

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Application of Chemometrics and Fast GC-MS Analysis for the Identification of Ignitable Liquids in Fire Debris Samples

Author(s): Michael Sigman, Ph.D., Mary Williams, M.S.

Document No.: 240684

Date Received: December 2012

Award Number: 2008-DN-BX-K069

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

**Application of Chemometrics and Fast GC-MS Analysis for the Identification of
Ignitable Liquids in Fire Debris Samples**

Award No. 2008-DN-BX-K069

Author/ Principal Investigator	Michael E. Sigman, Ph.D. Department of Chemistry and National Center for Forensic Science University of Central Florida PO Box 162367 Orlando, FL 32816-2367 407-823-6469
Co-Author	Mary Williams, M.S. National Center for Forensic Science University of Central Florida PO Box 162367 Orlando, FL 32816-2367

Abstract

The goal of the research conducted under this grant was to develop a chemometric method of data analysis that would facilitate the identification of GC-MS patterns associated with ignitable liquid classes, as designated under ASTM E 1618-10. The objective of the research was to develop a data analysis method that would classify ignitable liquid residue in the presence of background interferences found in fire debris. Pattern recognition and classification methods available at the onset of this research did not explicitly take into account background interference issues. A novel method was developed under this research to classify ignitable liquid residues into the ASTM classes, even in the presence of a strong background signal, without *a priori* knowledge of the background signature. The method makes use of target factor analysis (TFA) in combination with Bayesian decision theory. The use of Bayesian decision theory provides results in the form of posterior probabilities that a set of samples from a fire scene contain an ignitable liquid of a specific ASTM class. Error rates are not currently available for fire debris analysis, other than extrapolations from proficiency tests. The method was further refined by introducing a sensitivity parameter which made the method very conservative in its predictions, and gave a true “soft” classifier. Soft classifiers allow classification of a sample into multiple classes and afford the possibility of not assigning the sample to any of the available classes. In order to achieve the goals, this work was broken down into three tasks.

Task I of the research focused on the development of the combined TFA and Bayesian decision theory, and testing of the method using computationally constructed data sets prepared using total ion spectra (TIS) from ignitable liquids and pyrolyzed substrates from the Ignitable Liquids Reference Collection and Database and the Substrate Databases respectively, both of which are NCFS/TWGFX databases. Task II tested the method using laboratory generated burn samples. Task II of the research generated data sets by burning common building/furnishing materials with differing amounts of applied ignitable liquid and by varying both the amount of applied liquid and the relative amount of substrate materials. Task III further tested the method on large-scale burn samples generated specifically for the purpose of testing the method. The large-scale burns were conducted on Konex shipping containers that had been built out on the inside and furnished with commercially available household furnishings. Each container was burned with the aid of one or two flammable liquids. Following each burn, fire debris samples were collected from on and off the ignitable liquid pour trails, and the samples were analyzed at the National Center for Forensic Science, on the campus of The University of Central Florida. Task III burns were conducted at the Bureau of Fire Standards and Training, Ocala, FL, in collaboration with the Bureau of Forensic Fire and Explosives Analysis, Tallahassee, FL.

Results from Task I demonstrated that ignitable liquid contributions can be detected, even as a minor component, in the presence of undefined background contributions from pyrolysis of common building materials. Laboratory-scale burns conducted in Task II demonstrated an 80 – 85% success rate for classification when assigning the ASTM class based on the class having the largest posterior probability. Similar results were achieved for the large-scale burns conducted in Task III.

The 15% incorrect classifications include those samples where the ignitable liquid residue was heavily weathered and, in some cases, the liquid may have been completely evaporated. The

correct classification rate of 80 - 85% is nearly identical to the 85% correct classification rate previously reported by our research group for ignitable liquids, in the absence of any interference, based on TIS similarity. These results indicate that ignitable liquid classification can be achieved in the presence of *a priori* unknown interferences from pyrolysis. The methodology developed under this research grant is directly applicable to fire debris analysis and numerous other areas of forensic trace analysis where the analyte must be identified in the presence of an unspecified background signature. The broader applicability of the method also broadens the potential impact on forensic practice. The results from the research have been presented at numerous forensic science conferences, two Forensic Science M.S. theses have been written, a comprehensive manuscript on the work is under peer-review at Forensic Science International, and one patent application on the combined TFA-Bayesian method has been filed.

.

Contents

Executive Summary	5
The Problem:.....	5
Purpose of the Research:.....	6
Research Design:	6
Findings and Conclusions:.....	8
Implications for Policy and Practice:	9
I. Introduction:	9
Statement of the problem:.....	9
Literature review:.....	10
Hypothesis or rationale for the research:	14
II. Methods	15
Experimental Methods:.....	15
Statistical Methods:.....	19
III. Results.....	27
Simulations:	27
Laboratory-scale burns:	33
Large-Scale Burn Studies:	36
IV. Conclusions.....	38
Discussion of findings:	38
Implications for policy and practice:	39
Implications for further research:.....	40
V. References.....	40
VI. Dissemination of Research Findings	42
Publications and Thesis:	42
Patents:.....	42
Presentations:	43
VII. Appendix.....	44

Executive Summary

The Problem: It has been said that fire debris analysis is among the most difficult forensic challenges. This problem is one of identifying the chromatographic pattern produced by hundreds of organic analytes from an ignitable liquid in the presence of possibly hundreds of interfering compounds, some of which are also components of the ignitable liquid. Gasoline, for example, can contain several hundred organic hydrocarbons. The compounds that interfere with identifying the ignitable liquid pattern come primarily from pyrolysis of materials contained in the structure. The heat from a structure fire will cause the organic materials (woods, plastics, foams, etc.) to breakdown into comparatively simple hydrocarbons. The number and identities of hydrocarbons formed from different organic materials is highly varied, and in many cases they will be identical to the hydrocarbons comprising a given ignitable liquid. There may be many organic materials present in any structure, i.e., building materials and furnishings. The pattern recognition problem is further complicated by evaporative processes resulting from the heat of the fire, which also distort the ignitable liquids chromatographic profile.

The method currently used to identify an ignitable liquid chromatographic profile is primarily one of visual pattern recognition. When the ignitable liquid residue is present in much higher concentrations than the interfering compounds, this procedure works well and the fire debris analyst can call a sample positive for ignitable liquid residue with confidence. However, when the ratio of ignitable liquid residue to interfering compounds decreases, recognition of the ignitable liquid residue becomes more challenging and the use of extracted ion chromatographic profiles and target compound analysis take on greater importance. As the ratio of ignitable liquid to substrate continues to decrease, eventually the analyst must decide if they are comfortable calling a sample positive for ignitable liquid residue. The point at which the analyst goes from certain to unsure will vary with each sample and is subjective at best. Consequently, when fire debris evidence is introduced at trial, the fire debris analyst does so without an estimated or known error rate. Given the lack of known error rates, the fire debris analyst must be conservative in calling a sample positive for ignitable liquid, that is to say that the ignitable liquid pattern must be strong and clearly recognized. A mathematical approach to address this issue and assist the fire debris analyst in reaching a decision regarding case samples does not currently exist, due in large part to the complicated nature of the problem. The research reported here directly addresses this issue.

To be clear, the problem that the fire debris analyst faces is not an identification problem, it is a classification problem. The analyst is not trying to identify a specific ignitable liquid, (i.e., 83 octane gasoline from the station on the corner). The analyst is attempting to identify a pattern that is consistent with a “class” of ignitable liquids. The standard method followed by most analysts in the United States is the American Society of Testing and Materials, ASTM E1618, which defines the classes of commercial ignitable liquids. The ASTM classes include gasoline (Gas), petroleum distillates (PD), *iso*-paraffinic products (ISO), aromatic products (AR), naphthenic paraffinic products (NP), normal alkane products (NA) and oxygenated solvents (OXY). The classes may be subdivided based on carbon range as light ($C_4 - C_9$), medium ($C_8 - C_{13}$) and heavy ($C_8 - C_{20+}$). The PD class is frequently broken into the light, medium and heavy sub-classifications; LPD, MPD and HPD respectively. The OXY class is a collection of all ignitable liquids that contain a major oxygenated component (i.e., ketone, alcohol, ester, etc.).

The miscellaneous (MISC) category does not possess class characteristics and is simply a collection of those ignitable liquids that do not fit into the other defined classes.

The problem addressed in this research is the development of a chemometric method to aid the fire debris analyst in determining if an ignitable liquid residue is present in a fire debris sample by providing a statistical evaluation that will assist the analyst in reaching his/her decision. The method must be robust enough to work without *a priori* knowledge of the interferences' chromatographic signature. In a broader context, detection of an analyte or class of analytes in the presence of an unspecified interference is a very challenging analytical problem. This research also addresses that broader problem. Fortunately, in both the forensic-specific and broader problem, the number of analyte classes or the number of analytes of interest are significantly smaller in number than the number of possible backgrounds. It is this common feature of the problem that facilitates a solution.

Purpose of the Research: The purpose of this research is to devise a chemometric method that aids the fire debris analyst in detecting the presence of an ignitable liquid in the presence of a substantial and undefined background signature. *It is important to note that the decision to call a sample positive for ignitable liquid residue lies with the analyst and it is not the intent of this research to circumvent that responsibility or to replace the analysts decision with a "black box" algorithm.* The product from this research should aid the analyst in reaching the decision to call a sample positive or negative for ignitable liquid residue and to provide some statistical information regarding the ASTM class of ignitable liquid that may be present in fire debris determined to contain ignitable liquid residue.

Research Design: The goals of this research were addressed through the development of an appropriate theoretical approach and a series of three experimental tasks. Currently, chemometric multivariate approaches to pattern recognition do not explicitly address the problem of interferences. The most common approach is to use a controlled training procedure where the training data set includes analyte samples that are contaminated with interferences. This approach has been met with varying degrees of success. The problem with this approach is that it is impossible to include conceivable interference and combination of interferences.

In the research reported here, data from multiple samples from a fire scene are analyzed as a whole, rather than analyzing a single sample. The data from multiple samples are concatenated into a single data matrix for analysis. Chromatographic retention times vary from one laboratory to another and even within a single laboratory. Therefore, chromatographic retention times do not form a strong basis for pattern recognition. The average mass spectrum taken across the chromatographic profile (also referred to as the total ion spectrum or TIS) has been shown to form a strong basis for identification of complex mixtures, such as commercial ignitable liquids. The TIS from the GC-MS analyses of extracted ignitable liquid residues from fire debris are used in this research. The chemometric methodology developed for this work involves a combined approach of target factor analysis (TFA) and Bayesian decision theory. The data matrix comprised of multiple fire debris sample TIS is subjected to abstract factor analysis (mathematically equivalent to principal components analysis) and the dimensionality of the data is reduced to give a set of latent variables or principal factors. A set of target TIS are projected into the principal factor space and then back-transformed into the original data space. A direct

comparison is made between the target TIS and recovered TIS (i.e. by Pearson correlation). The target TIS are taken from the Ignitable Liquids Reference Collection and Database (ILRC) maintained by the National Center for Forensic Science. Each entry in the ILRC has been assigned to one of the classes defined by ASTM E1618-10. The correlations from the ILRC TIS that were projected into the principal factor space are divided into their assigned ASTM classes and a class-conditional kernel probability distribution is calculated for each class. The probability of obtaining a correlation, greater than a specified limit, is calculated for each class by integrating the area under each class-conditional probability curve over the region between the specified limit and one. Those probabilities are used to calculate the posterior probabilities by the standard Bayesian approach. In cases where the integrated area for a given class-conditional probability falls below a sensitivity limit α , typically 0.05, the posterior probability for that class is set to zero. If the integrated areas for all classes fall below α for all classes, the sample cannot be classified. The method can be made very conservative by adjusting α and the integration lower limit, and the method meets the requirements of a soft classifier.

The first experimental task was to test the method using computationally generated data sets. A series of test calculations were conducted on data sets containing 10 spectra. Each spectrum contained one or two ignitable liquids from different ASTM classes, a set of five or 10 substrate pyrolysis TIS and five percent random noise. A second test set was calculated with two contributing ignitable liquids from classes ISO and MPD and five substrate pyrolysis TIS, along with 5% random noise. In the final test, a single ignitable liquid class, Gas, was combined in contributions up to 70% with 10 substrate pyrolysis TIS. This test is significant because the number of components (ignitable liquid plus substrates) exceeds the number of spectra analyzed (i.e., the problem is under determined). The last test is also significant because in real samples, it is impossible to tell if the number of spectra analyzed exceeds the rank of the resulting data matrix. The target TIS for these tests were taken from the ILRC and limited to the three classes involved in the tests.

The second experimental task involved investigation of a series laboratory-scale burn samples. Each experiment was comprised of a set of samples prepared by heating a set of substrates and a single ignitable liquid in quart paint cans. The samples in each set contained varying volumes of the ignitable liquid and/or variable substrates. Passive headspace samples were collected from each burn sample and analyzed following a defined analytical protocol. The TIS from analysis of each set of samples were compiled and analyzed following the TFA-Bayesian method. The tests utilized all ASTM classes of ignitable liquids and the target TIS were taken from the ILRC. The results from the tests were evaluated to determine if the highest posterior probability corresponded to the class of ignitable liquid used in the burn.

The third experimental task was to conduct large-scale burns with known ignitable liquids, to take multiple samples from the post-burn scene and test the TFA-Bayesian method. The large scale burns were conducted at the Bureau of Fire Standards and Training, Ocala, FL, in collaboration with the Bureau of Forensic Fire and Explosives Analysis, Tallahassee, FL. Konex shipping containers were built out on the inside with wallboard and commercial flooring. The containers were furnished with commercially available household furnishings and then burned with the aid of one or two ignitable liquids (500 mL of each liquid). The fires were allowed to burn for 5 – 15 minutes before being extinguished with water. All of the containers went to

flashover and were allowed to cool for a period of several hours before samples were collected. Fire debris samples were taken along the pour trails and off the trails. Samples were returned to UCF for analysis by gas chromatography – mass spectrometry (GC-MS). Data from each container burn was combined and analyzed by the TFA-Bayesian method. A total of 12 container burns were performed in three sets of four burns.

Findings and Conclusions: The TFA and Bayesian theories used in this research are carefully described in this report. The theoretical approach in this work differs from those previously described. Because of this difference and the need for forensic scientists to fully understand the method, a full derivation of the vector projection method is given along with a careful description of how the TFA results are combined with Bayesian decision theory to obtain a soft classification system. The combination of these two well-known methods is novel and constitutes a major portion of this research. The results from applying the TFA-Bayesian method to complicated data samples proved promising for current application and future research.

In the first experimental testing, gasoline was successfully detected at contributions as low as 1 – 5% in the presence of 90 – 94% substrate. Posterior probabilities greater than 0.9 were obtained for the Gas class, while the other two classes not included in the sample, ISO and MPD, had much lower posterior probabilities. Similarly, when the ISO and MPD classes were combined in concentrations as low as 1 – 5% each, in the presence of 85 – 93% substrate, the posterior probabilities for each of these two classes approached 0.5, while the posterior probability for Gas was very low. In the underdetermined problems, Gas was identified with a high posterior probability; however, the correlations between the target and recovered TIS were significantly lower than observed in tests that were not underdetermined.

