower detection limits, analysis speed, etc. By demonstrating these expanded techniques in a broad sense, we have not constrained large sections of the forensic population to purchase highly specialized, single purpose, or extremely expensive hardware.

1.4. Multivariate Data Analysis

1.4.1. Data Scaling

Prior to performing factor analysis, GC-MS data must be appropriately scaled so that it approximates the assumptions of the factor analysis technique. The factor analysis techniques used here are based on the method of least squares, which assumes that the errors are independently and identically distributed (*i.i.d.*) normal. Since these GC-MS data are collected as counts from a quadrupole mass spectrometer, the first principles assumption is that the data are actually Poisson distributed.[15-17] SNL has extensive experience with optimal scaling of multivariate Poisson-distributed data.[18-20] Briefly, the data can be scaled using the inverse of the square root of the mean mass spectrum.

Consider some GC-MS data in the $m \times n$ matrix $\mathbf{D}$ oriented as mass spectral domain by chromatographic domain with mean *m/z* spectrum $\bar{\mathbf{d}}_m$ given by

$$\bar{\mathbf{d}}_m = \frac{1}{n}\mathbf{D}\mathbf{1}_n \tag{1}$$

where $\mathbf{1}$ is an n-vector column of ones. Now, the data can be scaled in $\mathbf{D}$ using the diagonal matrix $\mathbf{H}$ whose diagonal elements are $\bar{\mathbf{d}}_m^{-1/2}$ using

$$\tilde{\mathbf{D}} = \mathbf{H}\mathbf{D} \tag{2}$$

where $\tilde{\mathbf{D}}$ is the data scaled for Poisson statistics. This scaling decreases the effect of large variations in the data due solely to noise in intense spectral regions. It is important for subsequent factor analysis as it effectively down-weights the effect of variance due to noise in intense spectral features and concomitantly up-weights minor spectral features, which in the raw data are smaller in magnitude than noise elsewhere.

1.4.2. Principal component analysis (PCA)

PCA is a statistical method that decomposes a matrix into two sets of orthogonal of basis vectors, ordered by decreasing variance, that model the row and column spaces of the matrix.[21-22] It is often used as an initial data reduction method, whose subspace representation may be readily factor-analyzed by additional statistical treatments.[18] PCA can be represented in matrix form as

$$\tilde{\mathbf{D}} = \tilde{\mathbf{T}}\mathbf{P}^{\mathrm{T}} + \tilde{\mathbf{E}} \tag{3}$$

where $\tilde{\mathbf{T}}$ is an $m \times p$ matrix which describes the row (or mass spectral) space of the scaled data in $\tilde{\mathbf{D}}$, $\mathbf{P}$ is an $n \times p$ matrix describing the column (or chromatographic) space of $\tilde{\mathbf{D}}$, $\tilde{\mathbf{E}}$ is an $m \times n$ matrix of scaled residuals or noise, and the superscript "T" indicates the transpose of the preceding matrix or vector. We use p to define the size, or pseudorank, of the subspace model that describes the chemically meaningful information contained in $\tilde{\mathbf{D}}$; simply put, the number of distinguishable chemical species in $\tilde{\mathbf{D}}$. $\tilde{\mathbf{T}}$ is orthogonal and $\mathbf{P}$ is orthonormal having the properties:

$$\begin{aligned} \mathbf{\Lambda} &= \tilde{\mathbf{T}}^{\mathrm{T}}\tilde{\mathbf{T}} \\ \mathbf{I} &= \mathbf{P}^{\mathrm{T}}\mathbf{P} \end{aligned} \tag{4}$$

where $\mathbf{I}$ is a $p \times p$ identity matrix and $\mathbf{\Lambda}$ is a $p \times p$ diagonal matrix of eigenvalues ordered from largest to smallest. One can also compute a "full set" of eigenvalues of length $min(m, n)$ for fairly low computational cost. These eigenvalues can be used to estimate the pseudorank, p, in numerous ways.[21] Commonly, a semi-logarithmic plot of eigenvalue versus factor number is produced and the number of factors selected where a "knee" occurs in the plot.[23]

There is a variety of methods to compute the PCA, among these are nonlinear iterative partial least square (NIPALS),[24] eigenanalysis,[25-26] and singular value decomposition (SVD).[26-27] SVD is very convenient since it decomposes the matrix $\tilde{\mathbf{D}}$ as

$$\tilde{\mathbf{D}} = \tilde{\mathbf{U}}\mathbf{S}\mathbf{V}^{\mathrm{T}} + \tilde{\mathbf{E}} \tag{5}$$

where **U** and **V** are, respectively, the $m \times p$ and $n \times p$ matrices of orthogonal left and right singular vectors and **S** is the diagonal matrix of singular values. The singular values are the square roots of the eigenvalues, *viz.*

$$\mathbf{\Lambda} = \mathbf{S}^2 \tag{6}$$

We can combine equations (3) and (5) to show that

$$\begin{aligned} \tilde{\mathbf{T}} &= \tilde{\mathbf{U}}\mathbf{S} \\ \mathbf{P} &= \mathbf{V} \end{aligned} \tag{7}$$

We have been careful in this section to utilize notation that indicates which data domain bears the scaling, in this case the mass-spectral domain. This is important since after factor analysis we will want to return those factors to their native scale, specifically

$$\mathbf{T} = \mathbf{H}^{-1}\tilde{\mathbf{T}} \tag{8}$$

1.4.3. Orthogonal Factor Rotation

After performing PCA, one will have an orthogonal, rank-*p* representation of the data that will probably not resemble any meaningful information to the chromatographer. Consequently, additional factor analysis is required to transform the PCA factors into interpretable factors. A reasonable approach to transforming the factors is to use a factor rotation method.

Factor rotation methods seek to maximize (or minimize) some criterion that is consistent with the nature of the data. In the case of GC-MS data, the varimax rotation[28-30] is an appropriate criterion for the chromatographic domain. The varimax criterion seeks an orthogonal rotation matrix, **R**, which maximizes the row (or time) variance of the orthonormal matrix **P**, thereby maximizing the "simplicity" of the rotated elution-time or chromatographic factors. In matrix form we have

$$\tilde{\mathbf{D}} = \tilde{\mathbf{T}}\mathbf{P}^{\mathrm{T}} = \tilde{\mathbf{T}}\mathbf{R}\mathbf{R}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}} = \underline{\tilde{\mathbf{T}}}\underline{\mathbf{P}}^{\mathrm{T}} \tag{9}$$

Generally, the factors of chromatographic domain are simple or sparse. For example, when a species elutes it produces a peak in the chromatogram, generating a chromatographic factor with a single peak, and zeros or noise at all other times. Unless another species co-elutes, all other factors will be zero-valued (or baseline-noisy) at the elution times encompassing that species peak. So, the chromatographic domain is, generally, sparse. Overlapping due to co-elution or the presence of a large background is a violation of this premise. By contrast, there is no reason to expect that the mass-spectral domain is sparse since many different compounds generate the same mass fragments, although not in the same *pattern*; so this violates the sparsity assumption.

1.4.4. Multivariate Curve Resolution (MCR)

If all species were to elute at different times such that none were overlapped, and no background arose from column packing loss and discharge, then PCA and varimax rotation would be sufficient to produce interpretable results. Unfortunately, one often has to deal with these and other problems, and so needs to find a method applicable to these issues. MCR, also called linear unmixing,[30] is a factor analysis method that utilizes an alternating least squares strategy

while employing constraints;[31-33] the most common constraint used being nonnegativity.[34-35] MCR seeks to solve

$$\tilde{\mathbf{D}} = \tilde{\mathbf{M}}\mathbf{C}^{\mathrm{T}} + \tilde{\mathbf{E}} \tag{10}$$

where $\tilde{\mathbf{M}}$ is the nonnegative $m \times p$ matrix modeling the row (or mass spectral) space of the scaled data in $\tilde{\mathbf{D}}$, $\mathbf{C}$ is the nonnegative $n \times p$ matrix modeling the column (or chromatographic) space of $\tilde{\mathbf{D}}$. Combining equations (10) and (9), we can form the relationship

$$\underline{\tilde{\mathbf{T}}}\underline{\mathbf{P}}^{\mathrm{T}} = \tilde{\mathbf{M}}\mathbf{C}^{\mathrm{T}} \tag{11}$$

which represents the dimension reduction of $\tilde{\mathbf{D}}$ as well as the imposition of non negativity constraints on the rotated PCA factors. Finally, following MCR, we would rescale the factors in $\tilde{\mathbf{M}}$ using the appropriate substitution into Eq. (8).

1.4.5. Trilinear Data Analysis

In its rudimentary data acquisition mode, GC-MS data is bilinear. That is, for an *R*-component chemical system, the data follow the model

$$d_{ij} = \sum_{p=1}^{R} m_{pi} c_{pj} \tag{12}$$

where d_{ij} is the intensity at mass element i and elution time j, m_{pi} is the concentration, or relative amount of the pth chemical species at mass element i, and c_{pj} is the unit-concentration response for the pth species at elution time j. While bilinear data is rich in information content, its decomposition into pure component concentrations and spectra is complicated by a mathematical peculiarity known as the rotation problem. The rotation problem arises because any solution for the factors matrices representing mass spectra and concentrations, *i.e.*, **M** and **C**. Simply stated, one can construct any invertible transformation matrix, $\mathbf{T}_i$, to operate on **M** and **C** producing an identical solution to the least-squares cost function. Thus, in a least squares sense

$$\mathbf{D} = \mathbf{M}\mathbf{C}^{\mathrm{T}} = \left(\mathbf{M}\mathbf{T}_i^{-1}\right)\left(\mathbf{T}_i\mathbf{C}^{\mathrm{T}}\right). \tag{13}$$

Often, we can get around the rotation problem using constraints such as nonnegativity and various equality and other constraints, as is done in MCR. However, there is no guarantee that the solution one obtains is the most accurate representation of the true underlying information. Fortunately, there exist data types and analysis methods that can overcome the rotation problem, even in the absence of constraints. Trilinear data and trilinear decomposition methods fall into this category. Since the TD/GC-MS allows us to add an additional acquisition mode, desorption temperature, which can vary linearly as do the mass spectral and concentration modes, producing information-rich, rotationally-unique trilinear data.

Trilinear data follow the model[36-41],

$$d_{ijk} = \sum_{p=1}^{R} m_{pi} c_{pj} f_{pk}\,, \tag{14}$$

where d_{ijk} is the d_{ij} measured at the kth thermal desorption temperature and f_{pk} is the concentration-independent desorption rate for species p evaluated at temperature interval k. The matrix or array representation of Eq. (14) is

$$\mathbb{D} = \otimes\left(\mathbf{M}, \mathbf{C}, \mathbf{F}\right), \tag{15}$$

where the fancy D is now a three-way array and the operator $\otimes$ represents the triple product of the pure component factor matrices **M**, **C**, and **F**. The solutions to the factor matrices can be estimated in a number of ways involving iterative methods and direct methods. This group has had success using an iterative method known as parallel factors analysis or PARAFAC, which uses an alternating least squares scheme similar to MCR. While MCR alternately estimates **M** given **C**, and then **C** given **M** until convergence; PARAFAC-ALS (alternating least squares) estimates **M** given **C** and **F**, then **C** given **M** and **F**, and finally **F** given **M** and **C**. This approach lends itself well to our established expertise and fast algorithms developed for use with MCR.

SNL has developed novel fast algorithms for analyzing large three-way data sets and have successfully applied them to hyperspectral imaging-confocal microscopy (HSI-CM) photobleach data. [42-43]

1.5. Effects of Age and Humidity

To investigate the effects of humidity on the aging of CEPS, swatches of common fabrics were spiked with neat surrogate and were placed in constant humidity chambers. After a set period of time, test samples were cut from the swatches and the coupons analyzed using thermal desorption with gas chromatography with mass spectrometry detection.

2. METHODS AND EXPERIMENTAL

2.1. Sample Preparation

2.1.1. Dilution procedure, sample deposition

CEPS (CAS 5535-49-9) was purchased from Sigma-Aldrich (St. Louis, MO), henceforth "New" CEPS, and another sample of Sigma Aldrich that was obtained from older laboratory stock, henceforth "Old" CEPS. Some CEPS samples were diluted using hexanes and stored in capped sample vials. Dilution factors were calculated by mass. When looking for low-level marker compounds in the CEPS solution, a high concentration standard (7123 ng/µl) was used. A 40 µl aliquot was deposited onto unsilanized glass wool inside the quartz desorption tubes and allowed to air dry for 30 seconds before sealing for analysis. This permitted the light hexane solvent to nearly completely evaporate prior to analysis, but retained the CEPS and other less volatile species. Alternatively, some tests were performed using neat CEPS deposited onto glass wool. This limited the complication of solvent-based trace contaminants corrupting the results, but is only qualitative as mass or volume measurement of sample quantity is difficult.

2.2. TD/GC-MS

2.2.1. Thermal desorption methods and instrumentation

Samples were placed in desorption tubes and inserted into a Perkin Elmer model TurboMatrix ATD thermal desorption unit. The automated thermal desorption (ATD) was operated with an inlet valve temperature of 50°C, while maintaining the transfer line at 240°C. The sample trap low temperature was set at -30°C in order to collect the sample, and then ramped to 300°C at a heating rate 40°C/min to revolatilize the sample. The system used a purge time of one minute, a desorb time one minute, and a trap hold time of 5.0 min. A cycle time 60 min was chosen for these experiments, for compatibility with the expected elution times. The inlet split was off and outlet split on, and two-stage desorb mode was employed. The column pressure was set at 14.5 psi, outlet split of 36.3 mL/min, with desorb flow 46 mL/min, and inlet split of 197 mL/min.

Several temperature profile experiments were performed to investigate the temperature steps that would provide additional thermal separation of components, while simultaneously generating three-way data suitable for PARAFAC analysis. After a number of "trial and error" experiments a desorption temperature series of 50 °C, 65 °C, 80 °C, 100 °C, and 125°C was chosen since it provided good analyte response and thermal separation of the minor component species observed in the CEPS samples.

2.2.2 GC methods and instrumentation

All separations were performed using an Agilent 6890N gas chromatograph with an Agilent model 5975 inert XL MSD mass spectrometer (Agilent Technologies, Santa Clara, CA). The instrument was equipped with a HP-1701 GC column, 60 m × 0.32 mm ID and 0.1 µm film thickness. Temperature profile parameters were 50°C, hold 2 min; ramp at 6°C/min to 230°C, hold 20 min, for a total time of 52 min. Mass spectrometer details were to scan 50 to 380 amu (4.27 scans/sec) with samples at 2, threshold equal to 0, and sampling rate 2^2.

The photo in Figure 3 shows the instrumentation used in this work. The large tube above the instruments is the insulated capillary transfer line between the thermal desorption autosampler

and the gas chromatograph's heating inlet. The glass sample tubes on the circular stage at right contain the glass wool on which the sample is deposited. The tubes are rotated into place and heated internally in the ATD unit and swept with helium to release chemical constituents. Within the ATD is a thermoelectrically cooled trap of small internal volume that cryogenically traps the constituents that are released. At a predetermined time, the trap is rapidly heated to release the constituents which travel through the transfer line into the gas chromatograph. Inside the GC oven is a long capillary separation column that separates the constituents in time and introduces them into the mass spectrometer detector (MSD) for detection.

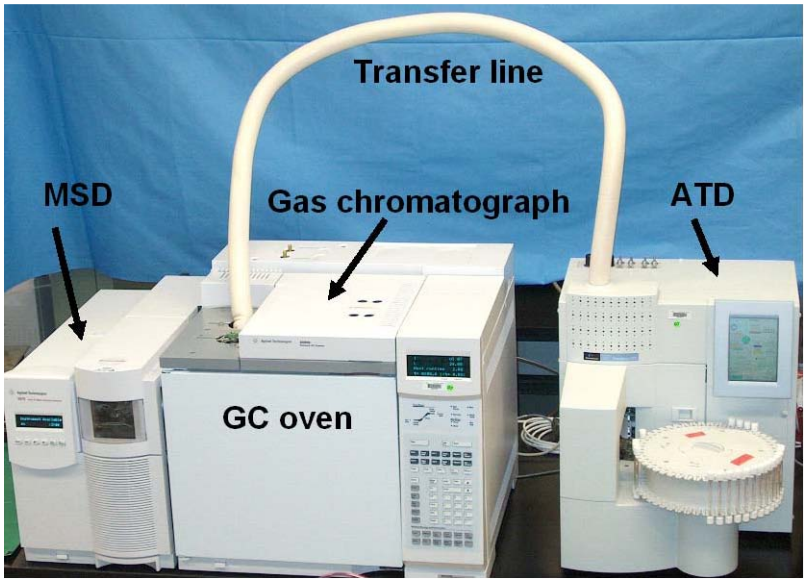

Figure 3: Photo of TD/GC-MS instrumentation used in this work. From left to right, mass spectrometer detector (MSD), gas chromatograph (GC), and automated thermal desorption (ATD).