Application of the TFA-Bayesian approach to the laboratory burn data resulted in four of the 20 samples where the maximum calculated posterior probability did not match the class of ignitable liquid used in the burn (i.e., 80% correct prediction of the correct class based on the largest posterior probability). Similarly, when the TFA-Bayesian method was applied to the large-scale burns, approximately 83% correct classification was obtained based on assigning the ignitable liquid to the class with the largest posterior probability. In both the laboratory- and large-scale burns, when the analysis was performed at a sensitivity level of 0.05, posterior probabilities were calculated for all of the burns. When the posterior probability calculation was repeated at $\alpha = 0.01$ for the large-scale burns, in one case none of the class probabilities exceeded α , and therefore the ignitable liquid residues from one burn could not be classified.

In considering this overall correct classification rate, several factors must be kept in perspective. The large-scale burns were all taken to flashover and there is absolutely no control over the amount of ignitable liquid residue remaining for analysis. Similarly, the degree of weathering is not controlled. In these regards, the data are very representative of real fire debris samples. This is, to our knowledge, the first multivariate classification model designed to work explicitly in the presence of an unknown and unspecified background signature. The correct classification rate of 83% is nearly identical to the 85% correct classification rate for ignitable liquids (in the absence of any interference) based on TIS similarity.[11] Finally, the addition of the sensitivity criterion makes the method conservative and by adjusting the lower correlation limit for integration in

combination with the sensitivity, the conservative nature of the method can be controlled. The disadvantages of the method are that it requires the analysis of data from multiple samples taken from a single fire scene, rather than a single sample, and the number of samples should exceed the “chemical rank” of the data, a condition that may be difficult to ensure in real samples. In this research, less than 15 samples were analyzed in each fire scene.

Implications for Policy and Practice: Fire debris analysts are currently allowed to opine that a sample is positive for ignitable liquid residue without the requirement to offer a known error rate for this assessment. It is reasonable to assume that at some point the admission of expert opinions in fire debris analysis is likely to meet a Daubert challenge based on the lack of a known error rate. This study represents one approach to obtaining error rates. Other approaches are also available from NIJ-sponsored research. In addition, this method provides a set of probabilities that the analyst may take into consideration when determining whether to call a sample positive or negative for ignitable liquid residue. Samples containing a significant signature for an ignitable liquid residue will result in large probabilities for the appropriate ignitable liquid class, and the probabilities may be helpful in providing statistical support for the analyst. We see perhaps a larger benefit coming when the ignitable liquid signature is weaker. The TFA-Bayesian method can identify the presence of an ignitable liquid at low levels, as shown in simulation studies, where gasoline was identified at contributions as low as 1 – 5%.

I. Introduction:

In many analytical chemistry and forensic science applications, a significant challenge exists in determining the presence of an analyte of interest when there is a significant and highly variable background signature. The difficulty of this problem is increased when the analyte in question is comprised of a complex mixture of chemical species. Fire debris analysis exemplifies this problem and the complexities it entails.

Statement of the problem: The research reported here directly addresses the broader analytical problem of detecting an analyte in the presence of a strong background signal for the purpose of determining the class of analyte present, but not the specific identity of the analyte. This problem is of general interest in the areas of signal processing, sensor development and forensic analysis. The forensic application specifically addressed in this research is the detection of ignitable liquid residue in the presence of pyrolysis background contributions and the classification of any ignitable liquid present into the classes defined under ASTM E 1618. We consider the solution to this problem to be the “Holy Grail” of forensic fire debris analysis. This problem is not readily solved by simply applying existing chemometric methods because: (1) classification methods that are currently available do not explicitly address the problem of background contributions to the overall signature of the sample and (2) most classifiers are defined as hard classifiers, while the existing soft classifiers do not directly provide probabilities of membership in multiple classes, nor do they offer the option of not assigning a sample to any class. The key to successfully addressing this problem is the recognition that in most analytical problems the number of analytes, or the number of classes of analyte, that are of interest are significantly less than the number of possible background interferences. Detection and classification of ignitable liquid residue in fire debris is representative of the “limited analyte

classes – multiple interferences” problem. The approach taken to solve the problem, as addressed here, is to identify ignitable liquids that potentially contribute to a sample by testing a library of ignitable liquids from the ASTM E1618 classes to determine if they contribute to a fire debris sample. More specifically, we seek the probability that a sample contains contributions from one or more ASTM class of ignitable liquid, and this information must be obtained irrespective of the background signature.

Literature review: Most of the ignitable liquids encountered in forensic fire investigations are commercial products that are comprised of multiple components. Varying amounts of the more volatile components are frequently lost through evaporation during the fire, a process referred to as “weathering” by fire debris analysts. The complexity of commercial products, coupled with weathering complicates the analysis of ignitable liquid residues in fire debris samples. The process is further complicated by the addition of background or interference compounds. In fire debris, these background components often arise from pyrolysis. The term pyrolysis is used here as defined by Webster: “chemical decomposition or other chemical change brought about by the action of heat regardless of the temperature involved.”[1] The term is often used in the chemical literature to refer to decomposition in the absence of oxygen; however, the amount of oxygen present at any given time in a fire is difficult to define but absent from Webster’s definition and therefore the term pyrolysis is used as defined by Webster. Pyrolysis converts common building and furnishing materials into a complex mixture of hydrocarbons, and many of the pyrolysis products may be coincidental with those compounds found in commercial ignitable liquids. When the amount of pyrolysis products is large compared to the ignitable liquid residue, the problem is to identify the pattern of the ignitable liquid in the total ion chromatogram (TIC) of the fire debris sample. On the other hand, when the amount of pyrolysis products is small, the TIC from the fire debris may closely match the TIC from a reference ignitable liquid and the problem can simplify to one of explaining the origin of a few extraneous peaks from the fire debris sample. In the United States, common practice requires that the ignitable liquid pattern of the TIC be readily discernible by visual pattern recognition following the American Standard Testing and Materials (ASTM) E1618-10 Standard Test Method.[2]

Despite the level of weathering and interference from pyrolysis products, the problem in fire debris analysis is not one of identifying a specific ignitable liquid or commercial product. Instead, the problem is to determine if a chromatographic profile and the chemical composition leading to that profile is consistent with one of the seven “classes” of products, or the miscellaneous category, as defined under ASTM E 1618. The ASTM classes include gasoline (Gas), petroleum distillates (PD), *iso*-paraffinic products (ISO), aromatic products (AR), naphthenic paraffinic products (NP), normal alkane products (NA) and oxygenated solvents (OXY). The classes may be subdivided based on carbon range as light ($C_4 - C_9$), medium ($C_8 - C_{13}$) and heavy ($C_8 - C_{20+}$). The PD class is frequently broken into the light, medium and heavy sub-classifications; LPD, MPD and HPD respectively. The OXY class is a collection of all ignitable liquids that contain a major oxygenated component (i.e., ketone, alcohol, ester, etc.). Under this definition, the OXY class may also be considered a derivative class wherein a member of any other class may conform to the OXY definition simply by addition of an oxygenated component. The miscellaneous (MISC) category does not possess class characteristics and is simply a collection of those ignitable liquids that do not fit into the other defined classes.

While fire debris analysts rely heavily on the TIC and extracted ion chromatograms for the classification of ignitable liquids, they do so by visual pattern recognition. In order to automate the process of classification so that it may be applied across multiple laboratories, it is necessary to either define a method of establishing fixed retention times across laboratories or move to a system that is based on a parameter with little or no inter-laboratory variability. Robust classification schemes cannot be built on highly variable parameters.[3] One approach would be to base matches on the similarity between covariance maps of GC-MS data sets. [4,5] Covariance maps are time-independent data representations that have been demonstrated to be useful for statistical comparisons of ignitable liquids. Covariance mapping,[6,7,8,9] and coincidence measurements,[10] have also been applied to time-of-flight mass spectrometry to resolve correlated events. Calculating covariance maps from GC-MS data sets is a CPU and memory intensive process. Alternative approaches to time-independent representations of GC-MS data have been investigated.

The total ion spectrum (TIS) is an average mass spectrum measured over the chromatographic profile, and has been shown to give 85% correct classification into the primary ASTM classes in an inter-laboratory test of neat ignitable liquids. [11] The TIS has also been shown to provide database search results similar to covariance mapping, and the calculation of a TIS and comparisons between two TIS require fewer computational resources than required for covariance maps. Figure 1 shows a heat map comparison of the similarities between TIS from over 500 ignitable liquids. High similarities are colored red and low similarities are colored blue. The lower left to upper right diagonal reflects the similarity of each liquid with itself, which is defined as a similarity of one. The off-diagonal elements show the similarities between different ignitable liquids. Similarities are clustered such that the most similar row-wise and column-wise patterns are grouped. The result is the heat map which shows areas of red along the diagonal, with the highest similarities corresponding to the PD and NP classes of ignitable liquid (upper right corner). Smaller clusters corresponding to MPD, NP, HPD and LPD can be seen within the larger PD/NP clusters. Near the center of the heat map, the NA cluster can be observed with two clusters of ISO along the diagonal on either side of NA. A small cluster of the MISC category and OXY class occurs between the ISO/NA and a small cluster of the Gas class. Below the Gas cluster is a cluster of primarily HPD ignitable liquids that show slightly weaker similarities. It is not clear why these HPDs cluster away from those in the upper right-hand corner of the heat map. Finally, along the diagonal in the lower left-hand corner of the heat map are a series of small clusters corresponding to the heavy, medium and light members of the AR class. The AR class show very low similarities with members of other classes of ignitable liquid. Interspersed among the clusters are primarily MISC and OXY ignitable liquids. The heat map shown in Figure 1 was constructed based on an algorithm using the Manhattan distance with a single linkage model. Heatmaps based on algorithms using other distances (i.e., Euclidean, Minkowski, etc.) and different linkages (i.e., centroid, complete, etc.) give slightly different results.

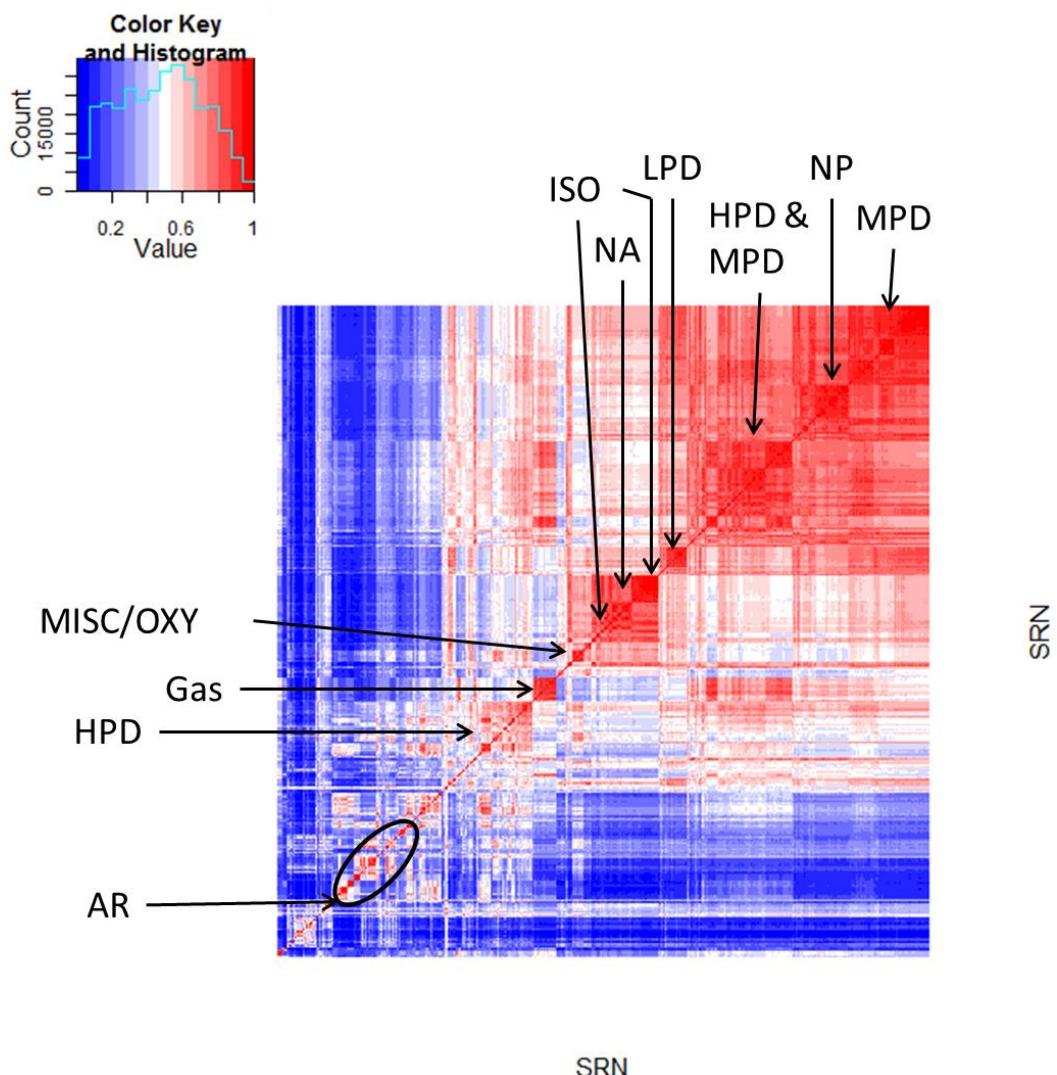


Figure 1. Heat map showing similarities between the TIS of different ignitable liquid classes designated by a sample reference number (SRN).

When a similar comparison is made between the same set of ignitable liquids and TIS from pyrolysis products from 60 common commercial building materials, the resulting heat map is shown in Figure 2. The heat map is now asymmetric and the dendrogram for the substrate clustering is shown on the top border of the graph. The heat map is dramatically different from the ignitable liquids heat map. Most of the ignitable liquid TIS are seen to have similarities with pyrolysis products of less than approximately 0.5. The only area of high similarity observed in the heat map is between the PD/NP classes and a group of pyrolysis products comprised of polyester and olefin. This indicates that the TIS may be useful in discriminating between pyrolysis products and ignitable liquids from most ASTM classes. The TIS from complex ignitable liquids are coupled with a multivariate statistical technique and Bayesian decision theory to form the classification method described in this work.