2.3. Data Processing and Analysis

Agilent data were converted to MATLAB® file format using MassTransit software (Palisade Corp, Newfield, NY). All data processing and analysis were performed using SNL authored programs written in MATLAB®[44] m-file language. All computations were performed using Dell Precision 690 equipped with two, dual-core, 3.2 GHz Xeon processors and 4.0 Gbyte RAM; operating under Windows XP Professional Version 2; and running MATLAB® version 7.8.0.347 (R2009a) and later. PCA was performed using a freely available SVD algorithm written in MATLAB® m-file format.[45]

2.4. Fabric Spiking and Aging

CEPS (Sigma-Aldrich, St. Louis, MO) was used as received. Undyed swatches of cotton, ripstop nylon, polyester, and silk were purchased from a local fabric store and used as received. Humidity chambers were constructed using desiccator jars with a water-saturated salt in the base: potassium acetate (23% RH), magnesium nitrate (56% RH) and potassium chloride (85% RH). Relative humidity was confirmed with a VWR Thermo-Hygro hygrometer.

Samples were prepared by cutting a 1.125 inch diameter circle from the selected fabric as shown Figure 4.

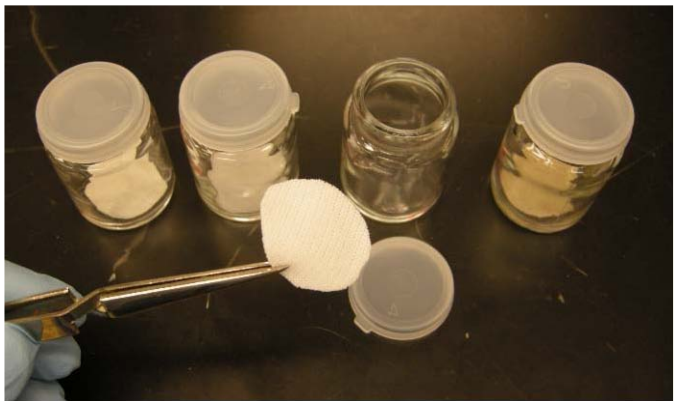

Figure 4. Swatch Samples for CEPS Aging Study

After spiking, each swatch was placed in a glass jar, and each jar, without a lid, was placed in a constant humidity chamber for aging. Three 2×4 mm samples were cut from each swatch at Days 0, 2, 7, 14, 21, and 28.

Thermal desorption was performed using a Perkin-Elmer TurboMatrix thermal desorption unit operated in the combined inlet and outlet split mode. Thermal desorption parameters are shown in Table 1. Samples for thermal desorption were centered in a glass desorption tube using glass wool. Prior to use, the desorption tube had been conditioned at 380°C and was demonstrated to be free of organic species.

Table 1. Aging Study Thermal Desorption Parameters

Parameter	Value
Desorption Temperature	200°C
Desorption Time	5 minutes
Desorption Flow	31.6 mL/min
Trap Temperature, Trapping	-30.0°C
Trap Temperature, Injection	300°C
Injection time	5 minutes
Inlet split	150 mL/min
Outlet split	51.0 mL/min
Valve Temperature	240°C
Transfer Line Temperature	240°C

Gas chromatography-mass spectrometry was performed using an Agilent 6890 N gas chromatograph coupled with an Agilent 5975 mass spectrometer (Agilent Technologies, Santa Clara, CA) operated in the electron ionization mode. An Agilent Technologies HP-1701 column, 60 m × 0.32 mm with a 1.0 µm film thickness, was used in the analysis. Analysis parameters are shown in Table 2.

Table 2. Aging Study GC-MS Analysis Parameters

Parameter	Value
Initial Temperature	100°C, hold for 3 minutes
Ramp 1	30°C/min
Temperature 2	280°C, hold for 5 minutes
Ramp 2	20°C/min
Temperature Final	300°C, hold for 10 minutes
Solvent Delay	3.30 minutes
Scan range	33 to 380 amu
Scan rate	4.08 spectra per second
Source Temperature	230°C
Quadrupole Temperature	150°C

The peaks observed were tentatively identified by computer comparison of the resulting mass spectrum with a NIST mass spectral library, and were confirmed by retention time matching using known materials where available.

3. RESULTS

3.1. TD/GC-MS Chemometric Analysis

Data were collected essentially as an array of mass counts and organized as mass spectral mode by chromatographic elution time by desorption temperature. Fractional mass channel data were summed into unit mass spectral elements. The total size of the arrays was then 331 spectral elements (50-380 amu) by 13,260 time elements (0-52 min, ~ 0.24 s increments) by five temperatures (50, 65, 80, 100, and 125 °C). Figure 5 contains the plot of the total ion chromatograms (TICs) for the five thermal desorption temperatures of a sample of CEPS diluted in hexane and then evaporated. Here, the sample is the "Old" CEPS. The chromatograms are dominated by a single peak at ~32 minutes, which is the CEPS peak.

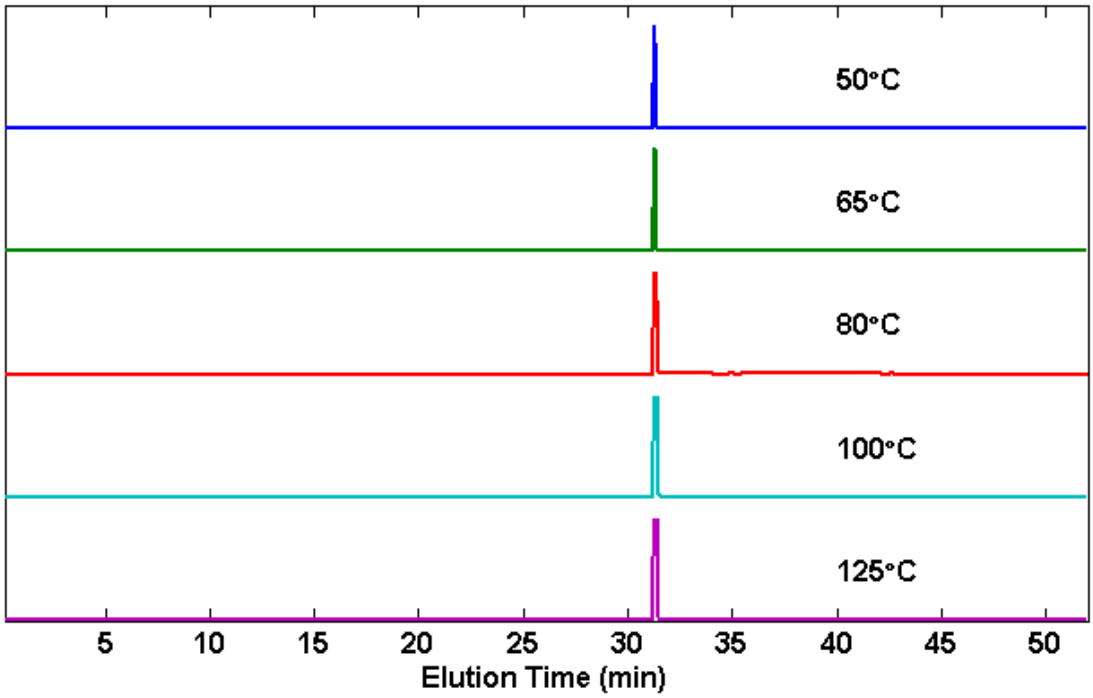

Figure 5. Total Ion Chromatograms of CEPS at Five Thermal Desorption Temperatures. Shown are total ion chromatograms (TICs) of CEPS after dilution in hexane. Dilution and evaporation was computed by summing all mass channels. From top to bottom are the TICs for thermal desorption temperatures 50, 65, 80, 100, and 125°C. Stagger on y-axis is 25M counts from baseline to baseline.

While it is not obvious in Figure 5, the CEPS peak indicates column saturation, which is readily apparent in Figure 6. Not only is the column saturated, where the number of theoretical plates is exceeded as is reflected in the severe distortion and spreading of the peak, the mass detector is also saturated. Mass detector saturation presents a very serious challenge for multivariate analysis. Under conditions of detector saturation, a single chemical species will need many factors to describe it mathematically; as a result it will appear to be many species during an analysis. One tactic for overcoming this problem is to simply eliminate the offending peak. Figure 7 is a plot of the data in Figure 5 restricted to the first 31.125 minutes and scaled up by a factor of 70. None of the peaks in this region of higher volatile compounds indicate column or detector saturation. In addition, they display peak height variability as a function of temperature. This behavior is desirable and bodes well for multiway analysis.

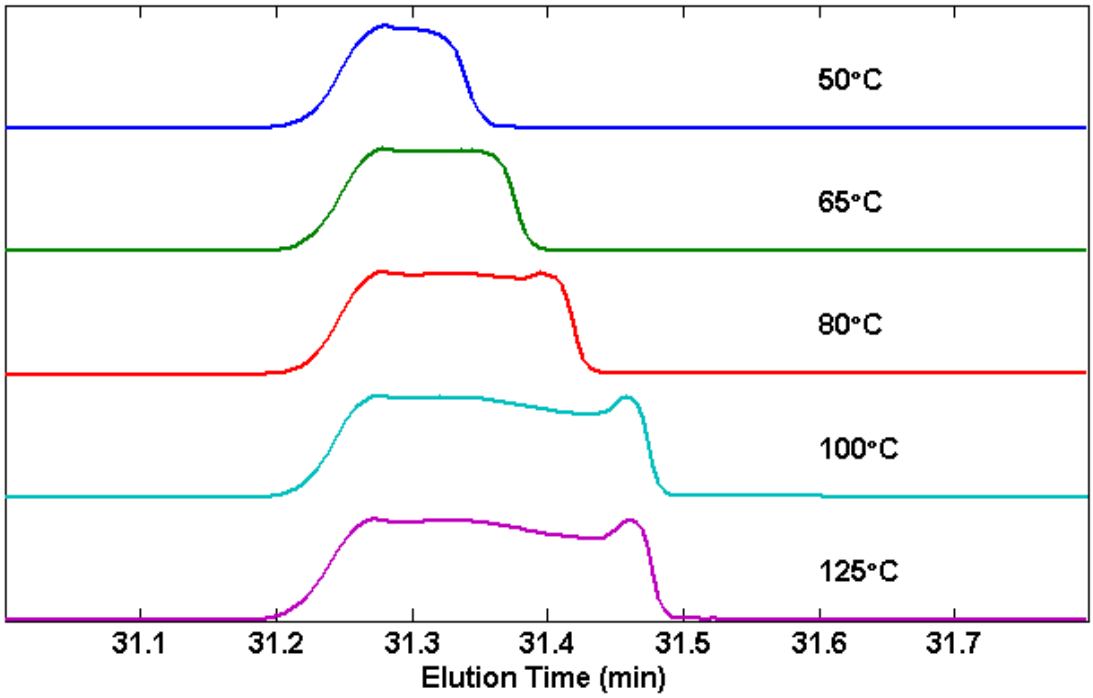

Figure 6. TICs of CEPS at Five TD Temperatures Showing CEPS Peak.
The TICs above are those from Figure 5 over a limited elution time range to highlight the CEPS peak. The progression through the desorption temperatures illustrates the extent of column saturation..

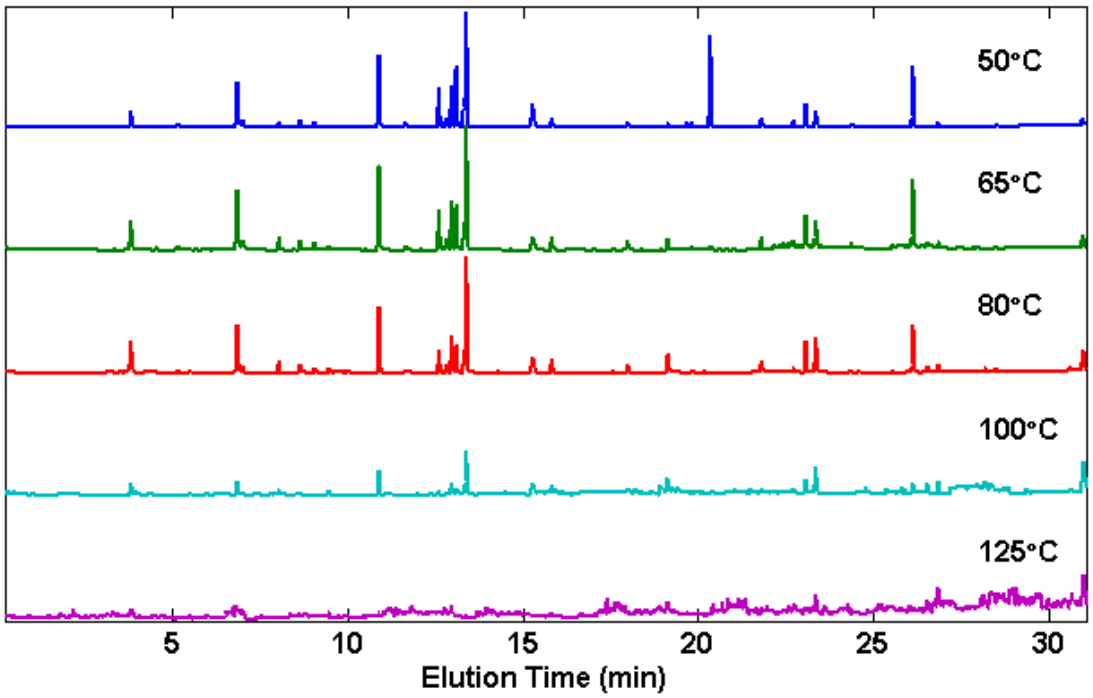

Figure 7. Higher Volatility Range TICs of CEPS at Five TD Temperatures.
The TICs above are those from Figure 5 expanded by a factor of 70. Note the variability in the ratios of peaks heights as a function of temperature. This clearly illustrates the effects of volatility and/or adsorption differences for the species represented by the peaks. Each chromatogram has 7916 elements. Stagger on y-axis is ~357k counts baseline to baseline.

Figure 8 contains an amplified-scale of the lower-volatility range of the previously described data. Again, there is no indication of saturation and there is good peak variability from temperature to temperature.

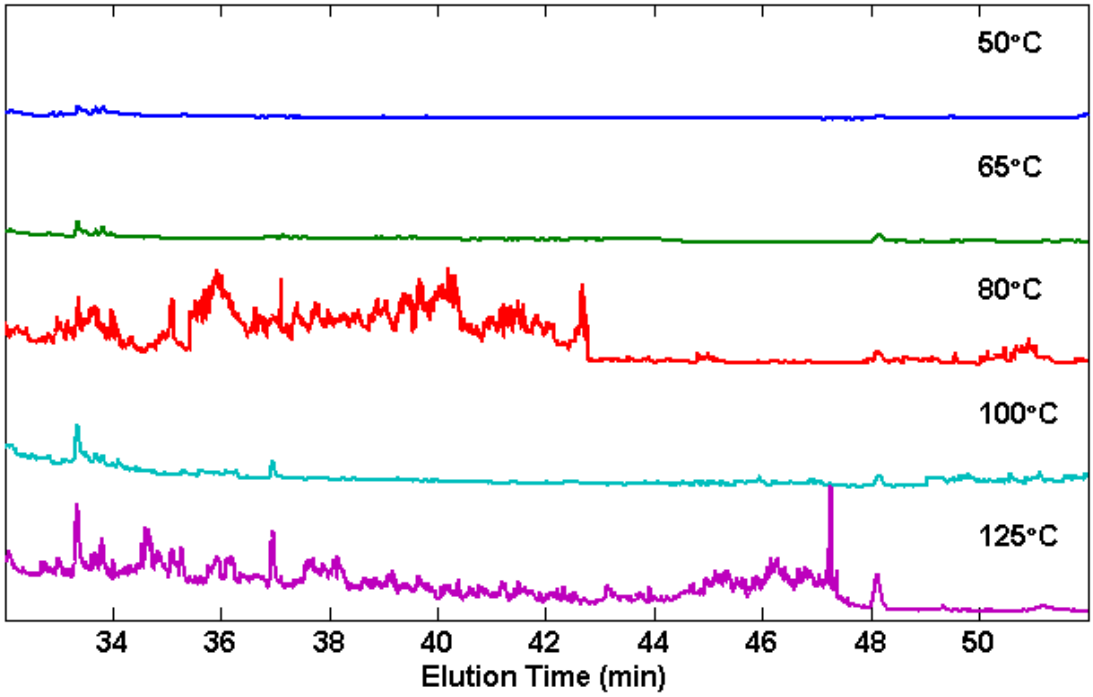

Figure 8. Lower Volatility Range TICs of CEPS at Five TD Temperatures.
The TICs above are those from Figure 5 expanded by a factor of 85. Again, note the variability in the ratios of peaks heights as a function of temperature. In contrast to the lower volatility region, peaks tend to "grow in" with increasing desorption temperature. Stagger on y-axis is ~294k counts.