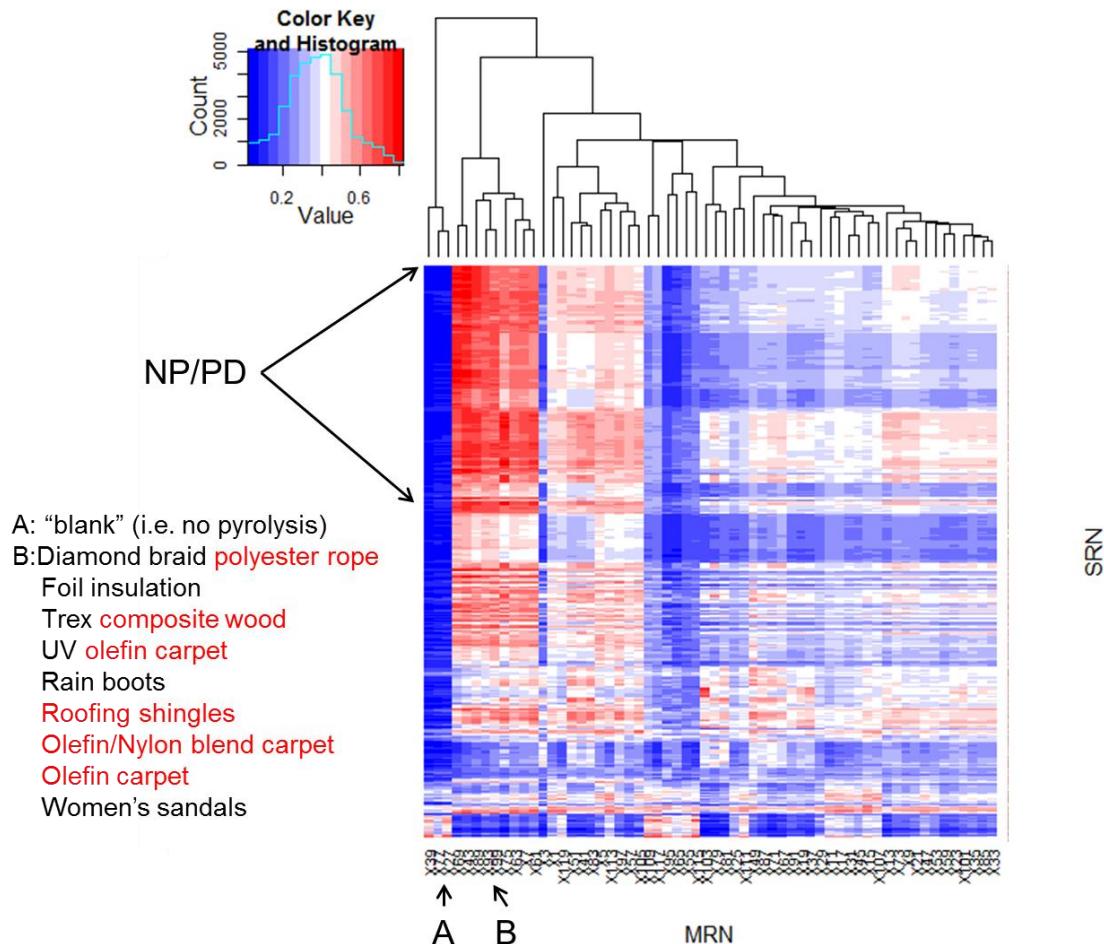


Figure 2: Heat map of similarities between ignitable liquid designated by sample reference number (SRN) and pyrolysis product TIS designated by material reference number (MRN).

Multivariate statistical techniques that have previously been applied to classification problems have not explicitly address the problem of interferences. There have been attempts to include interference into the training data, as described below, as a way of improving prediction quality. These attempts have been met with varying levels of success. The combined use of Principal components analysis (PCA) and linear discriminant analysis (LDA) has been applied to a set of gasoline samples for the purpose of discriminating between samples.[12,13,14] The method correctly classified 96% of the fresh gasoline samples, but only 51% of the evaporated gasoline samples. These studies used reference liquid gasoline samples rather than fire debris samples. Recent work from UCF (Award No. 2009-DN-BX-K227) has examined LDA and quadratic discriminant analysis (QDA) for the classification of ignitable liquids into the ASTM classes. [15] The method utilized TIS and was applied across all ASTM classes, except OXY and MISC. In addition, a single discrimination between samples containing ignitable liquid residue and those that did not contain ignitable liquid residue was included in the multi-step discrimination. Incorporating up to 20% pyrolysis products produced less than 10% false positive rate for the presence of ignitable liquid residue. PCA has also been combined with Pearson product moment correlation (PPMC) in an attempt to "associate" ignitable liquid residues with the corresponding neat ignitable liquids in the presence of interferences.[16] The study incorporated 12 ignitable liquids from six ASTM classes. The study demonstrated the difficulties of showing quantifiable

“associations” between neat ignitable liquids and ignitable liquid residues extracted from fire debris (i.e., burned nylon carpet substrate in this case). Principal components analysis is not an optimal classification technique. The goal of PCA is to reduce the dimensionality of a data set while maintaining the maximum possible variance in the minimum number of latent variables. Principal components analysis and soft independent modeling of class analogy (SIMCA) were tested for the classification of a set of 51 ignitable liquids chosen from five ASTM classes.[17] High correct classification rates were reported for SIMCA in these tests, which involved addition of solvents to previously burned substrates, thus avoiding much of the weathering process that characterizes fire debris analysis.

Some highly innovative approaches, although less conventional, have also been tested for classification of ignitable liquids. A comparison of gas chromatography-mass spectrometry (GC-MS) and gas chromatography-differential mobility spectrometry (GC-DMS) has been reported for the analysis of fire debris using projected difference resolution, a two-way data analysis method.[18] The number of prediction errors from one-way data analysis (i.e., comparison of chromatographic profiles), exceeded those from the two-way analysis. The research examined neat liquids, residues extracted from unburned substrate, and residue extracted from previously burned substrates. The abundant fragment ions in the integrated MS spectra (equivalent to the TIS) provided enough information to classify different ignitable liquids. The integrated MS spectra gave comparable correct classification to the two-way data for the neat and burned samples. An additional, somewhat unconventional approach, unsupervised self-organizing feature maps (SOFM) has successfully linked evaporated lighter fluids to the corresponding un-evaporated liquid for a set of 15 different samples.[19] Successful classification by lighter fluid brand required the use of hierarchical cluster analysis coupled with SOFM, whereas SOFM alone was reported to be more robust in identifying visual similarities and differences in the chromatographic data.

Hypothesis or rationale for the research: As described in the preceding section, multivariate statistical techniques that have been applied to the classification of ignitable liquids do not explicitly account for background interferences and many of the methods are hard classifiers that require each sample be assigned to one class. Current fire debris analysis practice does not utilize statistical methods. Current fire debris analysis methods have no known error rates, and rely totally on visual pattern recognition which is inextricably tied to the use of chromatographic data. While the goal of this research was to develop alternatives that can overcome some of the challenges in fire debris analysis, it must be pointed out that current protocols are not inherently flawed and that fire debris analysis is among the most complex of all forensic analyses. The goal of this research is not to replace or substantially change current practice, but rather, to provide statistical-based chemometric methods to enhance current practice. This research is based on the following hypotheses: (1) the signature of an ignitable liquid can be extracted from the signature of pyrolysis products in a significant percentage of cases, (2) classification can be based on time-independent data representations that are comparable between laboratories and (3) a multivariate soft classification method can be developed that will provide the analyst with probabilities of class membership as defined under ASTM E1618-10. These propositions were addressed by the synthesis of a chemometric method based target factor analysis and Bayesian decision theory. The time-independent total ion

spectrum of ignitable liquids reference samples are used as the target factors, giving the method inter-laboratory applicability.

II. Methods

Experimental Methods: The chemometric method of target factor analysis requires a library of target factors. In the problem addressed in this research, the target factors will be a set of TIS for reference ignitable liquids. The target factors will be drawn from the Ignitable Liquid Reference Collection (ILRC), created and maintained by the National Center for Forensic Science (NCFS) in collaboration with the Technical Working Group for Fire and Explosions (TWGFEX). Each ignitable liquid in the collection has been assigned to an ASTM E1618-designated class. The TIS was calculated for each of the over 500 commercial ignitable liquids in the database. Spectra in the ILRC library originated from 20 microliters of a reference ignitable liquid diluted with 1 milliliter of carbon disulfide and analyzed by GC-MS as described below. Similarly, GC-MS data for substrate pyrolysis products have been taken from the NCFS/TWGFEX Substrate Database. Spectra in the Substrate library originated from materials heated to produce pyrolysis and combustion products. These combustion and pyrolysis products were adsorbed onto activated carbon and then desorbed with carbon disulfide for GC-MS analysis, as described below.

Test fire debris samples were created in the laboratory by depositing ignitable liquids onto substrate materials prior to heating. Substrate materials such as building materials, flooring, and furniture were purchased from local home improvement and furniture stores. Decomposition products from these materials were produced by a modified destructive distillation method based on a procedure developed by the State of Florida Bureau of Forensic Fire and Explosives Analysis. The materials were cut to approximately 6 cm² in size, placed upside down inside a un-lined metal quart-size paint can (Best Containers, Eagle, ID), and then a perforated lid with nine 1 mm diameter holes was laid loosely onto the can, see Figure 3a. Heat was applied to the bottom of the can by a propane torch held at a distance of 4 cm from the bottom as shown in Figure 3b. Heating continued 2 minutes after smoke appeared from the holes in the can lid. Once the heat was removed, the perforated lid was replaced with a solid lid to retain vapors within the headspace and the can cooled to room temperature. After the can reached room temperature, an activated carbon strip was suspended in the headspace of the can to adsorb the resulting combustion and pyrolysis products along with any ignitable liquid residues. The cans were heated inside an oven at 66°C for approximately 16 hours. After the 16 hours, the cans were removed from the oven and cooled to room temperature. The adsorbed residues were desorbed from the activated carbon with carbon disulfide for GC-MS analysis. These conditions are based on procedures used for the Substrate Database. The target factor data analysis method requires that multiple samples, each potentially containing different contributions from an ignitable liquid and substrates, be analyzed as a composite dataset. Typically 6 – 12 samples were prepared for each ignitable liquid by varying the volume of (25 – 2000 µL) that was deposited onto the substrate or using various multiple substrates in each sample with a constant volume of ignitable liquid. The total ion spectra from the GC-MS analyses for a group of fire debris samples were concatenated into a single data matrix for further data analysis, as described below.

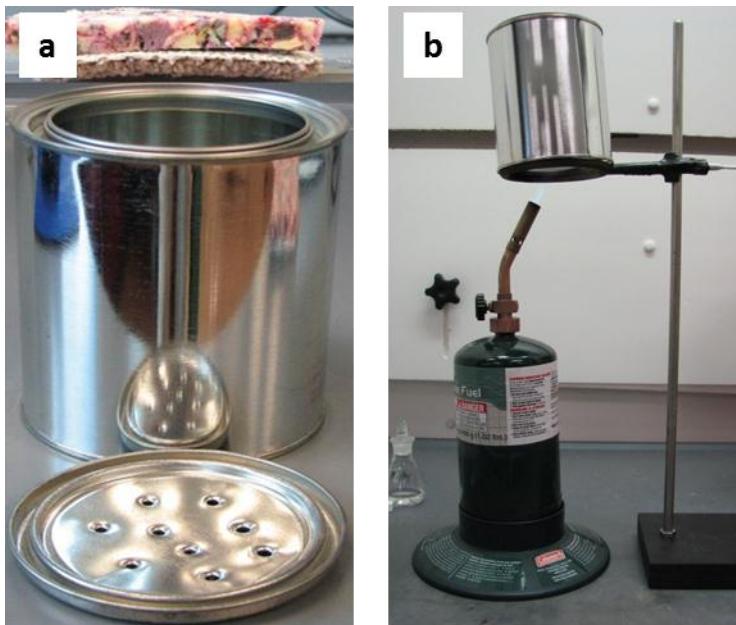


Figure 3. Laboratory burn setup. (A) Burn can showing lid with ventilation holes. (B) Ringstand and propane torch heating bottom of can.

Large-scale burns were conducted at the Florida Fire College using four $2.4 \times 2.4 \times 6.1 \text{ m}^3$ Konex shipping containers fabricated inside with sheetrock walls and ceilings to construct a two room structure. Each container had a three-sided $1.2 \times 1.2 \text{ m}^2$ opening at the rear of the container to resemble a window which allowed limited control of air to support the fire. The majority of the plywood floors were covered with carpet and carpet padding with a small area covered in either vinyl or wood laminate. The front room was furnished with a chair, sofa, and table to resemble a living room and the back room was furnished with a bed and dresser to resemble a bedroom. The building and flooring materials were purchased from Home Depot and the furnishing were purchased from either Wal-Mart or IKEA. Figure 4a shows a photograph taken inside one of the burn containers prior to ignition. Additional clothing, papers and plastic items were added to each container to increase the fuel load. Five hundred milliliters of ignitable liquid was poured from the front of the container at the couch and trailed to the back of the container at the bed, see Figure 5. The fire was initiated when a fire fighter placed a torch at the pour sight on the couch, see Figure 4b. The duration of the fires were typically 5 – 15 minutes, during which time, the containers reached temperatures as high as 870°C and the containers typically achieved a condition of “flashover”. These burns produced fire debris samples resembling those obtained from structural fires encountered by forensic laboratories. After extinguishment with water, see Figure 6, and subsequent cool down of the container, most samples were collected on or near the ignitable liquid pour trail with a few samples collected from areas remote of the pour trail, see sample markers in Figure 7. The fire debris samples collected were placed into individual one gallon metal paint cans. Any ignitable liquid, pyrolysis and combustion product residues were adsorbed onto activated carbon, as described in the following paragraph and subsequently desorbed with carbon disulfide for GC-MS analysis.

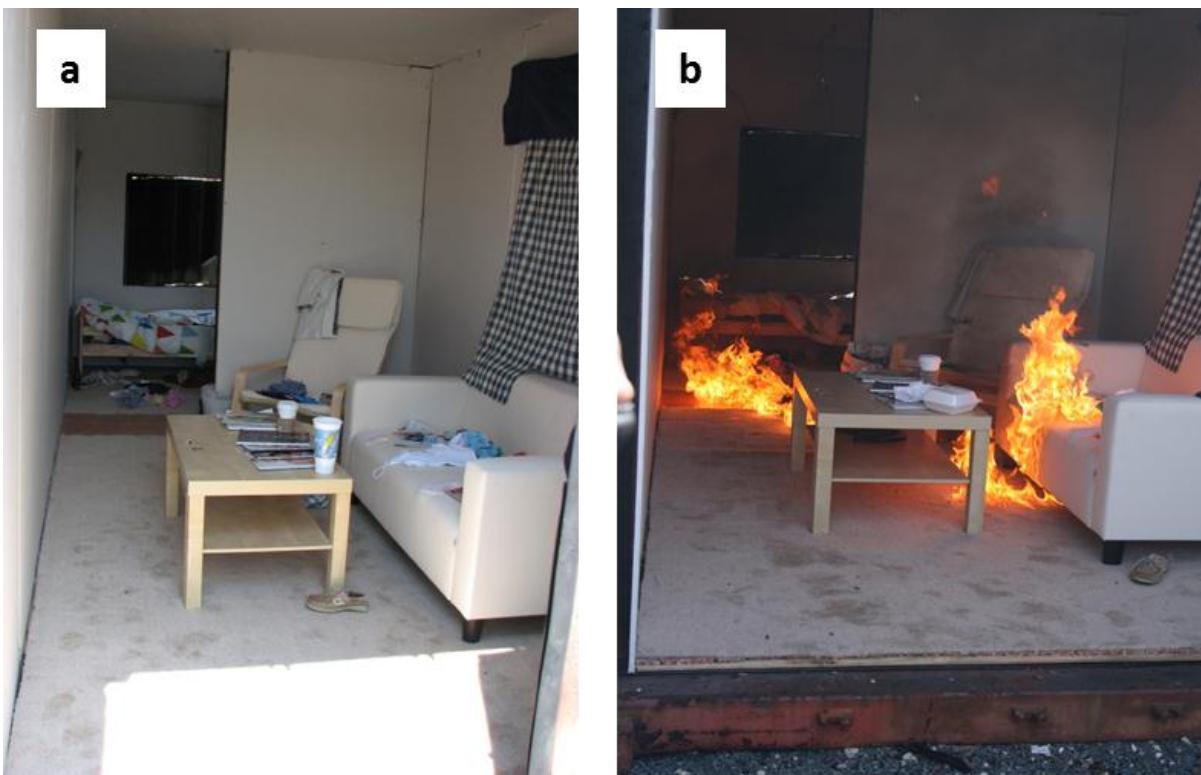


Figure 4. (a) Burn container before ignition and (b) after ignition.