Prior to multivariate analysis the data were Poisson-scaled using the mass spectral grand mean of the array as described in Eq. (2). Following scaling, the data were subjected to eigenanalysis to estimate the pseudorank. Since the reduced data is now 331×7916×5 this is not a daunting task, and is accomplished by reorganizing the data as a 331×39580 matrix. A plot of the eigenvalues for the higher-volatility data in Figure 7 is displayed in Figure 9. Ideally this type of plot would have a rather distinctive cutoff between the signal eigenvalues and the noise eigenvalues. Such a pattern would be represented by a decreasing set of signal eigenvalues from left to right, ending somewhere above a set of noise eigenvalues of essentially constant magnitude. This plot does not indicate a clean cutoff, and so rank estimation is more difficult. A close evaluation reveals approximately 48 factors for these data.

After estimating a rank for decomposition, the next step is to perform a data reduction and factor analysis step using PCA. Following PCA, the factors are rotated as described above using varimax rotation to impose elution time simplicity. A sample region of 12.7 to 13.7 minute containing five peaks (species) and five of the 48 factors of PCA-varimax analysis are displayed in Figure 10. These results indicate a very nice match of the data over this region.

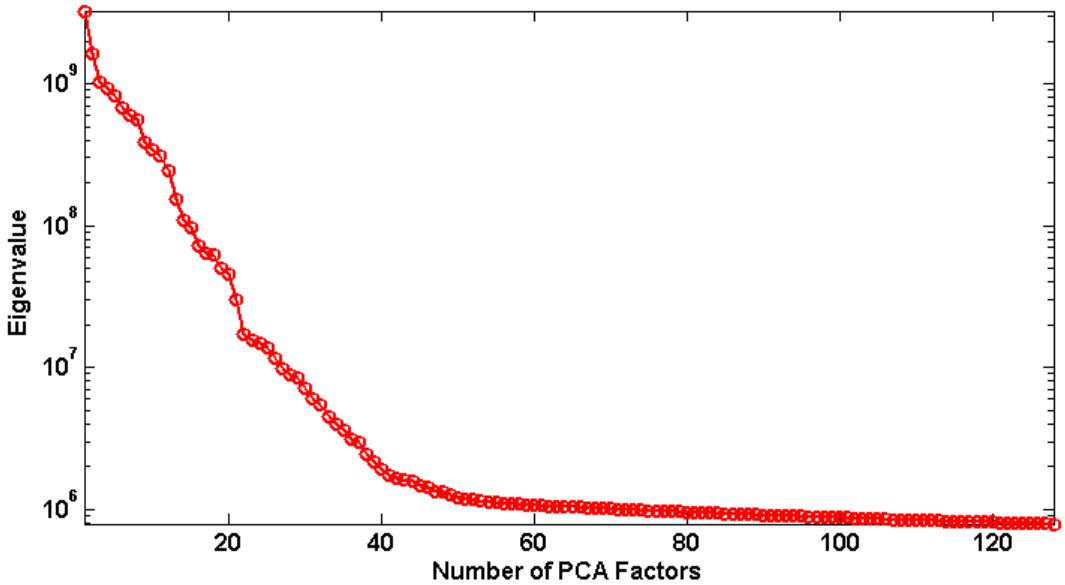

Figure 9. Eigenvalues of Poisson-scaled Data.
Eigenvalues of the data depicted in Figure 7 were computed using the 331×7916×5 data array arranged as a 331×39580 matrix. Only 128 of the computed 331 eigenvalues are shown here.

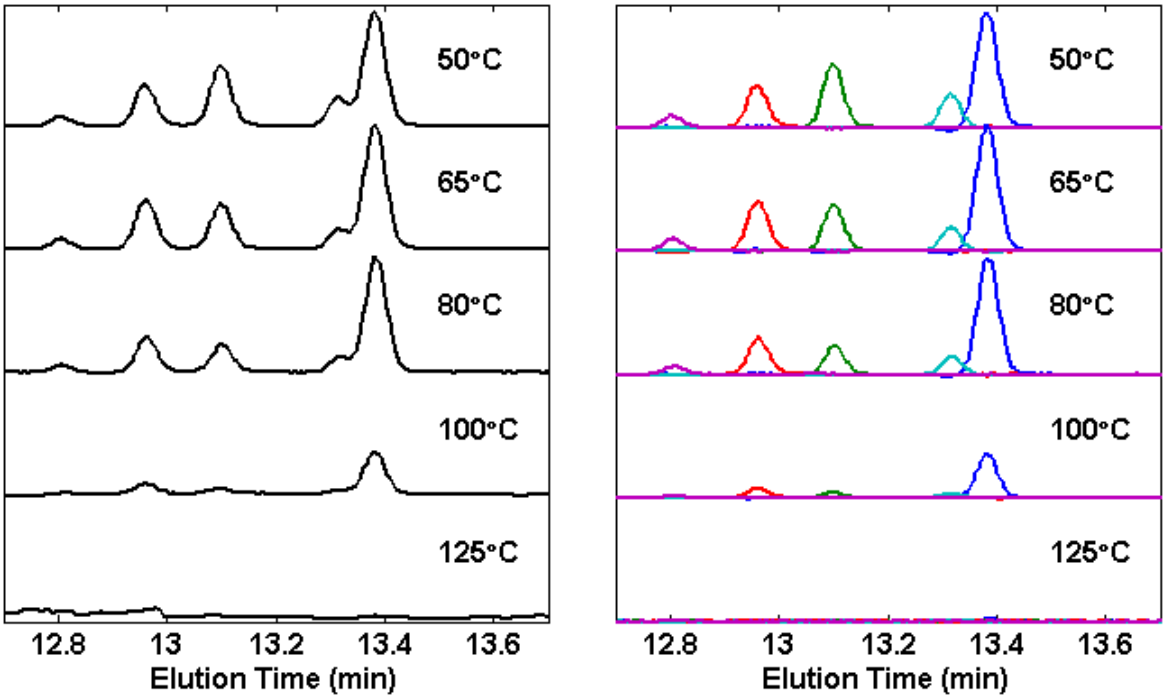

Figure 10. PCA-Varimax Solution Obtained for Region 12.7-13.7 Minutes.
On the left are plotted the expanded scale TIC of data displayed in Figure 7 of the elution time region12.7 to 13.7minutes. Five varimax-rotated PCA components that described the data over the same region are plotted on the right. Factor numbers for peaks from left-right are 30, 10, 6, 13 and 1.

Following PCA, the mass spectral factors were entered into the NIST mass spectrum matching routine. Table 3 is a partial list of the matches that exceeded a match value of 800 as well as those factors whose factor chromatograms are depicted in Figure 10.

Table 3. NIST Mass-spectral library matches for PCA-Varimax Model Factors*

PCA Factor	NIST Assigned Compound	Formula	Match value	CAS #
8	Ethene, chloro-	C_2H_3Cl	954	75-01-4
16	Benzene	C_6H_6	932	71-43-2
3	Cyclopentene, 3-methyl-	C_6H_{10}	922	1120-62-3
11	Benzene, (ethenylthio)-	C_8H_8S	898	1822-73-7
15	Benzenethiol	C_6H_6S	867	108-98-5
5	Octane, 1-chloro-	$C_8H_{17}Cl$	852	111-85-3
6	3-Hexanol	$C_6H_{14}O$	850	623-37-0
25	Decane, 1-chloro-	$C_{10}H_{21}Cl$	842	1002-69-3
1	Cyclopentanol, 1-methyl-	$C_6H_{12}O$	818	1462-03-9
9	Phenol	C_6H_6O	818	108-95-2
13	5-Hexen-3-ol, 2,2,4-trimethyl-	$C_9H_{18}O$	701	90676-50-9
10	5-Nonanone	$C_9H_{18}O$	699	502-56-7
30	2-Pentene, 4,4-dimethyl-, (E)-	C_7H_{14}	683	690-08-4

* Shaded rows are for factors depicted in Figure 10.

In addition to PCA and varimax rotation, factors were estimated using MCR and PARAFAC. The results for the same elution time region depicted in Figure 10 are presented in Figure 11. MCR was initialized using the PCA-varimax results and was performed using the reduced dimension data representation of the PCA-varimax factors rather than the full data set. Nonnegativity constraints were imposed in both the elution time and mass spectral modes. The results appear very similar to the PCA-varimax solution. The most notable difference is the elimination of the negative excursion of factor 1 (blue) around 13.3 min in PCA, which has a near-baseline value in MCR. Table 4 contains a list of the NIST mass spectrum matches that exceeded a match value of 800 as well as those factors whose factor chromatograms are depicted in Figure 10. Note that MCR tends to find more components with high scores than PCA. However, this is due more likely to the property of nonnegativity than the quality of the solution. All of the MCR mass spectral factors are all-positive while the PCA mass spectral factors have some small (and a few relatively large) negative values. These negative values certainly complicate the mass spectral recognition process.

The PARAFAC solution presented in Figure 11 was obtained while employing nonnegative constraints in all three modes during analysis. The plots exhibit an interesting feature with this methodology; there is substantial mixing of the peak at 13.4 min and the peak at 13.0 min. This mixing is due to the number of common mass spectral peaks observed in the peaks at 13.0 and 13.4 minutes, such as m/z 41, 57, 58, and 85. While the mixing may degrade identification of component peaks, it may actually increase the subsequent ability to determine the similarity (and therefore forensic quality) of data from different samples. The relative effectiveness, in a forensic context, of the MCR and PARAFAC solutions therefore needs to be assessed with

additional data. The additional computing burden of performing the PARAFAC analysis is minimal.

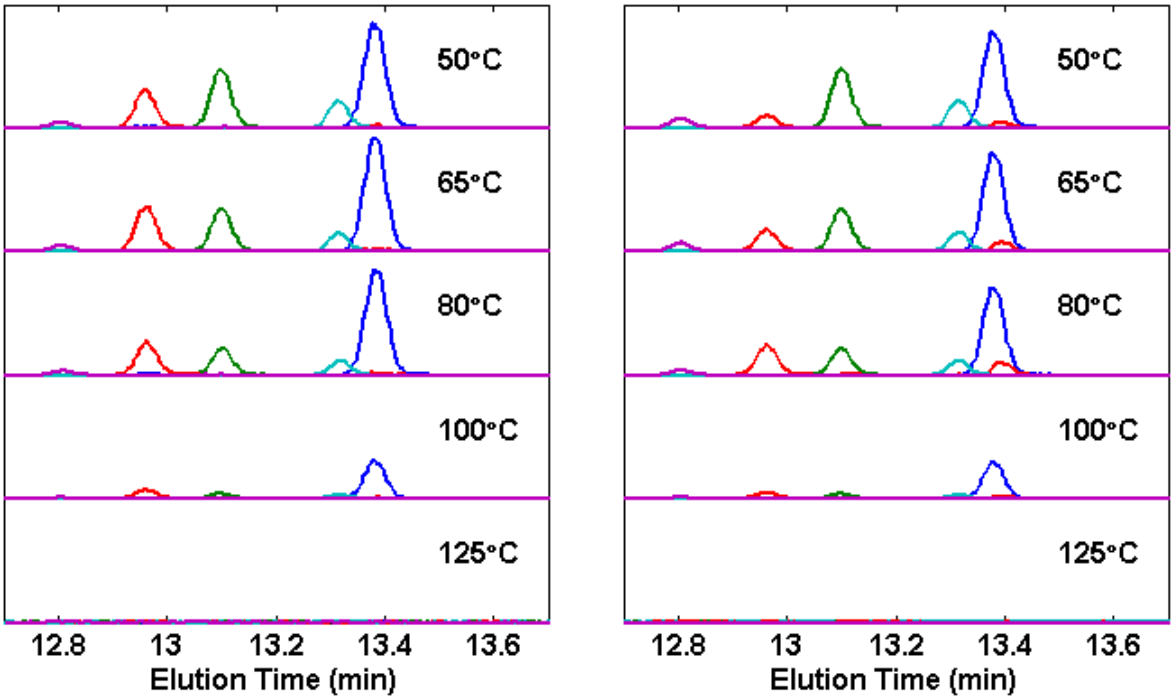

Figure 11. MCR and PARAFAC Solutions Obtained for Region 12.7-13.7 Minutes. Five MCR components (left) and PARAFAC components (right) that describe the data that are plotted on the right in Figure 10 are shown here. Factor numbers for peaks from left-right are 30, 10, 6, 13 and 1.

Table 4. NIST Mass-spectral library matches for MCR Model Factors*

MCR Factor	NIST Assigned Compound	Formula	Match value	CAS #
8	Ethene, chloro-	C_2H_3Cl	958	75-01-4
7	Biphenyl	$C_{12}H_{10}$	920	92-52-4
11	Benzene, (ethenylthio)-	C_8H_8S	910	1822-73-7
15	Benzenethiol	C_6H_6S	904	108-98-5
16	Benzene	C_6H_6	898	71-43-2
3	Cyclopentene, 1-methyl-	C_6H_{10}	897	693-89-0
14	2-Benzothiophene #	C_8H_6S	888	270-82-6
9	Phenol	C_6H_6O	847	108-95-2
25	Decane, 1-chloro-	$C_{10}H_{21}Cl$	842	1002-69-3
6	3-Hexanol	$C_6H_{14}O$	826	623-37-0
1	Cyclopentanol, 1-methyl-	$C_6H_{12}O$	808	1462-03-9
5	Octane, 1-chloro-	$C_8H_{17}Cl$	808	111-85-3
27	(R)-(+)-3-Methylcyclopentanone	$C_6H_{10}O$	805	6672-30-6
30	2-Propenoic acid, anhydride	$C_6H_6O_3$	784	2051-76-5
13	4-Octanol, 2-methyl-	$C_9H_{20}O$	701	40575-41-5
10	5-Nonanone	$C_9H_{18}O$	681	502-56-7

* Shaded rows are for factors depicted in Figure 11.

The MCR factors provide higher library match scores, in general; but also more factors that "mix" multiple peaks with similar mass spectral components. While this hinders some identification, it may actually benefit the differentiation of samples.

After evaluating the solvent extracted CEPS, the next step was to look at neat CEPS. In this case, the two CEPS sources were used, "Old" and "New." Figure 12 contains a plot of the higher volatility range for five thermal desorptions of the neat "New" CEPS. The most striking difference between this and the data in Figure 7 is the reduced complexity of the data. One may conclude that use of a solvent to extract the CEPS will make the analysis much more complicated and the solvent must also be analyzed carefully for contaminants, regardless of the quality or the grade. Figure 13 is a corresponding plot of the progressive temperature TD/GC-MS of the "Old" sample.

There is an obvious problem in the 65°C plot in the "Old" CEPS data in Figure 13, where there is a large irregularly structured artifact in the time range 15 to 18 minutes. An inspection of the mass spectra for the artifact reveals a large magnitude, nearly uniformly distributed spectrum. This may be a mass detector or electronic problem in the instrument or some external electronic noise. In any case, it is certainly not from a chemical analyte source. Fortunately, the region in which it is found is devoid of any real signal in the other TD chromatograms. Since, this would have a real influence on the analysis; that region was simply zeroed-out in the 65°C data, while the other TD chromatograms were untouched.

The two sets of plots look very similar with only some minor differences in proportions of peaks, which is certainly important in discriminating one from another. The rank analyses indicate that the Aldrich sample has a rank of 14 while the SNL sample has a rank of 17. In this sense, rank is the number of independent mass spectral components that are found in the data. It may not be equal to the number of chemical species if, say, two species have the same mass spectrum.

Tables 5 and 6 report the NIST mass spectrum matches that exceed a match value of 800 for the PCA-varimax analysis of the neat "New" and "Old" CEPS samples, respectively. The "New" sample analysis shows that eight of the 14 mass spectra were identified with a fairly high score, while the "Old" sample had 11 of 17 identified with reasonable confidence. One will note that one of the compounds (C_8H_8S) has two associated peaks. The compounds identified by factor 6 for the "New" and "Old" sample have very similar library mass spectra. Because of this high level of similarity, the spectral identification is somewhat ambiguous. It is likely that both compounds are present in both samples, and while the mixing degrades absolute identification it may as mentioned increase the ability to differentiate samples. In comparing the PCA model factors with higher than 800 library match values; both samples have a benzothiophene (isomer) peak identified at 26.5 minutes, 6 identical matches, and a mixed factor with very similar library spectra.