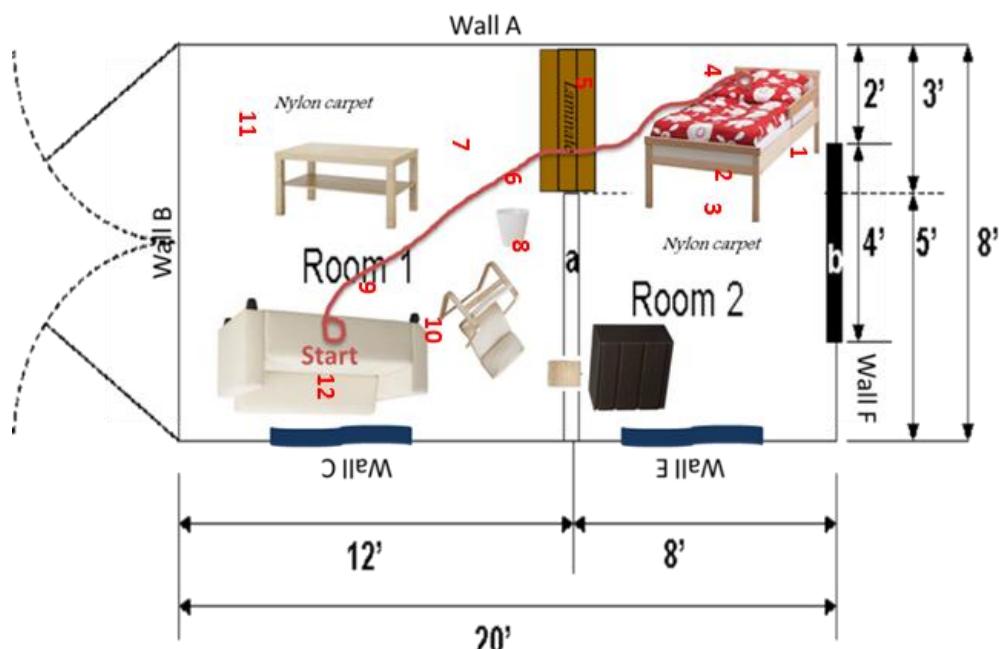


Figure 5. Diagram of the burn container showing front room with small sofa, chair, table and waste basket. Back room contained bed and small chest-of-drawers. Partition between rooms left a 3 foot opening. Back wall of container had a 4 foot “window” cut on three sides to allow it to open approximately 1 foot for air flow. The double doors on the front of the container were held at approximately 6 inches open during the burn. The red line in the diagram shows the pour-trail.



Figure 6. Extinguishing of large-scale burns was accomplished with water after the containers had reached flashover.



Figure 7. Post-burn debris photo showing the placing of sample markers along the pour trail.

The ignitable liquid residue along with pyrolysis and combustion products from the laboratory and large-scale burns were collected from the fire debris by passive headspace adsorption onto activated carbon following the ASTM E1412-07 standard method.[20] A 10 x 22 mm² activated carbon strip (Albrayco Technologies, Inc., Cromwell, CT) was suspended into the headspace of the can by a paperclip and un-waxed dental floss. The sealed can was heated for 16 hours at 66°C. Once cooled to room temperature, the activated carbon strip was removed from the can and cut in half lengthwise. One half was archived and the other was deposited into a vial with 1 ml of low benzene carbon disulfide (Fisher Scientific) for GC-MS analysis.

Gas chromatography-mass spectrometry analysis was performed in accordance with ASTM E1618-10 [2] utilizing an auto-sampler on an Agilent 6890 gas chromatograph interfaced to a 5973 mass spectrometer. One microliter of the CS₂-diluted ignitable liquid was injected into a 250°C injection port. The compounds were separated by a 100 % dimethylpolysiloxane (HP-1) capillary column with a film thickness of 0.50µm, a nominal diameter of 200 µm, and 25 m length. Helium gas was maintained at a constant flow rate of 0.8 ml/min with an average velocity of 36 cm/sec. The injection was split in a 50:1 ratio. The initial oven temperature of 50°C was held for 3 minute, followed by a temperature ramp of 10°C/min to a final temperature of 280°C, which was held for 4 minutes. The mass spectrometer transfer line temperature was 250°C with a source temperature of 230°C and a quadrupole temperature of 150°C. Mass spectra were scanned between 30 and 350 mass to charge ratio at an acquisition rate of 2-3 scans/second. The detector was turned on at 2.00 minutes after solvent elution.

The 3D GC-MS data from each sample was exported as a comma-separated values file and the total ion spectrum (from 30 – 200 m/z) for each sample was calculated by summing each mass channel over the chromatographic profile and normalizing the data such that the intensities across the entire m/z range summed to one. Each row of the data matrix corresponded to a sample and each column corresponded to a variable (i.e., m/z ratio). An example can be found in the appendix. The library spectra from the ILRC were prepared the same way and compiled for target factor analysis. The TIS were analyzed as described in the following section.

Statistical Methods: The classification methods discussed in “Section 2; Literature citations and review”, operate by classifying one sample at a time. Target factor analysis works by the simultaneous analysis of a set of data (i.e. data from multiple samples). In the case of interest here, the TIS from multiple samples collected from a single fire scene would be concatenated into a data matrix and analyzed. In this work, the number of spectra analyzed is typically 10 – 15. The significance of the number of spectra will become apparent upon further discussion of the mathematical approach. A set of spectra, each composed of multiple contributing factors, may be analyzed by target factor analysis (TFA). Target factor analysis begins with a reduction in the dimensionality of the data by abstract factor analysis, a process highly analogous to PCA, to identify a limited set of latent variables that represent a significant portion of the variance in the data. In theory, each entry in the data matrix, d_{sv} , should conform to Equation 1, which defines the linear nature of the approach. In Eq. 1, there are n contributing terms in each sum, each corresponding to a contributing factor. The values of r_{sk} and c_{kv} can correspond to concentrations and response in sample s at variable v .

$$d_{sv} = \sum_{k=1}^n r_{sk} c_{kv} \quad (1)$$

While Eq. 1 defines the underlying data structure that is required for either abstract factor analysis or PCA, these methods are often applied without verification of the underlying data structure, especially in the social sciences. However, in the problem addressed here, Eq. 1 will hold in the absence of complicating factors that lead to nonlinearity, such as detector saturation which can be avoided by proper choice of experimental conditions.

In abstract factor analysis, the $s \times v$ data matrix $[D]$, composed of s samples and v variables, is decomposed into the product of $[R]$, an $s \times N$ matrix, and $[C]$, an $N \times v$ matrix, where $[C]$ consists of the original variable loadings on an orthogonal set of eigenvectors (latent variables) and $[R]$ contains the scores, which give the relative contributions required of each eigenvector to reproduce the data. The eigenvectors are achieved by singular value decomposition in this research; however, other methods are available.[21] The eigenvalues associated with each eigenvector is proportional to the fraction of variance in the data which is accounted for by the respective eigenvector. There are a maximum of $N = s$ or $N = v$ eigenvectors that account for 100% of the variance in $[D]$, with N corresponding to the smaller of s or v . The number (n) of eigenvectors needed to retain a given percentage of the variance in the data is selected and the data matrix reproduced from the n eigenvectors will be written as $[D^{\#}]$. Similarly, the reduced dimensionality scores and loadings matrices are designated as $[R^{\#}]$ and $[C^{\#}]$, respectively. Determination of the number of latent variables to retain may be accomplished by various methods, including the determination of rank by median absolute deviation (DRMAD), which is used in this work. [22] Malinowski defines target transformation of the scores matrix $[R^{\#}]$ by beginning with the following definitions. When the number of factors retained, n , is less than N , the reproduced data will differ from the experimental data. This difference is generally expressed as an error matrix $[E]$ which is of dimension $s \times v$. In the following equations, $[E]$ is dropped as an additive term and it is understood that the associated error has been removed from the recovered data, $[D^{\#}]$, as shown in Equation 2.

$$[D^{\#}] = [R^{\#}]_{\text{abstract}} [C^{\#}]_{\text{abstract}} \quad (2)$$

The subscripts “abstract” in Eq. 2 emphasize that the abstract factor solutions do not correspond to physically meaningful factors. The above equation can be rewritten as Eq. 3, where $[T][T]^{-1}$ equals the identity matrix, $[\Pi]$.

$$[D^{\#}] = [R^{\#}]_{\text{abstract}} [T][T]^{-1} [C^{\#}]_{\text{abstract}} \quad (3)$$

The product of the transformation matrix $[T]$ with $[R]_{\text{abstract}}$ (or $[T]^{-1}[C]_{\text{abstract}}$) corresponds to “real” (i.e., physically meaningful) solutions as shown in Eq 4 – 5.

$$[R^{\#}]_{\text{real}} = [R^{\#}]_{\text{abstract}} [T] \quad (4)$$

$$[C^{\#}]_{\text{real}} = [T]^{-1} [C^{\#}]_{\text{abstract}} \quad (5)$$

Malinowski uses the notation $[\bar{R}]$ (read as “r bar”) to denote the transformed scores matrix, Eq. 6, and $[\bar{C}]$ to denote the transformed loadings matrix, Eq. 7. In addition, the “abstract” subscript is dropped so that:

$$[\bar{R}] = [R^\#] [T] \quad (6)$$

$$[\bar{C}] = [T]^{-1} [C^\#] \quad (7)$$

In seeking to find the underlying factor structure of the data it is possible to identify the elements of either $[T]$ or $[T]^{-1}$. In his classic text, Malinowski has derived the equations for the transformation of $[R^\#]$ by finding the elements of $[T]$ through target transformation and maximum likelihood (in a least squares sense) oblique rotations. Similarly, it is possible to find the elements of $[T]^{-1}$ to accomplish target transformations of $[C^\#]$, although Malinowski does not address this approach. It is worth noting that results identical to the transformation of $[C^\#]$ can be obtained following Malinowski's classic text if one chooses to diagonalize the covariance matrix, $[Z]$, defined by pre-multiplying $[D]$ by its transpose, rather than defining $[Z]$ as $[D][D]^T$. However, if we choose to define $[Z]$ as $[D]^T[D]$, we must think of the eigenvectors as corresponding to combinations of sample composition, rather than spectra. It is convenient to consider the transformation of $[C^\#]$ while maintaining the conventional arrangement of $[D]$ where each row corresponds to a sample and each column corresponds to a variable. In the case examined here, the variables correspond to m/z ratio and the test vectors that will be available for target transformation are TIS from commercial ignitable liquids. The following derivation is given for finding the elements of $[T]^{-1}$ to accomplish targeted oblique rotations of $[C^\#]$. The results of the derivation are particularly simple and readily interpreted within the construct of PCA, which is likely to be more familiar to forensic science students and practitioners. The derivation is also provided to enhance understanding of the result and facilitate the implementation by forensic scientists.

We begin the derivation with Eq. 8, where the subscript “*abstract*” has been dropped, as described above.

$$[D^\#] = [R^\#] [T] [T]^{-1} [C^\#] \quad (8)$$

It is not necessary to find the entire transformation matrix $[T]^{-1}$. Instead, it is possible to find a single row of $[T]^{-1}$, as expressed in Eq. 9.

$$\bar{C}_l = T_l [C^\#] \quad (9)$$

The vector \bar{C}_l is the predicted vector of dimension $1 \times v$ and T_l is a row vector of dimension $1 \times n$ such that T_l is the l^{th} row of the transformation matrix $[T]^{-1}$. Similarly, \bar{C}_l is the l^{th} row of the transformed column matrix $[\bar{C}]$. We wish to find T_l that yields \bar{C}_l which most closely matches a test vector denoted as $\bar{\bar{C}}_l$. The transformation row vector T_l can be written as in Eq. 10.

$$T_l = t_{l1}, t_{l2} \dots t_{ln} \quad (10)$$

The i^{th} column of $[C^\#]$ is designated as given in Eq. 11:

$$C'_i = c_{1i}, c_{2i} \dots c_{ni} \quad (11)$$

The dot product of T_l with C'_i gives the projection of the i^{th} column of $[C^\#]$ onto the new axis, \bar{C}_{li} , and the dot product may be written as shown in Eq. 12.

$$\bar{C}_{li} = t_{l1}c_{1i} + t_{l2}c_{2i} + \dots + t_{ln}c_{ni} \quad (12)$$

Multiplying each column of $[C^\#]$ by T_l similarly gives the projection of each column of $[C^\#]$ onto the new axis, resulting in $\bar{C}_{l1}, \bar{C}_{l2} \dots \bar{C}_{lv}$. The projected vector is compared to the test vector $\bar{\bar{C}}_l$, having components $\bar{C}_{l1}, \bar{C}_{l2} \dots \bar{C}_{lv}$. The difference, $\bar{C}_{li} - \bar{\bar{C}}_{li}$, is denoted ΔC_{li} and is expressed as shown in Eq. 13.

$$\Delta C_{li} = \bar{C}_{li} - \bar{\bar{C}}_{li} = t_{l1}C_{1i} + t_{l2}C_{2i} \dots t_{ln}C_{ni} - \bar{\bar{C}}_{li} \quad (13)$$

To perform the least squares maximum likelihood optimization, we set the sum of the derivatives of squares of the difference from Eq. 13 equal to zero. The square of the derivative can be expressed as in Eq. 14. The multiplier 2 can be factored out of each term in Eq. 14 and the terms summed over the v columns following a simple rearrangement, as shown in Eq. 15.

$$\frac{d(\Delta C_{li})^2}{dt_{li}} = 2t_{l1}C_{li}^2 + 2t_{l2}C_{1l}C_{2i} + \dots + 2t_{ln}C_{1i}C_{ni} - 2C_{1i}\bar{\bar{C}}_{li} = 0 \quad (14)$$

$$\sum_{i=1}^v C_{1i}\bar{\bar{C}}_{li} = t_{l1}\sum_{i=1}^v C_{li}^2 + t_{l2}\sum_{i=1}^v C_{1l}C_{2i} + \dots + t_{ln}\sum_{i=1}^v C_{1i}C_{ni} \quad (15)$$

Similar expressions can be written for each column in order to optimize with respect to each element of T_l . The resulting equations can be written in the following matrix form, where A_l , T_l and $[B]$ are given by Eq. 16, 17 and 18 respectively.

$$A_l = T_l[B] \quad (16)$$

$$A_l = [\sum_{i=1}^v C_{1i}\bar{\bar{C}}_{li} \quad \sum_{i=1}^v C_{2i}\bar{\bar{C}}_{li} \quad \dots \quad \sum_{i=1}^v C_{ni}\bar{\bar{C}}_{li}] \quad (17)$$

$$T_l = [t_{l1} \quad t_{l2} \quad t_{l2} \quad t_{ln}] \quad (18)$$

$$[B] = \begin{bmatrix} \sum_{i=1}^v C_{1i}^2 & \sum_{i=1}^v C_{1i}C_{2i} & \dots & \sum_{i=1}^v C_{1i}C_{ni} \\ \sum_{i=1}^v C_{1i}C_{2i} & \sum_{i=1}^v C_{2i}^2 & \dots & \sum_{i=1}^v C_{ni}C_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^v C_{1i}C_{ni} & \sum_{i=1}^v C_{2i}C_{ni} & \dots & \sum_{i=1}^v C_{ni}^2 \end{bmatrix} \quad (19)$$

From Eq. 17, A_l can be rewritten as the product shown in Eq. 20, and $[B]$ can be rewritten as the product shown in Eq. 21, which simplifies to the identity matrix, $[I]$ by virtue of the orthonormal nature of $[C^\#]$.

$$A_l = \bar{\bar{C}}_l[C^\#]^T \quad (20)$$

$$[B] = [C^\#][C^\#]^T = [I] \quad (21)$$

The final formula for calculation of T_l which will give the least squares best fit \bar{C}_l , Eq. 22, is then given by substitution of equations 20 and 21 into equation 16.

$$\bar{C}_l[C^\#]^T = T_l \quad (22)$$

After the transformation vector, T_l , is found, Equation 9 is then used to calculate \bar{C}_l which is compared by linear regression to the target vector, $\bar{\bar{C}}_l$. In this work, the Pearson product moment correlation coefficient, r , is used to judge the similarity of the test and predicted vectors. The property of target factor analysis that allows the search for one transformation vector at a time can then be utilized to identify factors that contribute to a data set, even in the presence of interferences, as will be demonstrated below.