Table 7 contains the NIST mass spectrum matches for all of the MCR analysis of the neat "New" and "Old" CEPS samples, respectively. Again, MCR apparently performs better, however this is due, in part, to the nonnegative nature of the resolved mass spectra. In the MCR model the "Old" sample has an additional phenol peak. While the samples are highly similar, as expected given the common source, they do not appear to be identical as demonstrated by the PCA and

MCR factor models. In addition, processing the data with these algorithms allows for sample differentiation and chemical identification of individual peaks.

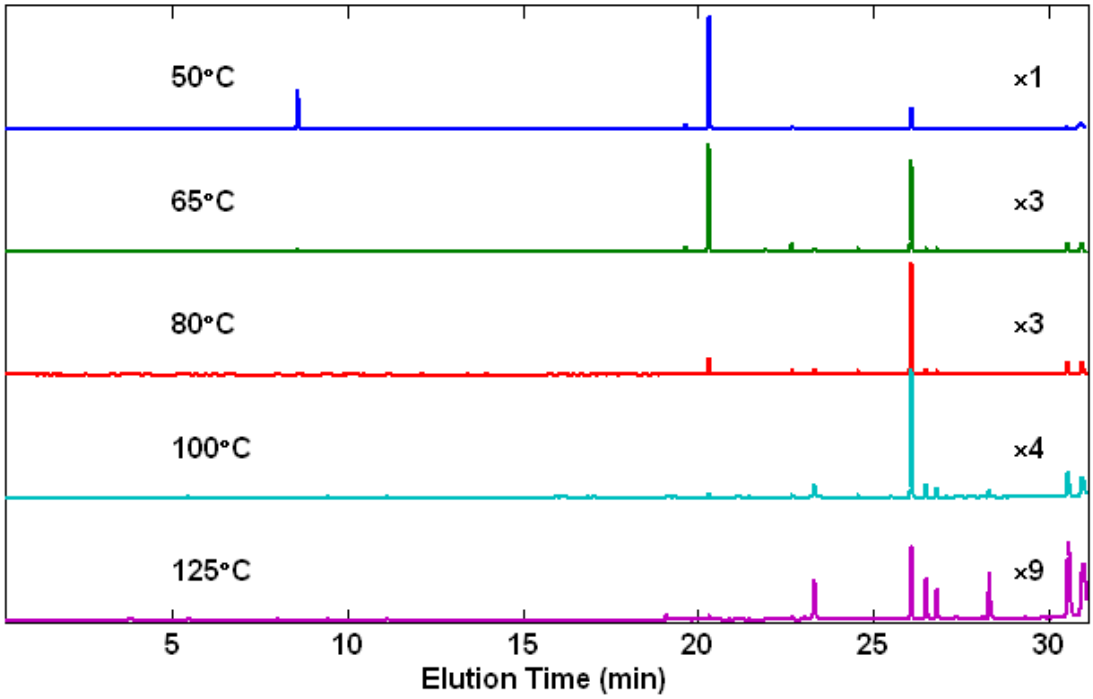

Figure 12. Higher Volatility Range TICs of Neat "New" CEPS at Five TD Temperatures.
TICs of neat the "New" CEPS measured for thermal desorption temperatures 50, 65, 80, 100, and 125°C.

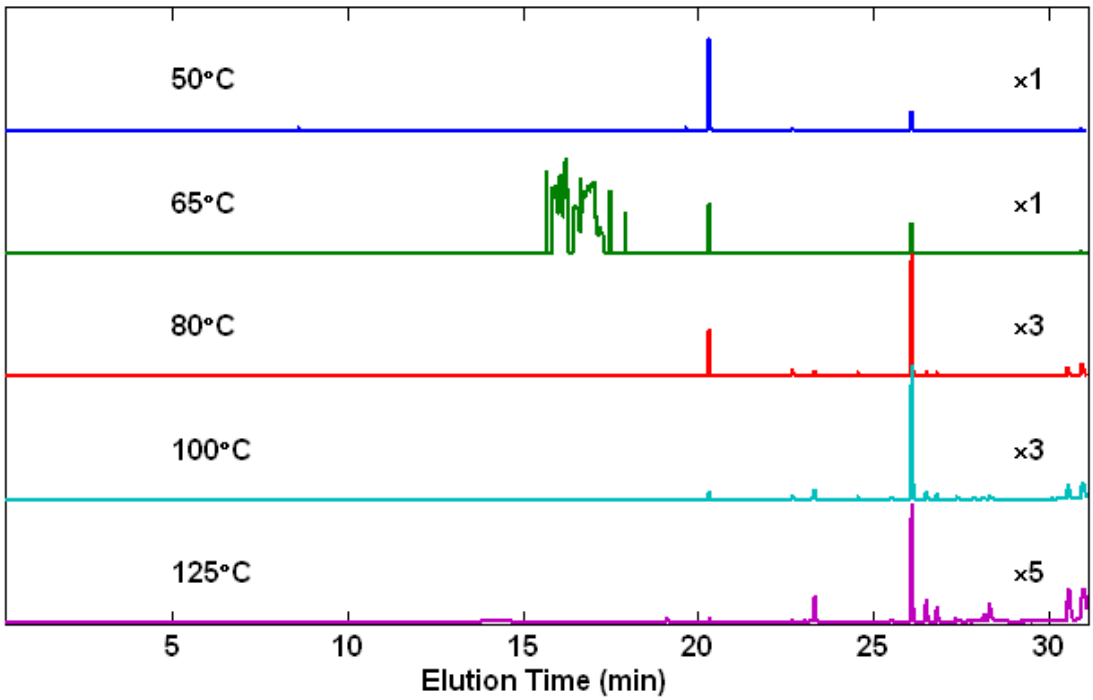

Figure 13. Higher Volatility Range TICs of Neat "Old" CEPS at Five TD Temperatures.
TICs of neat the "Old" CEPS measured for thermal desorption temperatures 50, 65, 80, 100, and 125°C. The 65°C TIC has a large artifact from approximately 15 to 18 minutes.

Table 5. Neat "New" CEPS library matches for PCA Model Factors

PCA Factor	NIST Assigned Compound	Formula	Match value	CAS #	Time(s)	
3	Biphenyl	$C_{12}H_{10}$	929	92-52-4	31.0	
5	Benzo[b]thiophene	C_8H_6S	923	95-15-8	26.5	
9	Benzene, (methylthio)-	C_7H_8S	908	100-68-5	22.7	
6	Benzo[b]thiophene, 2,3-dihydro-	C_8H_8S	894	4565-32-6	23.3	28.3
4	Acetaldehyde, (phenylthio)-	C_8H_8OS	887	66303-55-7	30.6	
2	Ethane, 1,2-dichloro-	$C_2H_4Cl_2$	881	107-06-2	8.6	
7	Decane, 1-chloro-	$C_{10}H_{21}Cl$	864	1002-69-3	26.1	
1	Octane, 1-chloro-	$C_8H_{17}Cl$	861	111-85-3	20.3	

Table 6. Neat "Old" CEPS library matches for PCA Model Factors

PCA Factor	NIST Assigned Compound	Formula	Match factor	CAS #	Time(s)	
1	Octane, 1-chloro-	$C_8H_{17}Cl$	950	111-85-3	20.3	
5	Decane, 1-chloro-	$C_{10}H_{21}Cl$	940	1002-69-3	26.1	
4	Benzo[c]thiophene	C_8H_6S	937	270-82-6	26.5	
2	Biphenyl	$C_{12}H_{10}$	925	92-52-4	31.0	
8	Benzene, (methylthio)-	C_7H_8S	924	100-68-5	22.7	
7	Ethane, 1,2-dichloro-	$C_2H_4Cl_2$	923	107-06-2	8.6	
3	Acetaldehyde, (phenylthio)-	C_8H_8OS	915	66303-55-7	30.6	
6	Benzene, (ethenylthio)-	C_8H_8S	910	1822-73-7	23.3	28.3
11	Octanal	$C_8H_{16}O$	891	124-13-0	19.6	
16	Octane, 1-chloro-	$C_8H_{17}Cl$	890	111-85-3	5.5	†
12	Benzene, nitro-	$C_6H_5NO_2$	840	98-95-3	24.6	

† chromatographic factor contains multiple small peaks

Table 7. Neat "New" and "Old" CEPS library matches for MCR Model Factors

"New" Time(s) min	"New" match value	NIST Assigned Compound	Formula	CAS #	"Old" match value	"Old" Time(s) min
26.5	933	Benzo[b]thiophene *	C8H6S	95-15-8	939	26.5
31.0	926	Biphenyl	$C_{12}H_{10}$	92-52-4	937	31.0
22.7	912	Benzene, (methylthio)-	C_7H_8S	100-68-5	910	22.7
23.3, 28.3	899	Benzene, (ethenylthio)-	C_8H_8S	1822-73-7	898	23.3, 28.3
30.6	891	Acetaldehyde, (phenylthio)-	C_8H_8OS	66303-55-7	922	30.6
30.9	876	2-Chloroethyl phenyl sulfide	C_8H_9ClS	5535-49-9		
8.6	874	Ethane, 1,2-dichloro-	$C_2H_4Cl_2$	107-06-2	940	8.6
26.1	867	Decane, 1-chloro-	$C_{10}H_{21}Cl$	1002-69-3	937	26.1
20.3	853	Octane, 1-chloro-	$C_8H_{17}Cl$	111-85-3	959	20.3
24.6	805	Benzene, nitro-	$C_6H_5NO_2$	98-95-3	864	24.6†
19.6	764	Octanal	$C_8H_{16}O$	124-13-0	940	19.6
N/A+	748	Endrin	$C_{12}H_8C_{l6}O$	72-20-8		
23.0†	720	p-hydroxyphenyl-Phosphonic acid	$C_6H_7O_4P$	33795-18-5		
26.8	678	1,2-Benzenedithiol, 4-methyl-	$C_7H_8S_2$	496-74-2	695	26.8
		Phenol	C_6H_6O	108-95-2	851	23.0†
		Pentane, 3-methyl-	C_6H_{14}	96-14-0	806	5.5†
		Benzenethiol	C_6H_6S	108-98-5	770	19.1
		Olean	$C_{30}H_{50}O_6$	15399-43-6	713	N/A+
		Ruthenium organometallic	$C_{14}H_{21}BO_3RuSeSi$	118772-38-6	687	N/A‡
		1-Decanol	$C_{10}H_{22}O$	112-30-1	681	28.2†

† chromatographic factor contains multiple small peaks
* "Old" NIST Assigned Compound: Benzo[c]thiophene
+ both "New" and "Old" results appear as rising structureless background, with noisy mass spectrum
‡ weakly structured chromatogram, with broad continuous mass spectrum

Independent library searching of the data (not shown) validates both the identification and the retention time of all but factor 16 in Table 6. For those factors with two elution times listed, the independent search matches the first listed time. Thus, the MCR analysis here is providing the same identification for these data as the commercial library search tool.

3.2. Fabric Aging Analysis

The analysis of the Day 0 sample from the cotton substrate is shown in Figure 14. The major species observed is CEPS, and this peak is saturated under the conditions of the analysis. This result is typical of all of the fabric substrates analyzed on Day 0.

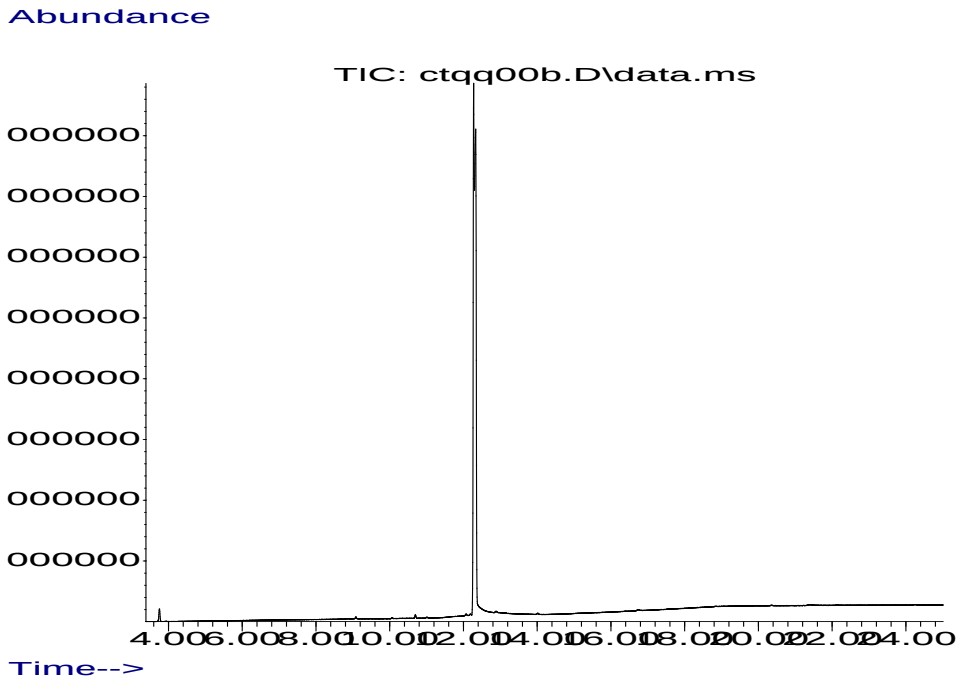

Figure 14. Day 0, CEPS on Cotton.

The first observed change was found in the Day 2 cotton sample from the 23% RH chamber. This total ion chromatogram is shown in Figure 15. In addition to CEPS, 2-(phenyl thio)ethyl acetate (CAS 20-93-1) is observed. This species was not observed on any of the other fabrics from 23% RH chamber, nor was it observed in the 56% RH and 85% RH samples.

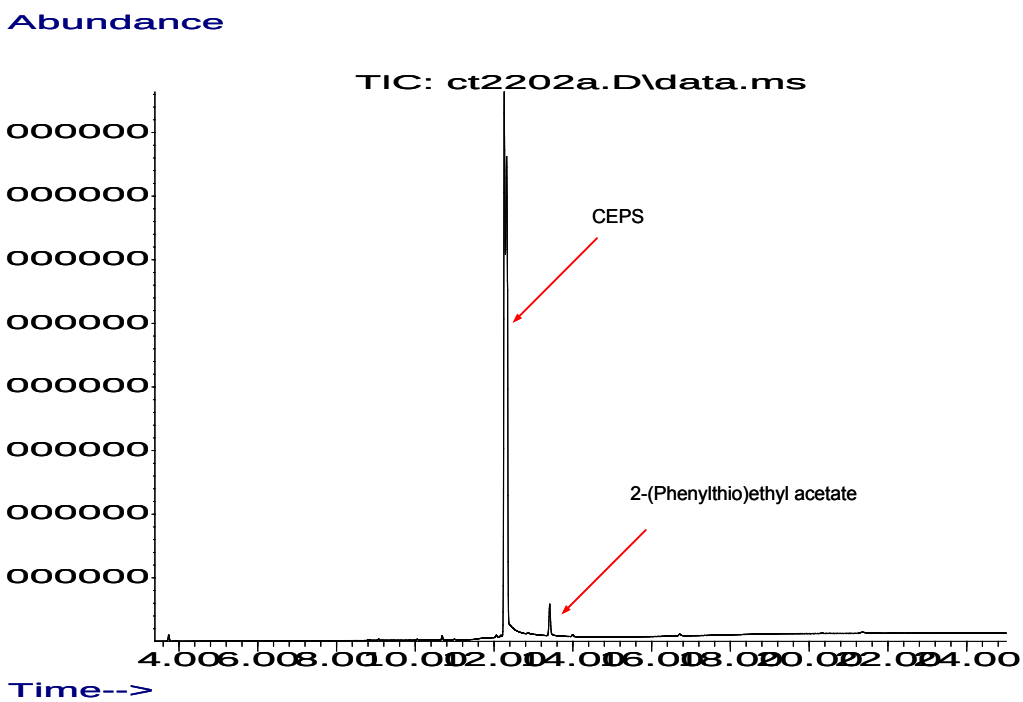

Figure 15. Day 2, CEPS on Cotton, 23% RH.

By Day 7, 2-(phenyl thio)ethyl acetate was detected on the nylon, polyester, and silk substrates from 23% RH chamber. The signal strength continued to increase with time on all 23% RH samples; the normalized response of 2-(phenyl thio)ethyl acetate from Day 2 to Day 28 is summarized in Table 8.

Table 8. Peak Area of 2-(phenyl thio)ethyl acetate, normalized to Day 2

Day	Peak Area Normalized to Day 2
2	100.0
7	196.9
14	387.8
21	636.6
28	675.2

The generation of 2-(phenylthio)ethyl acetate is due to the presence of acetic acid in the vapor phase of the 23% RH chamber reacting with the CEPS on the substrates; a gas phase sample of the 23% RH chamber revealed the presence of acetic acid in the vapor phase.

The second observed change was found in the Day 14 cotton and nylon samples from the 56% RH chamber. This total ion chromatogram for nylon is shown in Figure 16. In addition to CEPS, 2-(Phenylthio)ethanol (CAS 699-12-7) is observed.

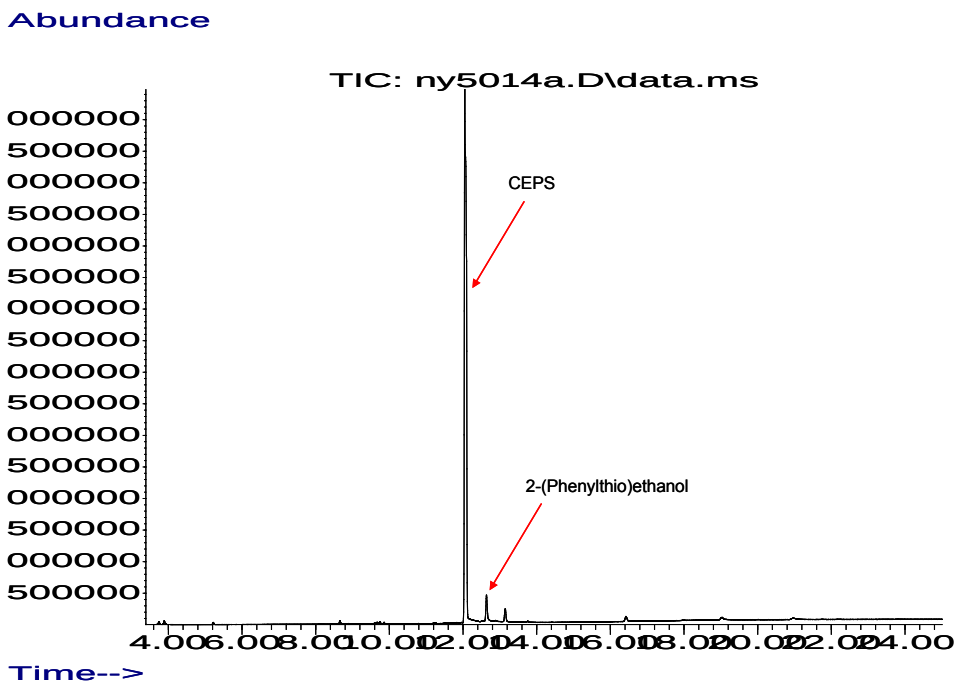

Figure 16. Day 14, CEPS on Nylon, 56% RH.

As the nylon aged, diphenyl disulfide (CAS 882-33-7), was identified as shown in Figure 17. No changes were observed in the 85% RH chamber samples until Day 28, when a trace amount of 2-(phenylthio)ethanol was observed on all samples.

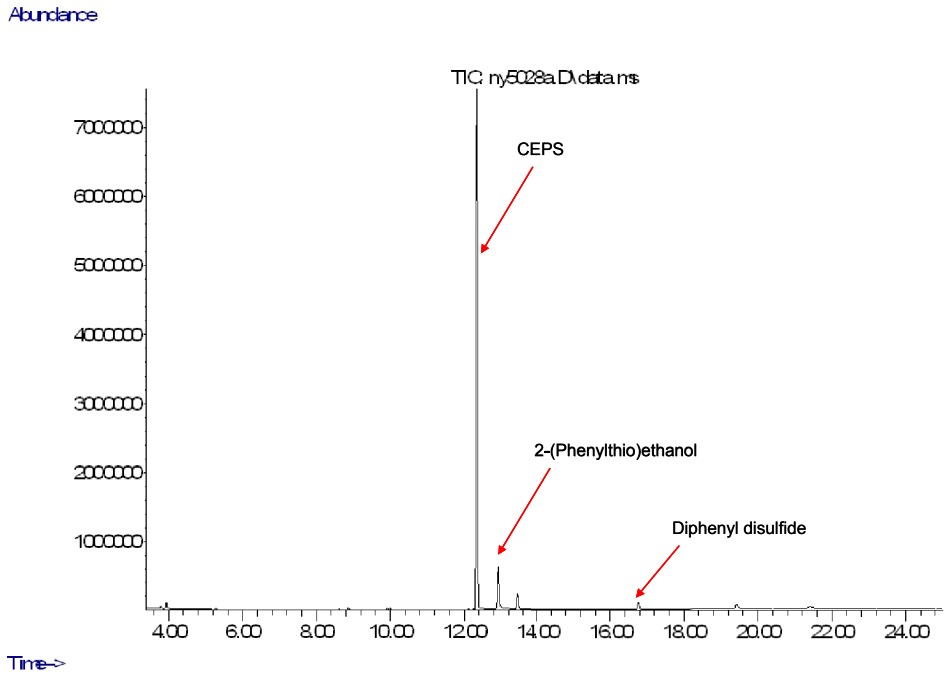

Figure 17. Day 28, CEPS on Nylon, 56% RH.

The major CEPS aging product observed in this study was 2-(phenylthio)ethyl acetated, which is a reaction product of CEPS and acetic acid. This product was only observed in the 23% RH chamber, which was maintained with a water-saturated potassium acetate salt mass in the desiccator base. No other product was observed in the 23% RH samples. In the higher humidity chambers, the hydrolysis product 2-(phenylthio)ethanol was observed earliest at 14 days in the 56% RH chamber, and at 28 days in the 85% RH chamber.

The use of the thermal desorption/gas chromatograph/mass spectrometer system produced reproducible results from the samples provided. This relatively inexpensive system was able to prepare and analyze samples and show differences in the aged materials without the use of extraction solvents.

4. CONCLUSIONS

We have presented results of a novel method for examining surrogates of a CWA. These methods may have the ability to discern different sources of the same compound. Our novel methods involve utilizing a stepped TD/GC-MS data acquisition scheme that may be almost totally automated, coupled with multivariate analysis schemes. These multivariate methods are extensible to other types of GC-MS analyses, as can be seen in the manuscript attached in Appendix A.

The algorithms used here were not developed specifically for GC/MS data, yet chemically relevant information is extracted as evidenced by the high match factor in the library searches. Broad search parameters were used, the full spectral library was searched (i.e. no CWA sublibrary was needed), no training set was needed, and the interaction between algorithm output and the NIST search was not optimized. No assumptions about peak width, quality, or overlap parameters that are necessary for conventional deconvolution software were needed, and no alignment of peaks across data files was required. In short, less user input was required, fewer data quality constraints were necessary, and yet very high quality chemical information was obtained. The results shown in the Tables demonstrate that these algorithms can process this type of data and show benefits over currently used techniques.

While it is true that good peak intensity was not a requirement for obtaining a high match factor (the peaks in Table 8 represent a range of intensities), further work includes dealing with noise issues to improve deconvolution of very low intensity peaks.

A chemical warfare agent surrogate's aging products may be observed using thermal desorption/gas chromatography/mass spectrometry. The aging products observed vary with the aging environment, and caution will be needed when interpreting any analytical result from the aftermath of an attack on a civilian structure if the ambient conditions after agent deployment are not known.

These studies provide an excellent introduction into the types of analyses that can be conducted in a standard forensics laboratory without the use of exotic equipment or exceptional expertise. Further development of these methods will facilitate technology transfer of these techniques to users in the field.

5. REFERENCES

1. F. Guthrie, "On some derivatives from the olefines," *Q. J. Chem. Soc.* **12**(109 - 126 (1860)
2. S. Budavari, M. J. O'Neil, A. Smith and P. E. Heckelman, Eds., *The Merck Index*, 11th ed., Merck & Co., Inc., Rahway, NJ (1989).
3. D. Evison, D. Hinsley and P. Rice, "Chemical weapons," *British Medical Journal* **324**(7333), 332-335 (2002)
4. J. S. Graham, *et al.*, "Wound healing of cutaneous sulfur mustard injuries: strategies for the development of improved therapies," *J Burns Wounds.* **4**(e1. (2005)
5. L. Szinicz, "History of chemical and biological warfare agents," *Toxicology* **214**(3), 167-181 (2005)
6. M. Gomberg, "Ethylene Chlorohydrin and β,β-Dichloroethyl - Sulfide," *J. Am. Chem. Soc.* **41**(9), 1414-1431 (1919)
7. V. Meyer, "Ueber Thiodiglykolverbindungen," *Berichte der deutschen chemischen Gesellschaft* **19**(2), 3259-3266 (1886)
8. K. Kehe and L. Szinicz, "Medical aspects of sulphur mustard poisoning," *Toxicology* **214**(3), 198-209 (2005)
9. Force Health Protection & Readiness Policy & Programs, "Mustard Disaster at Bari," http://fhp.osd.mil/CBexposures/ww2mustard.jsp, Date Accessed: 01/19/2010.
10. Department of the Navy -- Naval Historical Center, "Naval Armed Guard Service: Tragedy at Bari, Italy on 2 December 1943," (2006) http://www.history.navy.mil/faqs/faq104-4.htm, Date Accessed: 01/19/2010.
11. G. C. Oehler, "Continuing Threat from Weapons of Mass Destruction," *in Senate Armed Services Committee*, Central Intelligence Agency, Washington, D.C. (1996), https://www.cia.gov/news-information/speeches-testimony/1996/go_toc_032796.html, Accessed: 01/19/2010.
12. "Terrorist CBRN: Materials and Effects," Central Intelligence Agency, Washington, D.C., CTC 2003-40058, (2003).
13. S. L. Bartelt-Hunt, D. R. U. Knappe and M. A. Barlaz, "A review of chemical warfare agent simulants for the study of environmental behavior," *Critical Reviews in Environmental Science and Technology* **38**(2), 112-136 (2008)
14. V. Dubey, T. Parmar, A. Saxena and D. D. Agarwal, "Thiocompounds as Simulants of Sulphur Mustard for Testing of Protective Barriers," *J. Appl. Polym. Sci.* **111**(2), 928-933 (2009)
15. C. Hackett, *et al.*, "Laser ablation sampling for induction coupled plasma source quadrupole mass spectrometry of Cd-Zn-Te materials," *J. Electron. Mater.* **28**(6), 774-784 (1999)
16. I. S. Begley and B. L. Sharp, "Characterisation and correction of instrumental bias in inductively coupled plasma quadrupole mass spectrometry for accurate measurement of lead isotope ratios," *J. Anal. At. Spectrom.* **12**(4), 395-402 (1997)
17. B. El Abbassi, H. Mestdagh and C. Rolando, "Automatic extraction of relevant peaks and reconstruction of mass spectra for low signal-to-noise GC-MS data," *Int. J. Mass Spectrom. Ion Processes* **141**(2), 171-186 (1995)
18. M. R. Keenan, "Multivariate Analysis of Spectral Images Composed of Count Data," *in Techniques and Applications of Hyperspectral Image Analysis* H. F. Grahn and P. Geladi, Eds., pp. 89-126, John Wiley & Sons, Ltd, Chichester, West Sussex, England (2007).

19. M. R. Keenan and P. G. Kotula, "Optimal scaling of TOF-SIMS spectrum-images prior to multivariate statistical analysis," *Appl. Surf. Sci.* **231/232**(240-244 (2004)
20. M. R. Keenan and P. G. Kotula, "Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images," *Surf. Interface Anal.* **36**(3), 203-212 (2004)
21. E. R. Malinowski, *Factor Analysis in Chemistry*, John Wiley and Sons, Inc., New York (1991).
22. M. R. Keenan, "Maximum likelihood principal component analysis of time-of-flight secondary ion mass spectrometry spectral images," *J. Vac. Sci. Technol. A* **23**(4), 746-750 (2004)
23. M. L. Braun, "Spectral Properties of the Kernel Matrix and their Relation to Kernel Methods in Machine Learning," *Dissertation in Mathematics*, University of Bonn, Bonn (2005).
24. P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Anal. Chim. Acta* **185**(1-17 (1986)
25. R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, : Prentice Hall, Englewood Cliffs, N.J. (1992).
26. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore (1996).
27. R. M. Larsen, "Lanczos bidiagonalization with partial reorthogonalization," Department of Computer Science - Daimi, Aarhus University, Aarhus, Denmark, DAIMI PB-357, (1998).
28. R. J. Rummel, *Applied Factor Analysis*, Northwestern University Press, Evanston, IL (1970).
29. H. H. Harman, *Modern Factor Analysis*, University of Chicago Press, Chicago (1976).
30. M. R. Keenan, J. A. Timlin, M. H. Van Benthem and D. M. Haaland, "Algorithms for constrained linear unmixing with application to the hyperspectral analysis of fluorophore mixtures," *in Imaging Spectrometry VIII* S. S. Shen, Ed., pp. 193-202, SPIE - The International Society for Optical Engineering, Seattle, WA, USA (2002).
31. R. Tauler, E. Casassas and A. Izquierdo-Ridorsa, "Self-modelling curve resolution in studies of spectrometric titrations of multi-equilibria systems by factor analysis," *Anal. Chim. Acta* **248**(447-458 (1991)
32. M. H. Van Benthem, M. R. Keenan and D. M. Haaland, "Application of equality constraints on variables during alternating least squares procedures," *J. Chemom.* **16**(12), 613-622 (2002)
33. P. G. Kotula, M. R. Keenan and J. R. Michael, "Automated analysis of SEM X-ray spectral images: a powerful new microanalysis tool," *Microsc Microanal* **9**(1), 1-17 (2003)
34. R. Bro and S. DeJong, "A fast non-negativity-constrained least squares algorithm," *J. Chemom.* **11**(5), 393-401 (1997)
35. M. H. Van Benthem and M. R. Keenan, "Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems," *J. Chemom.* **18**(10), 441-450 (2004)
36. R. A. Harshman and M. E. Lundy, "Parafac : Parallel Factor-Analysis," *Computational Statistics & Data Analysis* **18**(1), 39-72 (1994)
37. R. Bro, "PARAFAC. Tutorial and applications," *Chemometrics Intell. Lab. Syst.* **38**(2), 149-171 (1997)
38. J. B. Kruskal, "Rank, Decomposition, and Uniqueness for 3-way and N-way Arrays," *in Multiway Data Analysis* R. Coppi and S. Bolasco, Eds., pp. 7-18, North-Holland, Amsterdam (1989).

39. B. C. Mitchell and D. S. Burdick, "Slowly Converging parafac sequences: Swamps and two-factor degeneracies," *J. Chemom.* **8**(155-168 (1994)
40. S. E. Leurgans and R. T. Ross, "Multilinear models: Applications in spectroscopy," *Statistical Science* **7**(3), 289-319 (1992)
41. G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Computational Statistics & Data Analysis* **50**(7), 1700-1734 (2006)
42. M. H. Van Benthem and M. R. Keenan, "Tucker1 model algorithms for fast solutions to large PARAFAC problems," *J. Chemom.* **22**(5), 345-354 (2008)
43. M. H. Van Benthem, *et al.*, "Trilinear Analysis Of Images Obtained With A Hyperspectral Imaging Confocal Microscope," *J. Chemometrics* (Published Online: Jul 9 2008)
44. "MATLAB," The MathWorks, Inc., Natick, MA (2009).
45. R. M. Larsen, "PROPACK," http://sun.stanford.edu/~rmunk/PROPACK/, Stanford, CA (2004).

APPENDIX A – MANUSCRIPT PENDING DHS RELEASE

FACTOR ANALYSIS OF GC/MS DATA WITH SELECTED ION DEPLETION

Mark H. Van Benthem, Theodore T. Borek III, Curtis D. Mowry, and Paul G. Kotula*

Sandia National Laboratories, Albuquerque, NM 87185-0886

Correspondence to: Mark H. Van Benthem
Sandia National Laboratories
MS0886
Albuquerque, NM 87185-0886
Voice: (505) 844-5443
Fax: (505) 844-2974
Email: mhvanbe@sandia.gov

ABSTRACT

We present a fast, robust, and automated multivariate statistical analysis method for the analysis of GC/MS data sets. Our method involves systematic elimination of undesired, saturated peak masses to yield data that follow a linear, additive model. The cleaned data are then subjected to a combination of principal component analysis (PCA) and orthogonal factor rotation followed by multivariate curve resolution (MCR) to yield highly interpretable results. Our analysis of two data sets on standard mixtures illustrates the power of the technique, highlighted by separation and identification of heavily overlapped peaks caused by an isotopic chlorine effect.