It is helpful to view the simple result of this derivation from the perspective of PCA, which is likely more familiar to forensic scientists. Equation 2 gives the general result of a PCA calculation. Equation 22 can be viewed as the projection of the data for a new sample (i.e., \bar{C}_l) into the principal component space represented by the solution given in equation 2. The resulting T_l can be viewed as the scores from the projection. In other words, T_l would define the combination of eigenvectors that best represent \bar{C}_l . Applying Eq. 9 with the calculated T_l gives \bar{C}_l , the best possible representation of the new data that can be achieved from the orthogonal basis set. A direct comparison of \bar{C}_l and $\bar{\bar{C}}_l$ by least squares gives the correlation between the two. A simple example of this result is shown in Figure 8, which depicts a data set (filled blue symbols) comprised of variables X, Y and Z. It can be seen from Fig. 8, that the data can be expressed as a function of two latent variables (i.e., two principal components), such that the data lies in a plane represented by the grid of black lines. If a new data point (pt1 shown by an open red circle) is projected into the PC-space, the scores for the projected point would be given by the coordinates of the blue cross symbol (pt1_r) in the coordinate system defined by the black grid. These scores would also correspond to the components of T_l , as discussed above. If the scores were projected from the 2D space back into the original 3D (XYZ) space, the coordinates would be given by location of the blue cross in the XYZ coordinate system, as shown in Fig. 8. Note the large distance between pt1 and pt1_r in the original space. A simple correlation between the XYZ coordinates of pt1 and pt1_r would yield a small correlation coefficient. In other words, pt1 cannot be accurately represented by the latent variables that define the 2D PC-space. A similar exercise is repeated for pt2 (green cross) and the recovered data, pt2_r, is shown by the open triangle in Figure 8. Now the distance in XYZ space between pt2 and pt2_r is quite small because it is possible to represent pt2 in terms of the latent variables. This relationship is easy to see when the original data space is composed of only three variables; however, when the original data space is comprised of tens or hundreds of variables and the PC space is defined by a large number of variables, the relationship is difficult, if not impossible, to visualize without the aid of linear algebra. It should be noted that the coordinates (scores) in the PC space are dependent on the phasing of the eigenvectors. The scores plot and loadings plot corresponding to the data in Figure 8 are shown in Figures 9a and 9b, respectively.

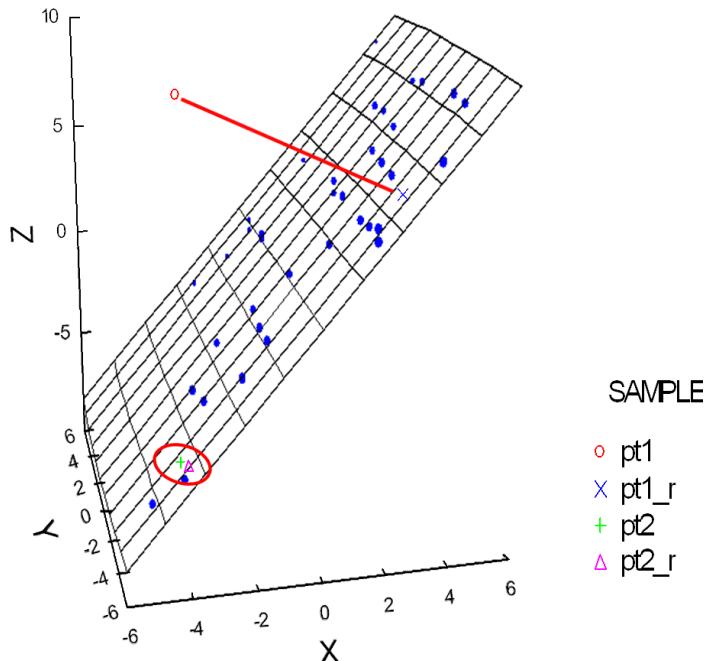


Figure 8. Example of data projection and recovery using a simple 3D data set.

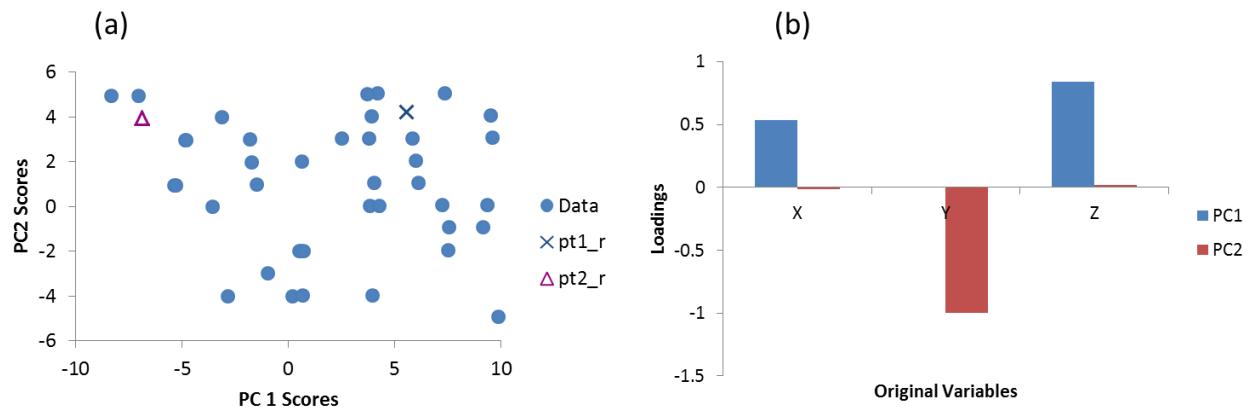


Figure 9. Traditional PC scores and loadings plot for data shown in Figure 8.

The theory and simple result outlined above can be applied by testing all of the members of a library of analytes to determine if they can be expressed as a combination of the latent variables derived from the abstract factor analysis of a data set. In the analysis of fire debris, the data set is comprised of the TIS from multiple samples coming from a fire scene. The library is comprised of the TIS from ignitable liquids in the Ignitable Liquids Reference Collection and Database (ILRC) maintained at UCF in collaboration with TWGFEX/SWGFEX. Each entry in the library has been classified into one of the ASTM E1618 classes by the TWGFEX Ignitable Liquids Committee. The question then becomes how to use this information, in combination with the target factor analysis results, to identify ignitable liquid residue in a fire debris sample. The connection is made through Bayesian decision theory.

Equation 23 forms the basis for Bayesian decision theory. In Eq. 23, the posterior probability, $P(\omega_i|r)$, is expressed as a function of the class-conditional probability, $p(r|\omega_i)$, and the prior probability, $P(\omega_i)$, of membership in class ω_i . [3] The probability distributions may be univariate or multivariate. In the discussion that follows, we treat the probabilities as univariate, where the variable is designated as r . Prior probabilities may be assessed in different ways, however, in the absence of a clear indication of the prior probabilities, they are set equal for each class. When the prior probabilities for each class are set equal, they cancel out of the numerator and denominator on the right-hand-side of Eq. 23. In the analysis of fire debris, the prior probabilities could be set based on previous frequencies of assigning ASTM classes to samples that were positive for ignitable liquid residue.

$$P(\omega_i|r) = \frac{p(r|\omega_i)P(\omega_i)}{\sum_i p(r|\omega_i)P(\omega_i)} \quad (23)$$

In a standard application of Bayesian decision theory, the class conditional probability curves are invariant and the value of r can be used directly to estimate class membership. For example, in Figure 10, if the two class conditional probability distributions are shown by the two curves, any sample having a value of r to the left of the vertical dashed line would be assigned to class A, and those to the right would be assigned to class B.

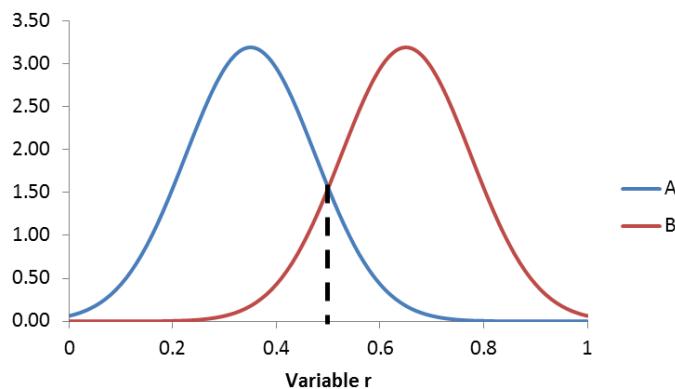


Figure 10. Simple two class classification example. Maximum probability of correct classification is achieved by assigning a sample to class A if it has an r less than the value where the vertical dashed line intersects the r axis. Otherwise the sample is assigned to class B.

The class conditional probability distributions are calculated from the correlation coefficients, r , obtained for each class as a result of the target factor analysis. There are no guarantees that the correlation coefficients for a given class of ignitable liquids will be normally distributed. Therefore, a kernel distribution function is used to calculate each class conditional probability distribution. Equations 24 – 26 give the kernel distribution function employed.

$$p(r|\omega_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{h_i \sqrt{2\pi}} \exp\left[-\frac{1}{2h_i^2}(r - r_j)^2\right] \quad (24)$$

$$h_i = 0.9A_i n_i^{1/5} \quad (25)$$

$$A_i = \min(s_i, \frac{R_i}{1.34}) \quad (26)$$

Equation 24, expresses the class-conditional probability as a superposition of Gaussian functions centered at the n_i correlation values for each class ω_i .[3] The value h_i is an adjustable bandwidth which may be calculated by Equation 25, where A_i is an adaptive estimate of the spread in the correlations for class ω_i .[23] In Eq. 26, s_i is the sample standard deviation of the correlations corresponding to class ω_i , R_i is the inter-quartile range of the correlations for class ω_i , and \min indicates that the minimum of the two values (s_i and $R_i/1.34$) is assigned to A_i . Equations 25 and 26 are reported to give a mean integrated squared error within 10% of the optimum for several commonly encountered non-normal distribution types.[23]

As an alternative to the conventional approach to Bayesian decision theory, the posterior probabilities may be utilized to estimate the degree of class membership, resulting in a Bayesian soft-classifier.[24] The soft classifier allows the user to view the probabilities and weight them against other evidence when making a decision. However, to truly be a soft classifier, the method must also have a way to reject any and all samples from being assigned to any of the classes.

The example shown in Figure 10 demonstrates the conventional approach to Bayesian decision theory. However, every time target factor analysis is performed on a set of samples from a fire scene, the positions and shapes of the class conditional probability curves change and the correlations for some classes may approach the upper bound of 1, as shown in Figure 11. Higher correlations correspond to closer matches between the target and recovered vectors. Consequently, one approach would be to calculate the posterior probability at r equal to one, where there is a perfect match between the target and recovered vectors. In the case shown in Figure 11, classes B and A would have posterior probabilities of 0.95 and 0.05, respectively, when the posterior probability is calculated at $r = 1$. With real samples, it is a very stringent standard to require perfect correlations.

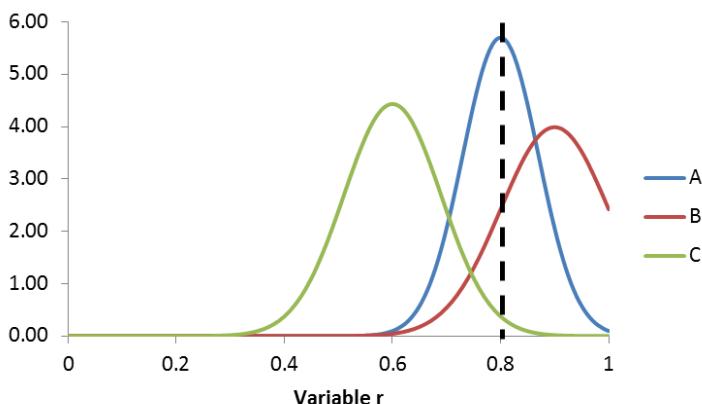


Figure 11. Class-conditional distribution functions for three classes (A, B and C) as they might result from target factor analysis. The vertical dashed line represents the correlation lower limit for integration (see text). The integrated area under the class-conditional probability over the limit range of -1 to r_{LL} gives the probability of obtaining a correlation coefficient for a given class that exceeds r_{LL} .

An alternative approach is to establish a lower acceptable limit for the correlation, r_{LL} . In this approach, the probability for obtaining a correlation greater than or equal to r_{LL} for each class is calculated by integrating the area under the class conditional probability distribution from r_{LL} to one. The correlation is bounded on the interval [-1,1], but the integral under curve B in Figure 11, when taken over the limits of -1 to 1, will not have an integrated area of 1. We avoid this boundary problem by defining the probability of obtaining a correlation greater than or equal to r_{LL} as the quantity $1 - I_i[-1, r_{LL}]$, where $I_i[-1, r_{LL}]$ is the integral for class i over the limits from -1 to r_{LL} . With this definition, we can rewrite the Bayesian equation as shown in Equation 27, where $p(r|\omega_i)$ has been replaced with $1 - I_i[-1, r_{LL}]$ and the posterior probability should now be read as the probability of class ω_i , given $I_i[r_{LL}, 1]$ and a sensitivity factor, α , which is described below.

$$P(\omega_i | I_i[r_{LL}, 1], \alpha) = \frac{(1 - I_i[-1, r_{LL}])P(\omega_i)}{\sum_i (1 - I_i[-1, r_{LL}])P(\omega_i)} \quad (27)$$

One last issue remains to be discussed. Given the formulation of the posterior probability, all of the probabilities will sum to one, as they should. However, it is possible that all of the classes would have very poor correlations between the test and transformed vectors, and yet their posterior probabilities would sum to one. The solution adopted in this work is to introduce a sensitivity parameter, α . If the integrated area under the class-conditional probability, $1 - I_i[-1, r_{LL}]$ for a given class did not exceed α , the posterior probability for that class would be set equal to zero. If the integrated area at correlations greater than r_{LL} did not exceed α for all classes, then a posterior probability would not be calculated and the sample could not be classified. This approach makes the classifier very conservative, so long as reasonable values are used for r_{LL} .