INTRODUCTION

Hyphenated chromatographic methods are some of the most powerful techniques for analysis of mixtures available.[1] Gas chromatography coupled with mass spectrometry (GC/MS) is one of the most widely utilized methods for the analysis of volatile species. A very powerful technique for improving the selectivity and signal to noise ratio (SNR) of GC/MS is selected ion monitoring (SIM).[2] SIM works by limiting the mass spectral filter to those ions associated with a chemical species of interest.[3] It can produce spectacular results when the analysis is limited to a few species. However, when the analysis is intended to find minor or unknown species in a mixture, use of SIM could result in missing important information.

There have been numerous methods developed to analyze GC/MS data over the years. One early method focused on identifying *m/z* indices which represented the pure chromatogram of each chemical component.[4] Knorr, et al.[5] took a different tack by modeling chromatogram peaks as skewed Gaussian functions while using a SIMPLEX minimization. Various modifications and advancements of the first theme[6-7] have led to a very popular and versatile approach,[8] which has

become known as AMDIS.[9] These approaches use the mass spectral information as well as assumptions about peak shape to "deconvolute" the chromatogram.

Factor analysis methods that assume no analytic model of the character of the data are often called soft models or matrix-based methods. Such self modeling methods typically make only limited assumptions about the nature of the data, but impose constraints such as nonnegativity to achieve physically meaningful results. Alternating least squares (ALS) or alternating regression (AR) methods have been employed successfully in GC/MS analysis.[10] Unfortunately, that algorithm was initiated with random starting points, which can lead to solutions trapped in local minima. We will present what we believe to be a better alternative.

Analysis of forensic samples presents a case where the identification of minor or unsuspected chemical species is very important. Minor components may present themselves as evidence of the source of a compound used in a criminal activity, such as an accelerant in arson or a byproduct from production of material used in a chemical attack. Separating the small but important components from a solvent or a dominant agent can be critical to identifying the source of the material. In the case of GC/MS, identifying minor components can be quite a challenge when the data are dominated by solvents or the major components, which may even saturate the column and/or detector. In this work, we will present a technique for performing a rapid analysis of GC/MS data containing large, interfering solvent peaks.

Our method for extracting information on minor components while reducing the impact of dominant components is two-pronged: first, we ignore the mass-spectral information originating from dominant species using a technique we call selected ion depletion (SID); and second, we

use factor analysis methods that take special advantage of the character of GC/MS data. These methods are fast, reasonably straightforward and easy to apply.

This paper is organized so as to walk a prospective user through a data analysis. We will initially discuss our method and strategy for SID and/or saturated peak depletion, followed by additional data preprocessing and factor analysis, finally we will present results from data acquired on commercially obtained standard mixtures.

THEORY

SELECTED ION DEPLETION

Selected ion monitoring is a technique that is implemented in the data acquisition phase of an experiment. It is accomplished by tuning the mass spectrometer to a suitable mass-charge (*m/z*) value or set of values and acquiring the current(s) as counts as a function of elution time.[2] In this way, the mass spectrometer becomes a very effective mass filter with high SNR at the expense of mass spectral range. By contrast, in our technique the mass spectrometer is scanned continually over an applicably wide *m/z* range sufficient to capture all of the fragments that one could reasonably expect in the sample. This may be seen as a more traditional, if brute force, method of obtaining data, where a larger range of *m/z* is collected at the expense of SNR. The advantage of collecting the full mass spectral range is that one has access to all mass fragments, which may include information on unexpected chemical species.

Often, chromatographic data contain peaks from dominant species or solvent that may be saturated either in the chromatographic domain or the mass spectral domain. Saturation, or overloading, in the chromatographic domain is frequently manifested as a distorted, non-Gaussian shaped peak with excessive tailing.[11] Saturation in the mass spectral domain can be

recognized as distorted mass spectra across the chromatographic peak, with the most populated mass channels saturating, resulting in reduced relative ratios to unsaturated peaks. These effects can make factor analysis more difficult by obscuring small peaks and increasing the complexity of the data. As intensity of the saturated peaks no longer increases with concentration, they are highly non-linear and thus linear factor analysis methods fail to produce easily interpretable results. Our solution for this type of problem is to remove the *m/z* channel data for major components and solvents via selected ion depletion (SID).

FACTOR ANALYSIS

Data Scaling

Prior to performing factor analysis, we must appropriately scale our GC/MS data so that it approximates the assumptions of the factor analysis technique. The factor analysis techniques we will use are based on the method of least squares, which assumes that the errors are independently and identically distributed (*i.i.d.*) normal. Since these GC/MS data are collected as counts from a quadrupole mass spectrometer, the first principles assumption is that the data are actually Poisson distributed.[12-14] Our group has extensive experience with optimal scaling of multivariate Poisson-distributed data.[15-17] Briefly, we can scale the data using the inverse of the square root of the mean mass spectrum.

Consider some GC/MS data in the $m \times n$ matrix **D** oriented as mass spectral domain by chromatographic domain with mean *m/z* spectrum **h** given by

$$\bar{\mathbf{d}}_m = \frac{1}{n}\mathbf{D}\mathbf{1}_n \tag{1}$$

where **1** is an *n*-vector column of ones. We can now scale the data in **D** using the diagonal matrix **H** whose diagonal elements are $\bar{\mathbf{d}}_m^{-1/2}$ using

$$\tilde{\mathbf{D}} = \mathbf{HD} \tag{2}$$

where $\tilde{\mathbf{D}}$ is the data scaled for Poisson statistics. This scaling decreases the effect of large variations in the data due solely to noise in intense spectral regions. It is important for subsequent factor analysis as it effectively down-weights the effect of variance due to noise in intense spectral features and concomitantly up-weights minor spectral features, which in the raw data are smaller in magnitude than noise elsewhere.

Principal component analysis (PCA)

PCA is a statistical method that decomposes a matrix into two sets of orthogonal of basis vectors, ordered by decreasing variance, that model the row and column spaces of the matrix.[18-19] It is often used as an initial data reduction method, whose subspace representation may be readily factor-analyzed by additional statistical treatments.[15] PCA can be represented in matrix form as

$$\tilde{\mathbf{D}} = \tilde{\mathbf{T}}\mathbf{P}^{\mathrm{T}} + \tilde{\mathbf{E}} \tag{3}$$

where $\tilde{\mathbf{T}}$ is an $m \times p$ matrix which describes the row (or mass spectral) space of the scaled data in $\tilde{\mathbf{D}}$, $\mathbf{P}$ is an $n \times p$ matrix describing the column (or chromatographic) space of $\tilde{\mathbf{D}}$, $\tilde{\mathbf{E}}$ is an $m \times n$ matrix of scaled residuals or noise, and the superscript "T" indicates the transpose of the preceding matrix or vector. We use p to define the size, or pseudorank, of the subspace model that describes the chemically meaningful information contained in $\tilde{\mathbf{D}}$, simply put the number of distinguishable chemical species in $\tilde{\mathbf{D}}$. $\tilde{\mathbf{T}}$ is orthogonal and $\mathbf{P}$ is orthonormal having the properties:

$$\begin{aligned} \mathbf{\Lambda} &= \tilde{\mathbf{T}}^{\mathrm{T}}\tilde{\mathbf{T}} \\ \mathbf{I} &= \mathbf{P}^{\mathrm{T}}\mathbf{P} \end{aligned} \tag{4}$$

where $\mathbf{I}$ is a $p \times p$ identity matrix and $\mathbf{\Lambda}$ is a $p \times p$ diagonal matrix of eigenvalues ordered from largest to smallest. One can also compute a "full set" of eigenvalues of length $min(m, n)$ for fairly low computational cost. These eigenvalues can be used to estimate the pseudorank, p, in numerous ways.[18] Commonly, a semi-logarithmic plot of eigenvalue versus factor number is produced and the number of factors selected where a "knee" occurs in the plot.[20]

There is a variety of methods to compute the PCA, among these are nonlinear iterative partial least square (NIPALS),[21] eigenanalysis,[22-23] and singular value decomposition (SVD).[23-24] SVD is very convenient since it decomposes the matrix $\tilde{\mathbf{D}}$ as

$$\tilde{\mathbf{D}} = \tilde{\mathbf{U}}\mathbf{S}\mathbf{V}^{\mathrm{T}} + \tilde{\mathbf{E}} \tag{5}$$

where $\mathbf{U}$ and $\mathbf{V}$ are, respectively, the $m \times p$ and $n \times p$ matrices of orthogonal left and right singular vectors and $\mathbf{S}$ is the diagonal matrix of singular values. The singular values are the square roots of the eigenvalues, *viz.*

$$\mathbf{\Lambda} = \mathbf{S}^2 \tag{6}$$

We can combine equations (3) and (5) to show that

$$\begin{aligned} \tilde{\mathbf{T}} &= \tilde{\mathbf{U}}\mathbf{S} \\ \mathbf{P} &= \mathbf{V} \end{aligned} \tag{7}$$

We have been careful in this section to utilize notation that indicates which data domain bears the scaling, in this case the mass-spectral domain. This is important since after factor analysis we will want to return those factors to their native scale, specifically

$$\mathbf{T} = \mathbf{H}^{-1}\tilde{\mathbf{T}} \tag{8}$$

Orthogonal Factor Rotation

After performing PCA, one will have an orthogonal, rank-*p* representation of the data that will probably not resemble any meaningful information to the chromatographer. Consequently, additional factor analysis is required to transform the PCA factors into interpretable factors. A reasonable approach to transforming the factors is to use a factor rotation method.

Factor rotation methods seek to maximize (or minimize) some criterion that is consistent with the nature of the data. In the case of GC/MS data, the varimax rotation[25-27] is an appropriate criterion for the chromatographic domain. The varimax criterion seeks an orthogonal rotation matrix, **R**, which maximizes the row (or time) variance of the orthonormal matrix **P**, thereby maximizing the "simplicity" of the rotated elution-time or chromatographic factors. In matrix form we have

$$\tilde{\mathbf{D}} = \tilde{\mathbf{T}}\mathbf{P}^{\mathrm{T}} = \tilde{\mathbf{T}}\mathbf{R}\mathbf{R}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}} = \underline{\tilde{\mathbf{T}}}\underline{\mathbf{P}}^{\mathrm{T}} \qquad (9)$$

Generally, the factors of chromatographic domain are simple or sparse. For example, when a species elutes it produces a peak in the chromatogram, generating a chromatographic factor with a single peak, and zeros or noise at all other times. Unless another species co-elutes, all other factors will be zero-valued (noisy) at the elution times encompassing that species peak. So, the chromatographic domain is, generally, sparse. Overlapping due to co-elution or the presence of a large background is a violation of this premise. By contrast, there is no reason to expect that the mass-spectral domain is sparse since many different compounds generate the same mass fragments, although not in the same *pattern*; so this violates the simplicity assumption.

Multivariate Curve Resolution (MCR)

If all species were to elute at different times such that none were overlapped, and no background arose from column packing loss and discharge, then PCA and varimax rotation would be sufficient to produce interpretable results. Unfortunately, we often have to deal with these and other problems, so we need to find a method applicable to these issues. MCR, also called linear unmixing,[27] is a factor analysis method that utilizes an alternating least squares strategy while employing constraints;[28-30] the most common constraint used being nonnegativity.[31-32] MCR seeks to solve

$$\tilde{\mathbf{D}} = \tilde{\mathbf{M}}\mathbf{C}^{\mathrm{T}} + \tilde{\mathbf{E}} \tag{10}$$

where $\tilde{\mathbf{M}}$ is the nonnegative $m \times p$ matrix modeling the row (or mass spectral) space of the scaled data in $\tilde{\mathbf{D}}$, $\mathbf{C}$ is the nonnegative $n \times p$ matrix modeling the column (or chromatographic) space of $\tilde{\mathbf{D}}$. Combining equations (10) and (9), we can form the relationship

$$\underline{\tilde{\mathbf{T}}}\underline{\mathbf{P}}^{\mathrm{T}} = \tilde{\mathbf{M}}\mathbf{C}^{\mathrm{T}} \tag{11}$$

which represents the dimension reduction of $\tilde{\mathbf{D}}$ as well as the imposition of nonnegativity constraints on the rotated PCA factors. Finally, following MCR, we would rescale the factors in $\tilde{\mathbf{M}}$ using the appropriate substitution into equation (8).

EXPERIMENTAL

In order to test the data analysis strategy described above, two known mixtures were analyzed. Commercially available volatile organic compound (VOC) mixtures, VOC Mix 6, and VOC Mix 8 (Supelco Inc., Bellefonte, PA) were analyzed as received. VOC Mix 6 (catalog number 4-8799), contains six compounds: dichlorofluoromethane, chloromethane, vinyl chloride, bromomethane, ethyl chloride, and trichlorofluoromethane. VOC Mix 8 (catalog number 4-

8803), contained 11 compounds: trans-1,2-dichloroethylene, cis-1,2-dichloroethylene, 1,2-dichloropropane, toluene, tetra-(per)chloroethylene (PCE), chlorobenzene, ethylbenzene, p-xylene, styrene, o-xylene, 1,2-dichlorobenzene. These mixtures were analyzed using an Agilent 6890 gas chromatograph/5975N mass spectrometer equipped with a J&W DB-5ms, 60m x 0.32 mm x 1.0μm film thickness, gas chromatography column. The analytical conditions for the mixture were modified for each mixture. The GC conditions used for VOC Mix 6 were isothermal at 35°C for 7 minutes, helium carrier gas flow rate of 1.2 mL/min, splitless injection of 0.2 μL of solution into 250°C injector. The mass spectrometer was operated in the full scan mode from *m/z*: 27-120. The GC conditions used for the VOC Mix 8 were isothermal at 35°C for 2 minutes, then ramped at 10°C/min to 200°C and held for 2 minutes, then 20°C/min to 290°C and held for 15 minutes. The helium carrier gas was kept at a constant flow rate of 1.4 mL/min. A splitless injection of 0.2 μL of solution into 250°C injector was used to introduce the mixture. The mass spectrometer was operated in the full scan mode from *m/z*: 27-120 to the 7 minute mark, then *m/z*: 33-320 through 18 minutes, then finally to *m/z*: 33-380 through the end of the analytical run.

Data were exported as netCDF files using Agilent ChemStation software version D.01.02.16 and subsequently imported into MATLAB® version 7.8.0.347 (R2009a)[33] using a freely available routine[34] modified by one of us to rapidly read these GC/MS data. Mass data were collected in increments of 0.1 amu, and upon importation binned to the nearest unit amu. Computations were performed using MATLAB® running under Microsoft Windows XP on a Dell Precision 690 computer, with two, dual-core, 3.2 GHz Intel® Xeon processors and 4.0 Gbyte RAM. PCA was performed using a freely available SVD algorithm written as MATLAB® m-file.[35] Analysis times were under ten seconds.

RESULTS AND DISCUSSION

VOC Mix 6

The VOC Mix 6 data were collected over a time range of approximately six minutes constituting 4287 time elements. The mass spectral domain covered *m/z* 27-119, whose non-zero entries consisted of 65 *m/z* elements. Of course, an all-zero column of mass spectral values is "depleted" and not analyzed since it contains no information. The top plot in Figure 1 is the total ion chromatogram of the raw data for VOC Mix 6; which is dominated by the methanol peak. The elution time range was truncated to a range of approximately 3-5.5 minutes to focus on the region of interest. In addition, the range *m/z* 28-33 represents the SID region for the methanol peak.[36] After this modification, the data size was 741 time elements by 56 *m/z* elements. The bottom plot in Figure 1 contains expanded-scale total ion chromatogram of the raw data as well as the SID plot with the methanol *m/z* eliminated. Most obvious here is complete loss of the methanol peak and its long tail, but an important consequence is the appearance of a peak (#4) that previously was concealed completely. Elimination of the methanol contribution is important for several reasons. Due to saturation effects, it makes the data much more complex than simply the addition of another chemical species. Its long tail complicates the analysis of overlapped peaks, since this behaves like a non-zero baseline, which prevents complete unmixing with PCA and presents special challenges for MCR.

Figure 2 is a scree plot of the eigenvalues of Poisson-scaled data VOC Mix 6. The plot shows nine or 10 significant PCA factors indicating nine or 10 possible analytes in the chromatogram. While the mixture purports to contain six species, the mixture and/or the solvent may be contaminated, or the column could have some carryover. We conducted analyses of both ranks nine and 10 on the data. The results of the rank-nine model, constituting over 99% of the total

variance, are presented below. The rank 10 analysis differs little from that of rank 9, with the additional component giving insignificant interpretive value.

The chromatographic factors for the varimax-rotated PCA analysis are presented in the top plot of Figure 3. While it is difficult to see, the extent of negativity in the factors is very small, even in the presence of overlap in the second and third peaks with centers at 3.67 and 3.71 minutes. For ease of viewing, we have expanded the scale and superpositioned the region describe here in the plot. Given the excellent performance of the algorithm, we required minimal "polishing" with MCR. The bottom plot of Figure 3 contains the MCR factors of the nine-factor model initialized using the varimax solution and completing only 10 iterations. Again, the expansion of the overlap region demonstrates the ability of MCR with nonnegativity constraints to "unmix" these factors. The MCR mass spectral factors for the nine components are displayed in Figure 4.

Identities of the six expected chemical species; with peak labels 1, 3, 5, 6, 7, and 8 in the Figures 1, 2 and 4 and Table 1; were confirmed using the mass spectral factors as inputs to the NIST mass spectral library search routine provided with the Agilent GC/MS software. In addition, the NIST routine identified peak 2 as dimethyl ether, a common contaminant of methanol; and peak 9 as acetonitrile, with a confirmatory match factor of >860. While the source of acetonitrile is mysterious, we did confirm its presence by comparing its elution time to that of an independent experiment involving only acetonitrile in methanol. The identity of the final species eluting as peak 4, was not revealed using the NIST mass spectral library search routine. This lack of library match is not unexpected because this factor is believed to represent ion-molecule reactions of methanol in the mass spectrometer source and thus not a separate distinct compound. The most abundant ion observed in the factors of peak 4 (see Figure 4) is m/z 34. For methanol there is a small (0.21%) isotopic contribution at m/z 34, likely which also has abundance from

protonated methanol. This factor demonstrates the sensitivity of the method to very low abundance chromatographic peaks.

VOC Mix 8

The original VOC Mix 8 data were collected over a time range of approximately 34 minutes constituting nearly 12,000 elements, and a mass range *m/z*: 27-371 comprising 240 nonzero mass channels. Initial data trimming reduced the time domain to 3-24 minutes (7554 elements) and the mass range *m/z*: 27-355 (148 mass channels), including SID of methanol *m/z* (28-33). The top plot in Figure 5 contains expanded-scale total ion chromatogram of the raw data as well as the SID plot with the methanol *m/z* eliminated. As in VOC Mix 6, complete loss of the methanol peak and its long tail are dramatic. The twelve major peaks have been numbered for convenience, and their expected identities, as per NIST MS library matching, are contained in Table 2. The table lists only the "best hit", which misassigned the dichloroethylene, xylene, and dichlorobenzene isomers. This is common for these types of isomers in many other types of data processing; and then GC/MS practitioners will employ retention time information or Kovats indices for confirmation. Browsing the NIST library matching results, the correct assignment in each case is within 1% of the "best hit" match score, and thus these misassignments are of minor concern. The important point is that the MCR is correctly modeling and separating the peaks. In addition to the eleven known constituents of the mix, the first peak is that of dimethoxymethane. Figure 6 is a scree plot of the eigenvalues of the Poisson-scaled data after trimming the time domain and removal of the methanol *m/z* channels. Unfortunately, in this plot there is not a distinct or obvious transition from data eigenvalues to noise eigenvalues. Upon expansion of the plot, we estimated that a 25-factor model would be appropriate.

The bottom plot in Figure 5 contains the rank-25 model, chromatographic factors from MCR, initiated using the varimax-rotated PCA solution. While it is difficult to make out all of the factors, we were able to model all of the major peaks, plus a significant number of minor peaks. For reasons which will become obvious, here we are using a lettering scheme for the *thirteen* major MCR factor peaks that model the *twelve* major data elution peaks. Species whose mass spectra are highly correlated pose a special problem; the results plotted in Figure 7 represent one such challenge.

The species giving rise to the peaks displayed in Figure 7 are, from left to right, ethylbenzene (8), p-xylene (9), styrene (10), and o-xylene (11). The first two peaks at times 15.3 and 15.5 minutes are well resolved, but have very similar, but discernable, mass spectra; so they are separable with PCA and MCR. The second two peaks are highly overlapped, but have very different mass spectra, so they are mostly separable with varimax-PCA with significant negative-going lobe so as to be orthogonal. The problem arises because the mass spectra of ethylbenzene, p-xylene, and o-xylene are so similar that the o-xylene peak is best modeled as a linear combination of ethylbenzene and p-xylene. We conducted a separate analysis of the data for all times greater than 15.8 minutes and were able to obtain a distinct peak for styrene at 15.99 minutes and another distinct peak for o-xylene at 16.02 minutes. While having to partition the data matrix in this manner not the most desirable situation, it does represent an option for identifying isomeric species whose mass spectra are nearly indistinguishable.

Perhaps the most impressive demonstration of the power of the techniques employed here is represented in Figure 8. Here in the top plot, we see in red the elution peak for PCE its *two* varimax-rotated PCA factors in blue and green. Due to orthogonality constraints the peaks have both positive and negative lobes, but more fundamentally one would ask, "Why two peaks?"

Even in the bottom plot, the MCR results nicely resolve the two peaks with Gaussian shapes. The answer to our question resides in Figure 9 which contains the peak-integrated mass spectrum, and the mass spectral factors of the first factor (left) and second factor (right) in the top, middle, and bottom panels, respectively. These plots reveal a chlorine isotope abundance dependence as a function of elution time. An analysis of these plots is found in Tables 3-5. The theoretical probability for the two, three and four chlorine fragments using a binomial probability distribution and a 24.22% natural abundance for ^{37}Cl.[37]

Table 3 is a comparison of the theoretical and actual proportions of C_2Cl_2 fragment for the possible combinations of ^{35}Cl and ^{37}Cl. A comparison of the theoretical proportions with that of the integrated peak of the raw data shows excellent agreement, as well as with the NIST Standard Data.[36] However, the first (left) peak appears to have enrichment in the lighter isotope since its 0/2 $^{37}Cl/^{35}Cl$ proportion is larger than the expected value while the 2/0 $^{37}Cl/^{35}Cl$ proportion is lower than expected. In contrast, the second (right) peak values show enrichment in the heavier isotope. This pattern is repeated in the subsequent tables: enrichment of ^{35}Cl in the left peak and complementary enrichment of ^{37}Cl in the right peak. This is entirely consistent with isotopic behavior in a diffusion controlled process.

CONCLUSIONS

We have presented a very powerful technique for analyzing GC/MS data using multivariate analysis. The method is based on established data analysis techniques that may be easily applied by most chromatographers using a small laboratory computer. We have demonstrated the method using some simple GC/MS data with excellent results. The ability of the analysis to extract very subtle details from the data is highlighted by the separation of heavily overlapped peaks due to an isotopic chlorine effect.

ACKNOWLEDGEMENTS

Funding for this project was provided in part by the Science and Technology Directorate of the U.S. Department of Homeland Security under a contract with Sandia National Laboratories. This work was supported by Sandia National Laboratories and the Department of Energy. Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-ACO4-94AL85000.

REFERENCES

(1) Cudiamat, G., GC–MS, In *E-Separation Solutions*; Advanstar Communications, Inc.: Woodland Hills, CA, 2009 http://chromatographyonline.findanalytichem.com/lcgc/Research/GCndashMS/ArticleStandard/Article/detail/623723, (retrieved 2 Nov 2009).

(2) Skoog, D. A.; Leary, J. J. *Principles of Instrumental Analysis*; Saunders College Publishing: Fort Worth, 1992.

(3) Herbert, C. G.; Johnstone, R. A. W. *Mass Spectrometry Basics*; CRC Press: Boca Raton, FL, 2003.

(4) Biller, J. E.; Biemann, K. *Analytical Letters* **1974**, *7*, 515 - 528.

(5) Knorr, F. J.; Thorsheim, H. R.; Harris, J. M. *Anal. Chem.* **1981**, *53*, 821-825.

(6) Dromey, R. G.; Stefik, M. J.; Rindfleisch, T. C.; Duffield, A. M. *Anal. Chem.* **1976**, *48*, 1368-1375.

(7) Colby, B. N. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 558-562.

(8) Stein, S. E. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770-781.

(9) Mallard, W. G.; Reed, J., "Automated Mass Spectral Deconvolution & Identification System AMDIS — USER GUIDE"; National Institute of Standards and Technology (NIST): Gaithersburg, MD, 1997.

(10) Karjalainen, E. J.; Leinonen, A.; Karjalainen, U. P. In *Advances In Mass Spectrometry,* Vol. 14; Karjalainen, E. J., Hesso, A. E., Jalonen, J. E., Karjalainen, U. P., Eds.; Elsevier: Amsterdam, The Netherlands, 1998; pp 595-609.

(11) Poole, C. F.; Poole, S. K. *Chromatography Today*; Elsevier Science: Amsterdam, The Netherlands, 1991.

(12) Hackett, C.; Hermon, H.; Cross, E.; Doty, P.; Tarver, E.; James, R. *J. Electron. Mater.* **1999**, *28*, 774-784.

(13) Begley, I. S.; Sharp, B. L. *J. Anal. At. Spectrom.* **1997**, *12*, 395-402.

(14) El Abbassi, B.; Mestdagh, H.; Rolando, C. *Int. J. Mass Spectrom. Ion Processes* **1995**, *141*, 171-186.

(15) Keenan, M. R. In *Techniques and Applications of Hyperspectral Image Analysis,* Vol.; Grahn, H. F., Geladi, P., Eds.; John Wiley & Sons, Ltd: Chichester, West Sussex, England, 2007; pp 89-126.

(16) Keenan, M. R.; Kotula, P. G. *Appl. Surf. Sci.* **2004**, *231/232*, 240-244.

(17) Keenan, M. R.; Kotula, P. G. *Surf. Interface Anal.* **2004**, *36*, 203-212.

(18) Malinowski, E. R. *Factor Analysis in Chemistry*; John Wiley and Sons, Inc.: New York, 1991.

(19) Keenan, M. R. *J. Vac. Sci. Technol. A* **2004**, *23*, 746-750.

(20) Braun, M. L. Dissertation, University of Bonn, Bonn, 2005.

(21) Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta* **1986**, *185*, 1-17.
(22) Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*; : Prentice Hall: Englewood Cliffs, N.J., 1992.
(23) Golub, G. H.; Van Loan, C. F. *Matrix Computations*, 3rd ed.; Johns Hopkins University Press: Baltimore, 1996.
(24) Larsen, R. M., "Lanczos bidiagonalization with partial reorthogonalization"; Report Number DAIMI PB-357; Department of Computer Science - Daimi, Aarhus University: Aarhus, Denmark, 1998.
(25) Rummel, R. J. *Applied Factor Analysis*; Northwestern University Press: Evanston, IL, 1970.
(26) Harman, H. H. *Modern Factor Analysis*; University of Chicago Press: Chicago, 1976.
(27) Keenan, M. R.; Timlin, J. A.; Van Benthem, M. H.; Haaland, D. M. *Proceedings of SPIE*, Seattle, WA, USA, July 8-10, 2002 2002; SPIE - The International Society for Optical Engineering; 193-202.
(28) Tauler, R.; Casassas, E.; Izquierdo-Ridorsa, A. *Anal. Chim. Acta* **1991**, *248*, 447-458.
(29) Van Benthem, M. H.; Keenan, M. R.; Haaland, D. M. *J. Chemom.* **2002**, *16*, 613-622.
(30) Kotula, P. G.; Keenan, M. R.; Michael, J. R. *Microsc Microanal* **2003**, *9*, 1-17.
(31) Bro, R.; DeJong, S. *J. Chemom.* **1997**, *11*, 393-401.
(32) Van Benthem, M. H.; Keenan, M. R. *J. Chemom.* **2004**, *18*, 441-450.
(33) Ver. 7.8.0.347 (R2009a); The MathWorks, Inc.: Natick, MA, 2009.
(34) Spencer, P.; http://www.mathworks.com/matlabcentral/fileexchange/15177: Natick, MA, 2007.
(35) Larsen, R. M., Ver. 1.1; http://sun.stanford.edu/~rmunk/PROPACK/: Stanford, CA, 2004.
(36) Stein, S. E., NIST Mass Spectrometry Data Center, In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg MD, 20899, 2008, http://webbook.nist.gov, (retrieved 24 Oct 2009).
(37) Lide, D. R., Ed. *CRC Handbook of Chemistry and Physics*, 84 ed.; CRC Press: Boca Raton, Florida, 2003.

FIGURES

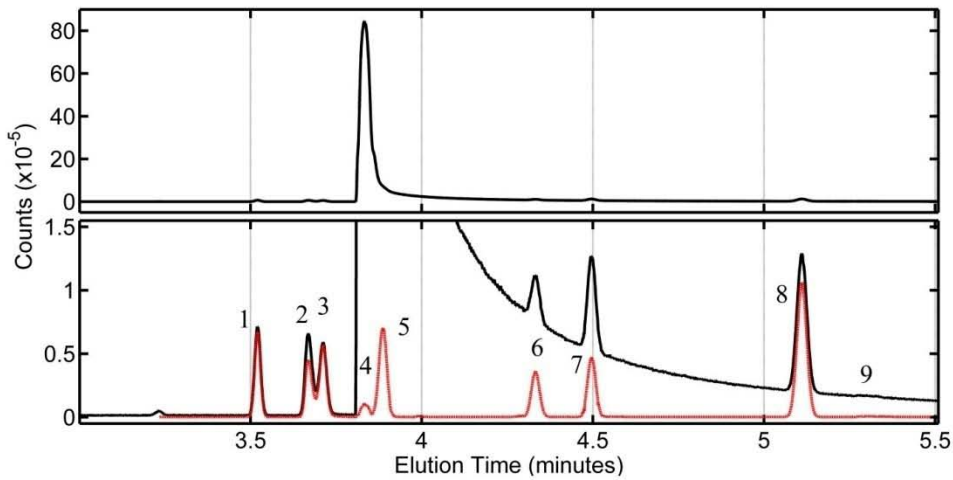

Figure 1. Top plot displays VOC Mix 6 total ion chromatogram prior to data pretreatment shows the dominance of the methanol solvent peak. Bottom plot is the comparison of the total ion chromatograms of the raw VOC Mix 6 GC/MS data (black) and of the same data after SID removing *m/z* 28-33 (red).

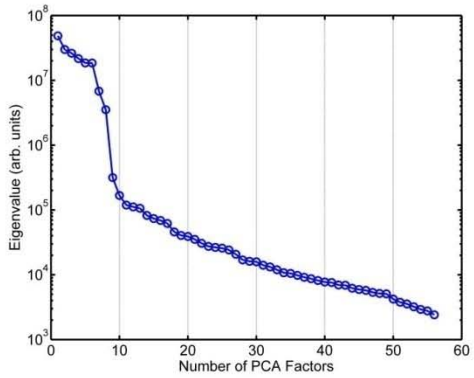

Figure 2. Eigenvalues of Poisson-scaled GC/MS VOC Mix 6 data plotted versus PCA factor number.

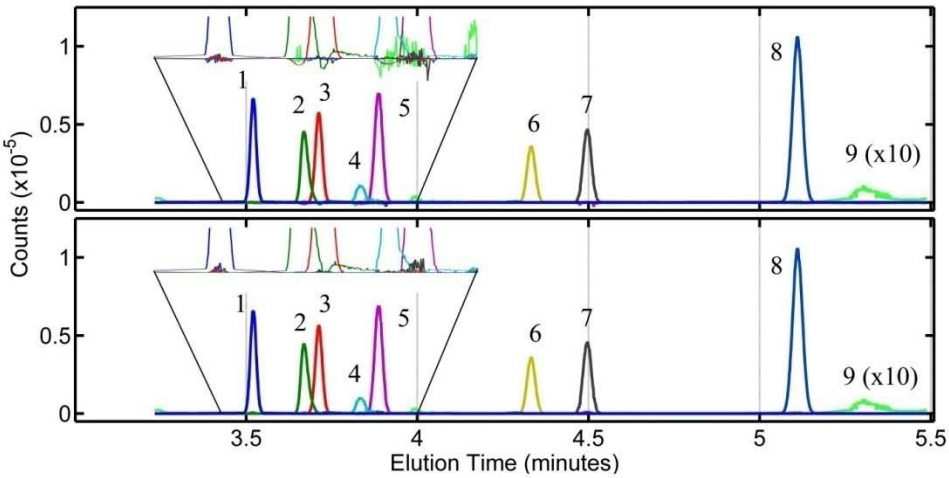

Figure 3. Top plot contains the eigenvectors for varimax-rotated PCA on the Poisson-scaled SID data. The plots are total ion chromatograms of the reconstructed individual components; which are essentially the scaled chromatographic factors. The inset shows the negative lobes of the overlapped peaks. Bottom plot contains results of ten iterations of MCR with nonnegativity applied in both domains and initialized using the varimax-rotated PCA factors. Peaks labels correspond to peak numbers in Table 1. The amplitude of peak 9 has been increased by a factor of 10 to improve visibility.