III. Results

Simulations: The TFA method was tested by constructing simulated data sets composed from the TIS from Gas, ISO and MPD samples and 10 substrates pyrolysis TIS, designated Sub1 – Sub10. The data sets each contained 10 spectra. Each spectrum in the set was composed of one or two ignitable liquid TIS, a single substrate TIS and random noise. When two ignitable liquids contributed to a data set, they were both mixed in each spectrum in amounts randomly chosen from a uniform distribution such that the total contribution from each ignitable liquid is in the range designated in Table 2. The identities and ignitable liquid and substrate samples, and their respective designations, are given in Table 1. The TIS for the three ignitable liquids and substrates 1 – 5 are shown in Figures 12 and 13, respectively. The compositions of the test sample data sets are given in Table 2. In the test labeled 1a, for example, each of the 10 spectra was comprised of gasoline in 1 – 15%, one of the substrates (Sub1 – Sub5) in 80 – 94% and random noise accounted for 5% of the total signal. The sum of all the peak intensities in each sample spectrum summed to a value of one. In test 1a, the data matrix had a rank of 6, corresponding to Gas and five substrate pyrolysis samples. Abstract factor analysis (or principal components analysis) found that six factors were required to reproduce the data, as determined by DRMAD, and the remainder of the factors accounted for the random noise. The eigenvectors

are not shown here, but the loadings in several vectors were large on the ions corresponding to m/z 91, 105, etc., as expected for gasoline. Examination of Table 2 reveals that the data matrices for tests 1a, 1b and 2a were found to have ranks corresponding to the number of factors (or components) used to make up the data matrix. In test 2b, two ignitable liquids (1 – 5% each) were combined with five substrates. The rank of the data matrix was determined by DRMAD to be eight, rather than seven which corresponds to the number of components contributing to the matrix. In this case the ignitable liquid contribution is very small, certainly less than an amount that would likely be called positive for an ignitable liquid in a fire debris sample. In real samples, it is preferable to retain more of the variance in the data, in order to ensure successful target transformation, as discussed below.

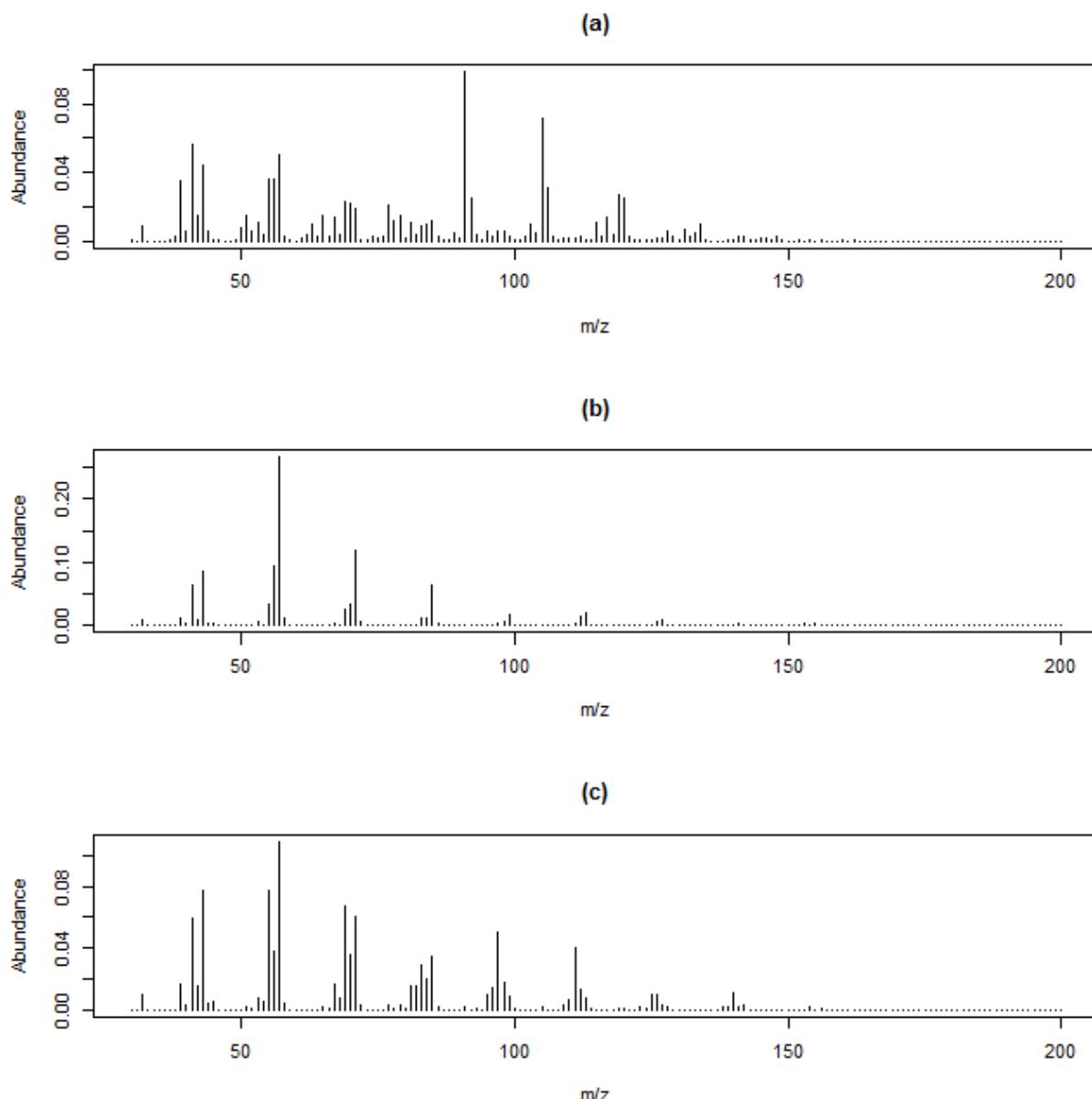


Figure 12: Total ion spectra used in simulation studies: (a) Gas, (b) ISO and (c) MPD.

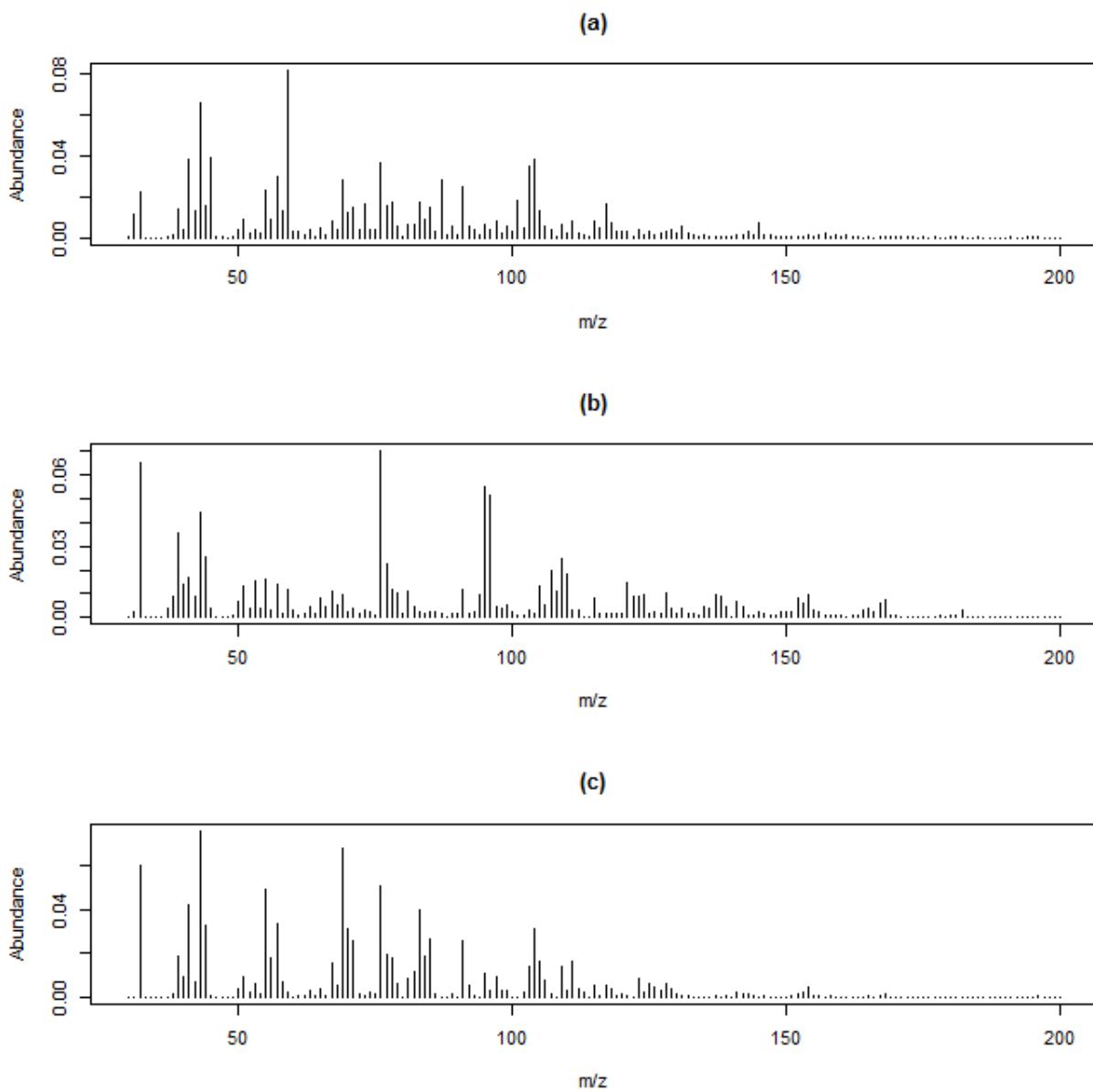


Figure 13: Total ion spectra for substrates 1 – 3 used in simulation studies are shown as a – c respectively. See Table 1 for identification of the substrates.

Table 1: Ignitable liquid and pyrolyzed substrate TIS used in simulations studies. Each TIS is referenced in the text by the identifier (ID) given in the first column.

ID	Ignitable Liquids
Gas	BP regular unleaded gasoline, 25% weathered
ISO	Sunnyside Odorless Paint Thinner
MPD	Ace Premium Quality Charcoal Lighter
Pyrolyzed Substrates	
Sub1	Carpet Padding
Sub2	Clear White Oak Laminate
Sub3	Olefin Nylon Blend Carpet
Sub4	Roofing shingles
Sub5	Yellow Pine
Sub6	PET polyester carpet and padding
Sub7	Street smart boots
Sub8	Vinyl floor
Sub9	Polyurethane foam mattress
Sub10	Cardboard box

Table 2: Composition of simulated data sets for testing TFA.

Test	IL	Pct. each	IL	Sub	Pct. Sub	Pct.	No. spectra	No. Factors	No. PC (MAD)
1a	Gas	1 - 15	Sub1 - 5	80 - 94	5	10	6	6	6
1b	Gas	1 - 5	Sub1 - 5	90 - 94	5	10	6	6	6
2a	ISO + MPD	1 - 15	Sub1 - 5	65 - 93	5	10	7	7	7
2b	ISO + MPD	1 - 5	Sub1 - 5	85 - 93	5	10	7	7	8
3a	Gas	1 - 5	Sub1 - 10	90 - 94	5	10	11	1	1
3b	Gas	1 - 30	Sub1 - 10	65 - 94	5	10	11	1	1
3c	Gas	1 - 50	Sub1 - 10	45 - 94	5	10	11	1	1
3d	Gas	1 - 70	Sub1 - 10	25 - 94	5	10	11	2	2

Samples 3a – 3d contained a single ignitable liquid, Gas, combined with 10 different substrates, in order to give a data matrix of rank 11; however, the number of spectra concatenated to produce the data matrix was only 10, one less than the rank of the matrix. In these cases, the problem is under determined, as it is required that the number of samples (rows in the matrix) exceed the rank of the matrix. It is nonetheless important to examine the behavior of this type of system because in a real fire debris analysis, it is not possible to determine if the number of spectra collected exceed the number of contributing components. In these tests, the number of principal factors (or components) predicted by DRMAD was limited to one, except in example 3d where the ignitable liquids was present in 1 – 70%. The results for 3a – 3d represent a potential challenge for the method; the number of samples must be greater than or equal to the

number of contributing components. Therefore, from a methodology perspective, it is important to maximize the number of samples analyzed. However, the predictive performance of the method has yet to be examined, and will be discussed in the following paragraphs.

The identification of an appropriate number of factors in the simulation data is not necessarily surprising. The more challenging aspects of the tests are to determine if the appropriate ignitable liquids can be identified through target transformations. The library of target factors for the simulation studies was comprised of the total ion spectra from samples of the ASTM classes Gas (23 samples), ISO (33 samples) and MPD (93 samples). Resulting vectors from target transformations of the eigenvectors from 1a with the 23 Gas test vectors gave correlations ranging from $r = 0.621$ ($n = 171$, 90% evaporated Gas) to $r = 0.930$ ($n = 171$, un-evaporated gasoline). These results are based on the retention of six eigenvectors, as shown in Table 2. Target transformation of spectra from the ISO and MPD classes gave poorer correlations between the test and predicted vectors. Class-conditional probability distributions for test 1a are shown in Figure 14a. The distributions are calculated based on the correlation coefficients for each class with bandwidths; $h_{Gas} = 0.035$, $h_{ISO} = 0.013$ and $h_{MPD} = 0.007$. The distributions reflect the overall larger correlations for library spectra from the Gas class. The posterior probabilities, $P(\omega_i|r_{LL},1)$, for the Gas, ISO and MPD classes were calculated to be 0.94, 0 and 0.06 respectively at $r_{LL} = 0.9$. The posterior probabilities are graphed as a function of r_{LL} in Figure 14b.

Posterior probabilities for tests 1 – 3, calculated at $r_{LL} = 0.9$, are given in Table 3. Results for test 1b, wherein the Gas contribution was in the 1 – 5% range, were nearly identical to the results for test 1a. Importantly, this very small contribution from Gas results in samples that are likely far too weak to call positive for ignitable liquid in real fire debris analysis. In tests 2a and 2b, the posterior probabilities approach 0.5 for both ISO and MPD, reflecting the presence of both liquids in the test data. The class-conditional probabilities for test 2a are shown in Figure 15. The graph in Figure 15 demonstrates increased correlations for the ISO and MPD classes, and decrease in correlations for the Gas class. Notably, tests 3a – 3d gave very high posterior probabilities for the Gas class, even though in these tests the problem was under determined and the number of principal factors was significantly under estimated. In these tests, the correlations for all three classes were significantly decreased; however, the ISO and MPD class correlations were decreased to the point that the integrated areas in the range from $r_{LL} - 1$ were below the α value and the posterior probabilities were set equal to zero. Although the correlations for the Gas class were low, the integrated areas in the range from $r_{LL} - 1$ were greater than α . Since Gas was the only class with a non-zero posterior probability and all probabilities must sum to one, $P(\omega_{Gas}|r_{LL},1)$ was assigned a value of one.

Table 3. Posterior probabilities $P(\omega_i|[0.9,1])$ from tests 1 – 3 in Table 2. Posterior probabilities less than 0.001 are reported as 0 and the posterior probability for analyte class in each test is indicated in bold text.

Test	1a	1b	2a	2b	3a	3b	3c	3d
IL	Gas	Gas	ISO + MPD	ISO + MPD	Gas	Gas	Gas	Gas
Pct. each								
IL	1 - 15	1 - 5	1 - 15	1 - 5	1 - 5	1 - 30	1 - 50	1 - 70
Gas	0.937	0.929	0.103	0.002	1.000	1.000	1.000	0.994
ISO	0.000	0.000	0.530	0.484	0.000	0.000	0.000	0.000
MPD	0.063	0.071	0.367	0.514	0.000	0.000	0.000	0.006

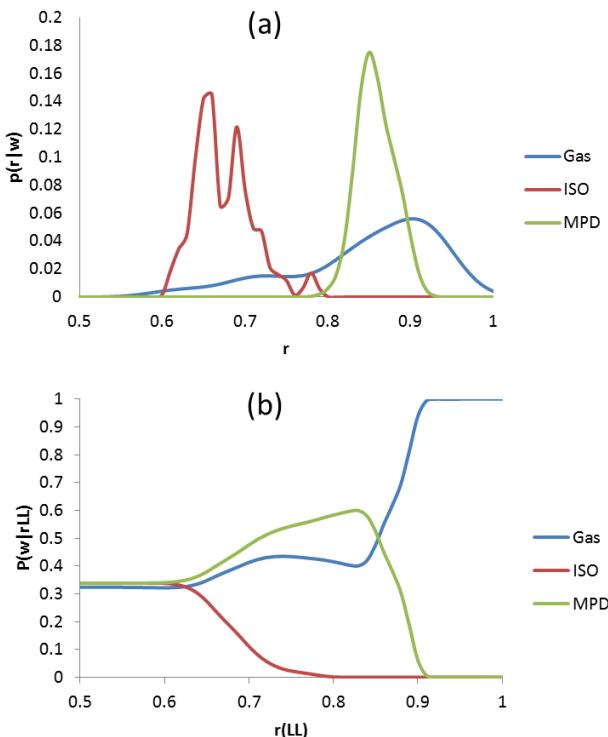


Figure 14. (a) Kernel probability distribution functions for test 1a (Table 2), (b) posterior probabilities for test 1a as a function of r_{LL} .