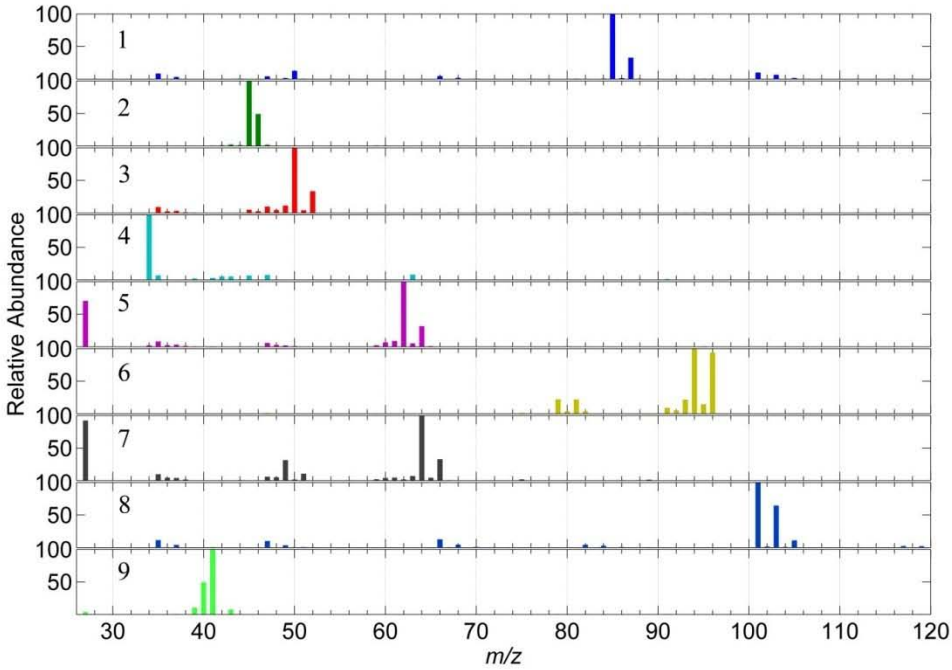

Figure 4. Mass spectral factors for 10 iterations of MCR initialized with varimax-rotated, nine-factor PCA model. MCR employed nonnegativity in both domains. Factors are color-coded to correspond to factors in the bottom plot of Figure 3.

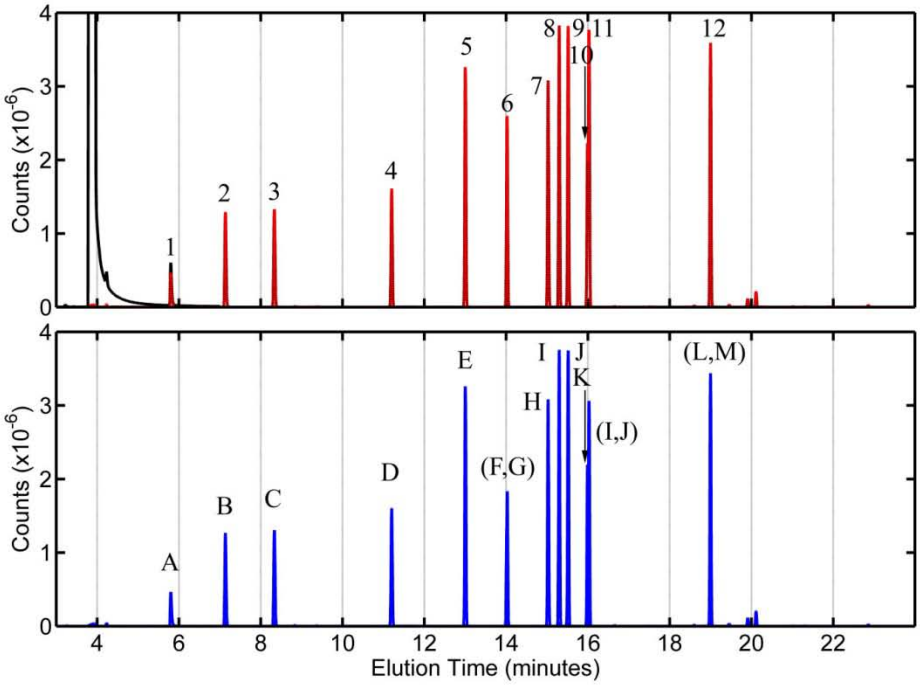

Figure 5. Top plot displays the comparison of the total ion chromatograms of the raw VOC Mix 8 GC/MS data (black) and of the same data after SID removing *m/z* 28-33 (red). Bottom plot contains results of ten iterations of MCR with nonnegativity applied in both domains and initialized using the 25-factor model, varimax -rotated PCA -factors. See text for species in mixture.

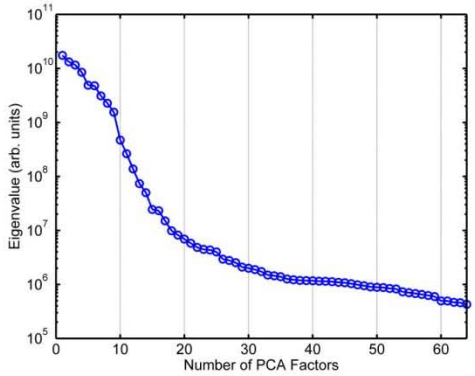

Figure 6. Eigenvalues of Poisson-scaled GC/MS VOC Mix 8 data plotted versus PCA factor number.

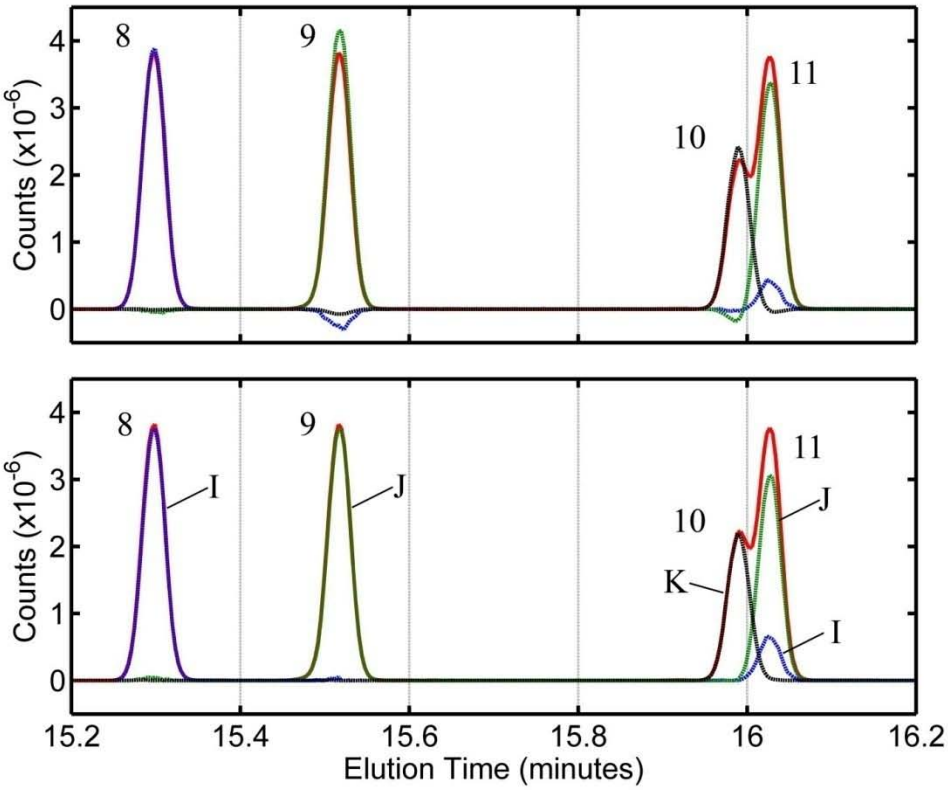

Figure 7. Top plot contains the chromatographic factors from the varimax-rotated PCA of the Poisson-scaled SID data for VOC Mix 8 over the elution time period 15.2-16.2 minutes. The red plot is the total ion chromatogram of the SID data, the blue, green, and black plots are scaled PCA chromatographic factors. Bottom plot is the chromatographic factors for 10 iterations of MCR initialized with varimax-rotated, 25-factor PCA model. MCR employed nonnegativity in both domains. The peaks from left to right are ethylbenzene (8), p-xylene (9), styrene (10) and o-xylene (11). Ethylbenzene, p-xylene and styrene are modeled by MCR factors I, J and K, respectively; while o-xylene is modeled as contributions from both factors I and J.

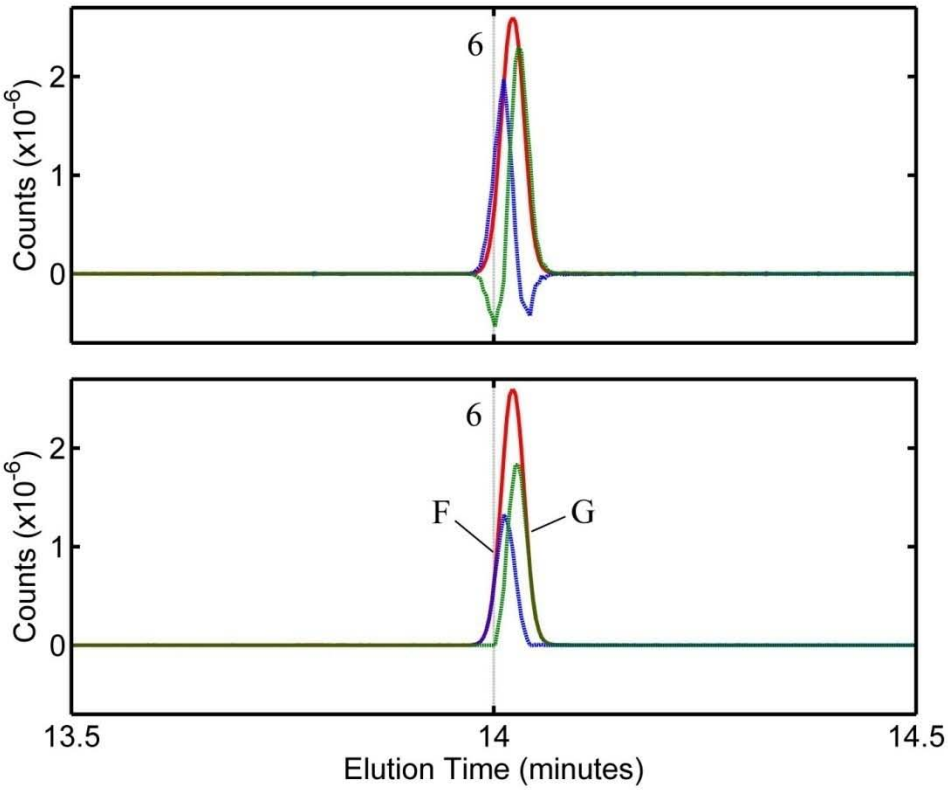

Figure 8 Top plot contains peak 6, PCE, and the chromatographic factors from the varimax-rotated PCA of the Poisson-scaled SID data for VOC Mix 8. The red plot is the total ion chromatogram of the SID data, the blue and green plots are scaled PCA chromatographic factors. Bottom plot is the chromatographic factors for 10 iterations of MCR initialized with varimax-rotated, 25-factor PCA model. MCR employed nonnegativity in both domains. Factor F is the lighter chlorine isotope PCE and G is the heavier isotope.

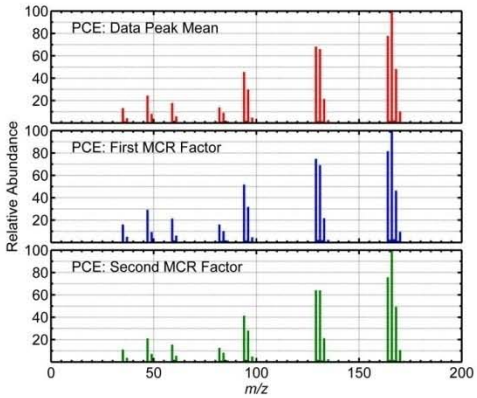

Figure 9. Mass spectra of average PCE peak (6) and as estimated from MCR. Top: Mean mass spectrum of SID data for the elution times 13.5-14.5 minutes. Middle: Mass spectral MCR factor for the early eluting PCE factor. Bottom: Mass spectral MCR factor for the later eluting PCE factor.

TABLES

Table 1. Chemical species assignments made from mass spectral factors derived from MCR analysis of VOC Mix 6 data. Notes: †Dimethyl ether believed to be a contaminant from methanol solvent. *Contaminant was unidentified. **Acetonitrile contaminant was from an unknown source, but may be column or system carryover.

Peak	Elution Time (min)	Chemical Compound (NIST Library Assigned)
1	3.52	Dichlorodifluoromethane
2	3.67	Dimethyl ether†
3	3.71	Chloromethane
4	3.83	Unknown*
5	3.89	Vinyl chloride
6	4.33	Bromomethane
7	4.50	Ethyl chloride
8	5.11	Trichloromonofluoromethane
9	5.31	Acetonitrile**

Table 2.

Peak	MCR Factor	Expected Chemical Compound	Elution Time (min)	Chemical Compound (NIST Library Assigned)
1	A	Dimethoxymethane	5.80	Dimethoxymethane
2	B	trans-1,2-Dichloroethylene	7.13	cis-1,2-Dichloroethylene
3	C	cis-1,2-Dichloroethylene	8.33	trans-1,2-Dichloroethylene

4	D	1,2-Dichloropropane	11.20	1,2-Dichloropropane
5	E	Toluene	13.00	Toluene
6	F	PCE	14.01	PCE
6	G	PCE	14.03	PCE
7	H	Chlorobenzene	15.03	Chlorobenzene
8	I	Ethylbenzene	15.30	Ethylbenzene
9	J	p-Xylene	15.52	m-Xylene
10	K	Styrene	15.99	Styrene
11	(I,J)	o-Xylene	16.02	
12	L	1,2-Dichlorobenzene	19.00	1,3-Dichlorobenzene
12	M	1,2-Dichlorobenzene	19.01	1,4-Dichlorobenzene

Table 3. Probabilities and actual proportions calculated for the two-chlorine fragments of PCE.

$^{37}Cl/^{35}Cl$ ratio:	0/2	1/2	2/0
Mass C_2Cl_2	94	96	98
Theoretical Prob. (%):	57.4	36.7	5.9
Raw Data Dist. (%):	56.9	37.1	6.0
Peak One Dist. (%):	58.8	36.1	5.1
Peak Two Dist. (%):	55.8	37.8	6.4
NIST data Dist. (%):	56.7	37.2	6.1

Table 4. Probabilities and actual proportions calculated for the three-chlorine fragments of PCE.

$^{37}Cl/^{35}Cl$ ratio:	0/3	2/1	2/1	3/0
Mass C_2Cl_3	129	131	133	135
Theoretical Prob. (%):	43.5	41.7	13.3	1.4
Raw Data Dist. (%):	43.2	41.7	13.5	1.5
Peak One Dist. (%):	44.6	41.2	12.8	1.3
Peak Two Dist. (%):	42.3	42.2	13.9	1.5
NIST data Dist. (%):	42.6	42.4	13.6	1.4

Table 5. Probabilities and actual proportions calculated for cationic PCE.

$^{37}Cl/^{35}Cl$ ratio:	0/4	1/3	2/2	3/1	4/0
Mass C_2Cl_4	166	168	170	170	172
Theoretical Prob. (%):	33.0	42.2	20.2	4.3	0.3
Raw Data Dist. (%):	32.8	42.2	20.3	4.3	0.3
Peak One Dist. (%):	34.3	42.0	19.4	4.0	0.3
Peak Two Dist. (%):	32.0	42.3	20.9	4.5	0.4
NIST Data Dist. (%):	33.3	41.2	20.7	4.5	0.3

DISTRIBUTION:

Electronic copy to:

1 Department of Homeland Security
Attn: S&T Directorate
TCA Branch
Dr. Robert L Bull VTA-8222
(202) 254 5344
Washington DC 20528

1	MS0886	Mark Van Benthem	01822
1	MS 0886	Theodore T. Borek, III	01822
1	MS 0886	Curtis Mowry	01822
1	MS 0886	Paul Kotula	01822
1	MS 0886	James Aubert	01822
1	MS 0826	Veronica Lopez	08523
1	MS 0825	Richard O. Griffith	01536
1	MS 9004	Duane L. Lindner	08120
1	MS0899	Technical Library	09532