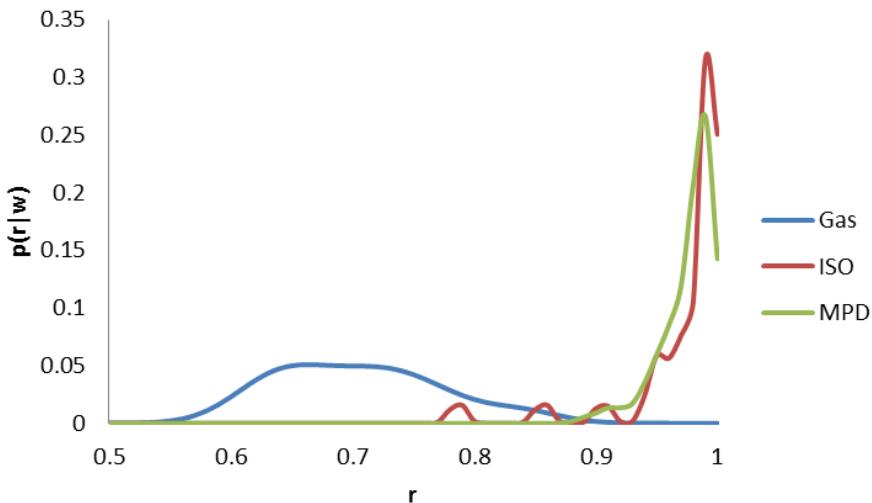


Figure15. Kernel probability distribution functions for test 2a (Table 2).

Laboratory-scale burns: A series of 20 laboratory-scale burn tests were conducted as described in the experimental section. The ASTM classes, ILRC sample reference numbers (SRN), and the substrates included in each set of burns are given in Table 4. Representative liquids from all ASTM classes were used in laboratory burns. The numbers and types of substrates were varied among the different test burns, as were the volumes of ignitable liquid and the number of burn data sets compiled into the data matrix for each test. Table 5 lists the range of volumes of ignitable liquid used in each test, the number of burn samples compiled for each test, the number of samples and number of substrates for each test, the ignitable liquid used in the burn and the posterior probabilities calculated at $r_{LL} = 0.95$. The ignitable liquids in the library used for target transformation comprised the ASTM classes and specifying light (L), medium (M) and heavy (H) sub-classifications for the petroleum distillate (PD) and aromatic (AR) classes. The library of 358 TIS contained 25 Gas, 3 HAR, 47 HPD, 34 ISO, 7 LAR, 23 LPD, 17 MAR, 95 MPD, 17 NA, 16 NP and 71 OXY ignitable liquids. Correct classification was optimized relative to r_{LL} by assigning each sample to the class having the highest posterior probability. The optimized correct classification rate was obtained at $r_{LL} = 0.94 - 0.97$, leading to a value of 0.95 used in Table 5. The sensitivity level, α , for the results in Table 5 was set at 0.05, corresponding to a 95% confidence level. The maximum posterior probability is shown in bold text for each test in Table 5. All $P(\omega_i|[r_{LL},1])$ values less than 0.001 are shown as zero in Table 5.

All of the Gas test burns, except one (test 8), were calculated to have a maximum posterior probability corresponding to the Gas class. In the one case where the largest posterior probability did not correspond to Gas, the rank of the data matrix (six substrates and gasoline) exceeded the number of burn samples analyzed (five samples analyzed). The volume of gasoline used in each of the burns in test 8 was only 0.025 mL. In this case the amount of gasoline residue was too small for successful classification, and gasoline pattern could not be discerned by visual inspection of the TIC. Other tests where the maximum calculated posterior probability did not correspond to the class of ignitable liquid used in the burn included both ISO samples and one OXY sample. One of the ISO samples was calculated to have the highest posterior probability

corresponding to the OXY class, while the other corresponded to the NA class. Similarities between the ISO and NA classes, see Figure 1, aid in understanding these results, where posterior probabilities of 0.436 and 0.520 were calculated for the ISO and NA classes respectively. One OXY test was calculated to have the largest posterior probability corresponding to the Gas class. No explanation can be offered for this result. The largest posterior probability for 16 out of 20 samples (80%) corresponded to the ignitable liquid used in the test.

Table 4. Laboratory-scale burn sample data.

Test No.	IL Class	ILRC SRN	Substrates
1	Gas	116	polyester carpet and padding
2	Gas	116	yellow pine, fiberglass insulation, polyurethane foam mattress, vinyl flooring, PET polyester carpet, and carpet padding
3	Gas	116	yellow pine, fiberglass insulation, polyurethane foam mattress, vinyl flooring, PET polyester carpet, and carpet padding
4	Gas	116	yellow pine, fiberglass insulation, polyurethane foam mattress, vinyl flooring, PET polyester carpet, and carpet padding
5	Gas	116	poplar, fiberglass insulation, olefin/nylon carpet
6	Gas	258	oak, nylon rope, and roofing shingles
7	Gas	385	teak, cotton tshirt, cork tiles
8	Gas	116	corkboard and vinyl
9	Gas	116	PET polyester carpet, carpet padding and cotton
10	HPD	206	polyester carpet and padding
11	ISO	87	polyester carpet and padding
12	ISO	87	Douglas Fir
13	AR	59	polyester carpet and padding
14	MPD	30	polyester carpet and padding
15	MPD	30	Douglas Fir
16	NA	241	polyester carpet and padding
17	NP	243	polyester carpet and padding
18	OXY	174	polyester carpet and padding
19	OXY	248	olefin carpet, cherry wood, vinyl
20	OXY	407	olefin carpet, oak, fiberglass insulation

Table 5: Posterior probabilities $P(\omega_i|[0.95,1])$ from laboratory-scale burn tests using the ignitable liquid (IL) classes and volumes reported. The number of samples analyzed and number of substrates (Sub.) used in each burn are also given. Posterior probabilities less than 0.001 are reported as 0 and the highest posterior probability for each burn is indicated in bold text.

No.	1	2	3	4	5	6	7
Vol. (mL)	0.1 - 2	1	0.2	0.05	0.2	0.2 - 2	0.2 - 2
Samples	9	9	6	6	5	5	5
Sub.	2	6	6	6	7	3	3
IL	Gas						
Gas	0.766	0.933	0.934	0.884	0.906	0.958	0.995
HAR	0.007	0	0	0	0	0	0
HPD	0	0	0	0	0	0	0
ISO	0	0	0	0	0	0	0
LAR	0.006	0	0	0	0.008	0	0
LPD	0	0	0	0	0	0	0
MAR	0.222	0.066	0.065	0.047	0.081	0.036	0
MPD	0	0	0	0	0	0	0
NA	0	0	0	0	0	0	0
NP	0	0	0	0	0	0	0
OXY	0	0	0.001	0.069	0.005	0.006	0.005

No.	8	9	10	11	12	13	14
Vol. (mL)	0.025	0.05	0.11 - 2	0.1 - 2	0.2 - 2	0.1 - 2	0.1 - 2
Samples	5	5	9	9	5	9	9
Sub.	6	5	2	2	1	2	2
IL	Gas	Gas	HPD	ISO	ISO	LAR	MPD
Gas	0.123	0.907	0.04	0	0	0.034	0
HAR	0	0	0.243	0	0	0	0
HPD	0	0	0.64	0	0.001	0	0.266
ISO	0	0	0	0	0.436	0	0.006
LAR	0	0	0	0	0	0.899	0
LPD	0	0	0	0.04	0.024	0.011	0.097
MAR	0.012	0.076	0	0	0	0.001	0
MPD	0	0	0	0	0.003	0	0.372
NA	0	0	0	0	0.52	0	0
NP	0	0	0.061	0	0	0	0.232
OXY	0.865	0.017	0.016	0.96	0.017	0.055	0.028

Table 5 continued.

No.	15	16	17	18	19	20
Vol. (mL)	0.2 - 2	0.1 - 2	0.1 - 2	0.1 - 2	0.2 - 2	0.2 - 2
Samples	5	9	9	9	5	5
Sub.	1	2	2	2	3	3
IL	MPD	NA	NP	OXY	OXY	OXY
Gas	0	0	0	0	0.804	0
HAR	0	0	0	0	0	0
HPD	0.274	0.004	0.014	0	0	0
ISO	0.006	0.053	0	0	0	0
LAR	0	0	0	0	0	0
LPD	0.067	0	0.016	0.059	0	0
MAR	0	0	0	0	0	0
MPD	0.341	0	0.117	0.228	0	0
NA	0	0.938	0	0	0	0
NP	0.288	0	0.836	0.311	0	0
OXY	0.024	0.005	0.018	0.402	0.197	1

When the laboratory-scale burn data was analyzed with $\alpha = 0.01$, posterior probabilities were not calculated for three of the tests (9, 11 and 18), because none of the class-conditional probabilities had integrated areas greater than α . Fourteen of the remaining 17 samples were correctly calculated to have a maximum posterior probability corresponding to the ignitable liquid used in the burn (82% correct).

Large-Scale Burn Studies: As described in the methods section of this report, 12 large-scale burns were conducted in Konex boxes that had been built out and furnished with common household items. In four of the burns, two ignitable liquids were used, whereas in the remaining eight burns a single ignitable liquid was used. Table 6 lists the number of samples taken from each burn, the ignitable liquid(s) used in the burn and the posterior probabilities for the ASTM classes and subclasses, as described for Table 5. A total of 500 mL of each liquid was used in each burn. The temperatures in each burn varied, but generally reached the 700 - 800°C range before the fire was extinguished. All of the containers went to flashover and were extinguished with only water. After the containers had cooled, evidence markers were placed where samples were to be collected and the locations documented. A total of 11 – 13 samples were taken from both on and off the pour trail for each burn. Assigning the ignitable liquid to the class with the highest posterior probability, 83% of the samples (10 of 12 burns) were assigned to either the single liquid used, or one of the two liquids used in the test, a result very similar to the results from the laboratory-scale burns.

Table 6: Posterior probabilities $P(\omega_i|[0.95,1])$ from large-scale burns using 500 mL of each of the ignitable liquid (IL) classes indicated. The number of post-burn samples analyzed are also given. Posterior probabilities less than 0.001 are reported as 0 and the highest posterior probability for each burn is indicated in bold text.

Container Samples	1	2	3	4	5	6
IL	Gas	Gas+MPD	MPD	OXY	Gas	Gas
Gas	0.925	0.017	0.032	0.254	0.511	0.895
HAR	0	0	0	0	0	0
HPD	0.002	0.418	0.248	0.037	0	0
ISO	0	0	0.009	0	0	0
LAR	0	0	0	0	0.275	0
LPD	0.017	0	0.093	0	0.041	0
MAR	0	0	0.043	0	0.084	0.049
MPD	0	0.511	0.302	0.006	0	0
NA	0	0	0	0	0	0
NP	0	0	0.247	0	0	0
OXY	0.056	0.054	0.028	0.703	0.089	0.057

Container Samples	7	8	9	10	11	12
IL	Gas + MPD	Gas + MPD	Gas + HPD	LPD	NP	ISO
Gas	0.038	0.001	0.004	0.012	0.002	0.3
HAR	0	0	0	0	0	0
HPD	0.248	0	0.724	0.021	0.159	0.02
ISO	0.003	0	0	0	0	0.359
LAR	0	0	0	0	0	0.002
LPD	0.063	0	0	0.442	0.03	0.011
MAR	0.019	0	0	0	0	0.058
MPD	0.3	0	0.173	0.453	0.331	0.047
NA	0	0	0	0	0	0.158
NP	0.31	0	0.027	0	0.455	0
OXY	0.019	0.998	0.073	0.072	0.024	0.046

Container 7 was burned with Gas and MPD; however, the highest posterior probabilities, 0.31 and 0.300, were calculated for the NP and MPD classes respectively. Container 8 was burned with Gas and MPD, but the largest posterior probability was calculated for the OXY class.

Container 2 was burned with Gas and MPD, and the posterior probability for MPD and Gas were calculated to be 0.511 and 0.17 respectively.

When the calculations are repeated at $r_{LL} = 0.95$ with a significance level of 0.01, only Container 4 was not classified due to none of the classes having an integrated area under the class-conditional probability curves greater than α . In nine of the 11 remaining containers (82%), the highest calculated posterior probability corresponded to the ASTM class of the ignitable liquid (or one of the two ignitable liquids) used to initiate the fire.

IV. Conclusions

Discussion of findings: The goal of the research conducted under this grant was to develop a chemometric method of data analysis that would facilitate the identification of GC-MS patterns associated with ignitable liquid classes designated under ASTM E 1618. The challenge of obtaining this goal is increased by the presence of a highly variable and unpredictable background signature arising from the pyrolysis products of building materials and furnishings. In order to attain this goal, the first objective of the research was to develop a data analysis method that would classify ignitable liquid residue in the presence of background contamination levels.

The first objective was achieved by applying the technique of target factor analysis to a matrix of data consisting of the total ion spectra from a series of samples taken from a fire scene. Target factors were provided by a library of total ion spectra derived from the analysis of commercially available ignitable liquids. The library of 358 TIS contained 25 Gas, 27 HAR, 47 HPD, 34 ISO, 7 LAR, 23 LPD, 17 MAR, 95 MPD, 17 NA, 16 NP and 71 OXY ignitable liquids. The TIS were calculated for ignitable liquids in the ILRC, which contains GC-MS data for commercially available materials and each entry has been given a consensus ASTM classification by the TWGFEILRC Committee. The target factors in each class were compared to the least-squares best fit projected vectors by Pearson product moment correlation coefficients. Results from the simulated data set studies demonstrated that target factor analysis of TIS from multiple samples would produce a distribution of correlation coefficients which was in general higher for the class(es) of ignitable liquids contained in the samples. Each sample contained a number of background contaminants that were derived from the NCFS Substrate database. The results further demonstrated that when ignitable liquids from two classes contributed to a sample, the correlation coefficient distributions for both class would be higher on average than the correlation coefficients for the non-contributing class.

The second objective of the study was to obtain statistical probabilities for the class assignments. This objective was achieved by analyzing the correlation coefficients for each class using a modified version of Bayesian decision theory. The Bayesian soft classification method combined with target factor analysis was tested on the analysis of simulated data and fire debris data. Unlike most classification methods, the TFA-Bayesian method is designed to work in the

presence of interfering background spectral contributions. The method provides the analyst with a posterior probability of ASTM class assignment and the associated risk, which should aid in the assignment of difficult samples. Applying the Bayesian method to the results from the simulated data samples provided high posterior probabilities for gasoline in the single ignitable liquid samples (Tests 1 and 3) and posterior probabilities approaching 50% for ISO and MPD classes in Test 2, which contained both ignitable liquids. Application of the TFA-Bayesian approach to the laboratory burn data resulted in four of the 20 samples where the maximum calculated posterior probability did not match the class of ignitable liquid used in the burn (i.e., 80% correct prediction of the correct class based on the largest posterior probability). Similarly, when the TFA-Bayesian method was applied to the large-scale burns, approximately 83% correct classification was obtained based on assigning the ignitable liquid to the sample with the largest posterior probability. In both the laboratory- and large-scale burns, when the analysis was performed at a sensitivity level of 0.05, posterior probabilities were classified for all of the burns; however, when the calculation was repeated at $\alpha = 0.01$ for the large-scale burns, posterior probabilities were not calculated for any of the classes in one burn sample.

While the overall correct classification rates are lower than expected, several factors must be kept in perspective. The large-scale burns were all taken to flashover and there is absolutely no control over the amount of ignitable liquid residue remaining for analysis. Similarly, the degree of weathering is not controlled. In these regards, these data are very representative of real fire debris samples. This is, to our knowledge, the first multivariate classification model designed to work explicitly in the presence of an unknown and unspecified background signature. From this perspective, a correct classification rate of 83% is very encouraging. Nonetheless, there is room for improvement, as discussed in the following sections. Finally, the addition of the sensitivity criterion makes the method conservative in that by adjusting r_{LL} and α , the method can be made to only calculate posterior probabilities for classes that have a reasonably high correlation for a significant portion of the library population of a given class. The disadvantages of the method are that it requires the analysis of data from multiple samples taken from a single fire scene, rather than a single sample, and the number of samples should exceed the “chemical rank” of the data, a condition that may be difficult to ensure in real samples.

Implications for policy and practice: Fire debris analysts are currently allowed to opine that a sample is positive for ignitable liquid residue without the requirement to offer a known error rate for this assessment. It is reasonable to assume that at some point the admission of expert opinions in fire debris analysis is likely to meet a Daubert challenge based on the lack of a known error rate. This study represents one approach to obtaining error rates. Other approaches are also available from NIJ-sponsored research. In addition, this method provides a set of probabilities that the analyst may take into consideration when determining whether to call a sample positive or negative for ignitable liquid residue. Samples containing a significant signature for an ignitable liquid residue will result in large probabilities for the appropriate ignitable liquid class, and the probabilities may be helpful in providing statistical support for the analyst. We see perhaps a larger benefit coming when the ignitable liquid signature is weaker. The TFA-Bayesian method can identify the presence of an ignitable liquid at low levels. Simulated data has shown that gasoline can be identified as low as 1 – 5% contribution.

Implications for further research: As described in the previous section, the TFA-Bayesian method was successful in giving correct classification rates in the presence of pyrolysis interference which were as high as the rates for correct classification of interference-free ignitable liquids based on similarity of the TIS. Improving on the 80 – 85% correct classification may be achievable based on an alternative representation of the GC-MS data set. This is the topic for future research. In addition, the TFA-Bayesian method described in this report is applicable in many areas of analytical chemistry, sensing and forensic analysis. Applications of the TFA-Bayesian approach to additional areas of forensic science constitute future research opportunities. Further research is also warranted on any method that allows the determination of error rates in fire debris analysis or any other area of forensic science where error rates do not currently exist.

V. References

-
- [1] “Webster’s Third New International Dictionary of the English Language Unabridged” P. B. Gove Ed., Merriam-Webster Inc., Springfield, MA, USA, 1986.
 - [2] ASTM (2010) Test Method for ignitable liquid residues in extracts from fire debris samples by gas chromatography-mass spectrometry (E1618-10). American Society for Testing and Materials.,” ASTM International, West Conshohocken, PA, 2006.
 - [3] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification, John Wiley & Sons, Inc., New York, 2001.
 - [4] M. E. Sigman, M. R. Williams, Application of Covariance Mapping to the Analysis of Complex Mixtures by GC-MS, Anal. Chem. 78 (2006) 1713 - 1718.
 - [5] M. E. Sigman, M. R. Williams, R. G. Ivy, Comparative analysis of gasolines by covariance mapping and gas chromatography – mass spectrometry, Anal. Chem. 79 (2007) 3462-3468.
 - [6] L. Frasinski, J. K. Codling, P. A. Hatherly Covariance mapping: A correlation method applied to multiphoton multiple ionization, Science 246 (1989) 1029-1031.
 - [7] P. Jukes, A. Buxey, A. B. Jones, A. Stace, Covariance images of the primary response from rare gas cluster ions to photoexcitation, J. Chem. Phys. 106 (1997) 1367-1372.
 - [8] M. Foltin,; Stueber, G. J.; Bernstein, E. R., Gas Phase Dynamics and Structure: Spectroscopy, Molecular Interactions, Scattering, and Photochemistry - Dynamics of neutral cluster growth and cluster ion fragmentation for toluene/water, aniline/argon, and 4-fluorostyrene/argon clusters: Covariance mapping of the mass spectral data, J. Chem. Phys. 109 (1998) 4342-4360.

-
- [9] A. B.Feldman, M .Antoine, J. S. Lin, P. A. Demirev, Covariance mapping in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* 17 (2003) 991-995.
- [10] M. J.Van Stipkonk, E. A. Schweikert, M. A. Park, Coincidence measurements in mass spectrometry , *J. Mass Spectrom.* 32 (1997) 1151-1161.
- [11] M. E.Sigman, M. R. Williams, J.A. Castelbuono, J. A. Colca, C. D. Clark, Ignitable Liquid Classification and Identification Using the Summed-Ion Mass Spectrum, *Instrumentation Science and Technology*, 36 (2008) 375-393.
- [12] P. M. L. Sandercock, E. DuPasquier, Chemical fingerprinting of unevaporated automotive gasoline samples, *Forensic Sci. Int.* 134 (2003) 1-10.
- [13] P. M. L. Sandercock, E. DuPasquier, Chemical fingerprinting of gasoline: 2. Comparison of unevaporated and evaporated automotive gasoline samples, *Forensic Sci. Int.* 140 (2004) 43-59.
- [14] P.M.L. Sandercock, E. DuPasquier, Chemical fingerprinting of gasoline: Part 3. Comparison of unevaporated automotive gasoline samples from Australia and New Zealand, *Forensic Sci. Int.* 140 (2004) 71-77.
- [15] E. E. Waddell, E. T. Song, C. N. Rinke, M. r. Williams, M. E. Sigman, “Progress toward te Determination of Correct Classification Rates in Fire Debris Analysis.” *J. Forensic Sci.*, 2012, in press.
- [16] J. M. Baernkopf, V. L. McGuffin, R. W. Smith, Association of ignitable liquid residues to neat ignitable liquids in the presence of matrix interferences using chemometric procedures, *J. Forensic Sci.* 56 (2011) 70-81.
- [17] B.Tan, J. K. Hardy, R. E. Snavely, Accelerant classification by gas chromatography/mass spectrometry and multivariate pattern recognition, *Anal. Chim. Acta* 422 (2000) 37-46.
- [18] Y. Lu, P. Chen, P. B. Harrington, Comparison of differential mobility spectrometry and mass spectrometry for gas chromatographic detection of ignitable liquids from fire debris using projected difference resolution, *Anal Bioanal. Chem.* 394 (2009) 2061 – 2067.
- [19] W. N. S. MatDesa, N. N. Daéid, D. Ismail, K. Savage, Application of unsupervised chemometric analysis and self-organizing feature map (SOFM) for the classification of lighter fuels, *Anal. Chem.* 82 (2010) 6395 – 6400.
- [20] ASTM (2007) Standard practice for separation of ignitable liquid residues from fire debris samples by passive headspace concentration with activated charcoal (E1412-07). American Society for Testing and Materials.,” ASTM International, West Conshohocken, PA, 2006.

-
- [21] K. Varmuza, P. Filzmoser “Introduction to Multivariate Statistical Analysis in Chemometrics”, CRC Press, Taylor & Francis Group, Boca Raton, FL, 2009.
- [22] E. R. Malinowski, Determination of rank by median absolute deviation (DRMAD): a simple method for determining the number of principal factors responsible for a data matrix, *J. Chemometr.* 29 (2009) 1 – 6.
- [23] B.W. Silverman, Density estimation for Statistics and Data Analysis, Chapman and Hall, New York, 1986.
- [24] J.R. Eastman, R. M. Laney, Bayesian Soft Classification for Sub-Pixel Analysis: A Critical Evaluation, *Photogrammetric Engineering & Remote Sensing* 68 (2002) 1149 – 1154.

VI. Dissemination of Research Findings

Publications and Thesis:

1. Williams, M. R.; Sigman*, M. E.; Lewis, J.; McHugh Pitan, K. “Combined Target Factor Analysis and Bayesian Soft-Classification of Interference-Contaminated Samples: Forensic Fire Debris Analysis,” *Forensic Sci. Int.*(2012) 222, 373-386.
2. DETERMINING THE PRESENCE OF AN IGNITABLE LIQUID RESIDUE IN FIRE DEBRIS SAMPLES UTILIZING TARGET FACTOR ANALYSIS by KELLY M. MCHUGH, A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Forensic Science in the College of Sciences at the University of Central Florida, Orlando, Florida, Spring 2010.
3. THE APPLICATION OF CHEMOMETRICS TO THE DETECTION AND CLASSIFICATION OF IGNITABLE LIQUIDS IN FIRE DEBRIS USING THE TOTAL ION SPECTRUM, by JENNIFER N. LEWIS, A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Forensic Science in the College of Sciences at the University of Central Florida, Orlando, Florida, Fall 2011.

Patents:

1. Sigman, Michael E, Williams, Marry R., "Classification of a Complex Mixture, or Components of a Complex Mixture, by Combining Target Factor Analysis with a Bayesian Classifier," Application No.: 13/088,980, Filing Date: April 18, 2011.

Presentations:

1. McHugh, K; Williams, MR; Sigman ME "Automated Searching of an Ignitable Liquids Library of Summed Ion Spectra by Target Factor Analysis" 62nd Annual Meeting of the American Academy of Forensic Sciences, Feb. 22 – 27, 2010, Seattle, WA.
2. Sigman, ME "Application of Chemometrics and GC-MS Analysis for the Identification of Traces of Ignitable Liquids in Fire Debris Samples (2008-DN-BX-K069)" The NIJ Conference, June 14 – 16, 2010, Arlington, VA.
3. Sigman ME, Williams MR "Application of Target Factor Analysis to the Classification of Ignitable Liquids from Fire Debris" 37th Federation of Analytical Chemistry and Spectroscopy Society Meeting, Oct. 17 – 21, Raleigh, NC.
4. Caitlin Rinke, Christopher Brown, Matthieu Baudelot, Martin Richardson, Michael E. Sigman "A New Paradigm for Substrate Independent Discrimination of Organic and Explosive Materials by Target Factor Analysis of Molecular Optical Signatures", Chemical and Biological Defense Science and Technology Conference (CBD S&T), Orlando, FL, Nov. 15, 2010.
5. Lewis, J.; Williams, M. R.; Sigman, M. E. "Identification of Ignitable Liquid Classes by Target Factor Analysis of Total Ion Spectra", 63rd Annual Meeting of the American Academy of Forensic Sciences, Feb. 21 – 26, 2011, Chicago, IL.
6. Williams, M. R.; Sigman, M. E. "Advancing Fire Debris Analysis through Chemometrics: An Overview of Research at the National Center for Forensic Science;" 63rd Annual Meeting of the American Academy of Forensic Sciences, Feb. 21 – 26, 2011, Chicago, IL.
7. Williams, M. R.; Lewis, J.; Waddell, E. E.; Rinke, C. N. "Advancing Fire Debris Analysis Through Chemometrics"; Pittsburg Conference and Exposition, March 13 – 18, 2011, Atlanta, GA.
8. Sigman, ME "Ignitable Liquids Reference Collection Panel: Present Research", NIJ Conference, Washington D.C., June 21, 2011.
9. Michael E. Sigman, "Matrix Independent Analyte Identification and Classification," University of Cadiz, Cadiz, Spain, May 4, 2012.

VII. Appendix

Representation of exported 3D data from Chemstation.

Scan	30	31	32	33	34	35	36	37	38	39	40	41	42	43	...	200
1	0	0	3067	0	0	0	0	0	1312	192	0	433	0	294		0
2	0	0	3278	0	0	0	0	0	1243	258	155	331	0	272		0
3	0	0	3043	0	0	0	0	0	1294	239	0	306	0	224		0
4	0	0	3027	0	0	0	0	0	1160	279	0	429	0	234		0
5	0	0	2798	0	0	0	0	0	1134	717	176	1375	476	315		0
6	0	0	2719	0	0	0	0	0	1238	1732	430	3580	1215	445		0
7	0	0	2645	0	0	0	0	0	1240	1759	461	3489	1310	572		0
8	0	0	2789	0	0	0	0	0	1120	1445	390	3088	1261	464		0
9	0	202	2595	0	0	0	0	0	1194	1265	370	2555	944	515		0
10	0	331	2744	0	0	0	0	0	941	739	193	1624	561	422		0
:																
3847	0	0	773	0	0	0	0	0	0	0	0	0	0	0		0

Calculate the total ion spectrum from exported 3D data.

Scan	30	31	32	33	34	35	36	37	38	39	40	41	42	43	...	200
TIS	692	183792	3413356	0	0	0	684	39627	151406	847525	159732	1622394	450391	3023710		0

Normalize the total ion spectrum

1. Sum the ion (m/z) intensity for the entire spectrum 84295880
 2. Divide each ion (m/z) intensity by the total ion intensity of the spectrum
- | Scan | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | ... | 200 |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|-------|
| normalized TIS | 0.000 | 0.002 | 0.040 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.010 | 0.002 | 0.019 | 0.005 | 0.036 | | 0.000 |

Representative data set of normalized TIS for nine samples.

m/z	30	31	32	33	34	35	36	37	38	39	40	41	42	43	...	200
Sample 1	0.000	0.002	0.040	0.000	0.000	0.000	0.000	0.000	0.002	0.010	0.002	0.019	0.005	0.036		0.000
Sample 2	0.000	0.002	0.033	0.000	0.000	0.000	0.000	0.001	0.002	0.010	0.002	0.019	0.005	0.033		0.000
Sample 3	0.000	0.003	0.037	0.000	0.000	0.000	0.000	0.000	0.002	0.009	0.002	0.021	0.006	0.030		0.000
Sample 4	0.000	0.001	0.034	0.000	0.000	0.000	0.000	0.000	0.001	0.008	0.002	0.017	0.003	0.024		0.000
Sample 5	0.000	0.002	0.033	0.000	0.000	0.000	0.000	0.001	0.002	0.010	0.002	0.021	0.006	0.033		0.000
Sample 6	0.000	0.001	0.037	0.000	0.000	0.000	0.000	0.001	0.002	0.013	0.003	0.028	0.008	0.045		0.000
Sample 7	0.000	0.000	0.034	0.000	0.000	0.000	0.000	0.001	0.002	0.013	0.002	0.024	0.006	0.043		0.000
Sample 8	0.000	0.001	0.029	0.000	0.000	0.000	0.000	0.001	0.002	0.009	0.002	0.020	0.005	0.030		0.000
Sample 9	0.000	0.001	0.029	0.000	0.000	0.000	0.000	0.001	0.002	0.009	0.002	0.019	0.005	0.027		0.000