2019

Chemometric Applications in Fire Debris Analysis: Likelihood Ratios from Naive Bayes and Frequency of Component and Pyrolysis Product Occurrence

Anuradha Akmeemana
University of Central Florida

CHEMOMETRIC APPLICATIONS IN FIRE DEBRIS ANALYSIS: LIKELIHOOD RATIOS FROM NAÏVE BAYES AND FREQUENCY OF COMPONENT AND PYROLYSIS PRODUCT OCCURRENCE

by

ANURADHA GAYATHRI AKMEEMANA
BSc. Open University of Sri Lanka, 2010
M.S. Eastern Illinois University, 2014
M.S.F.S. University of Central Florida, 2018

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Department of Chemistry
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2019

Major Professor: Michael E. Sigman

ABSTRACT

One of the major challenges in fire investigation is the determination of the cause of fire. The fire can be accidental or intentional. The determination of ignitable liquid residue (ILR) from fire debris helps the process and this process is called fire debris analysis in forensic science. This is one of the most complex areas in the field of forensics because of the evaporation of the ILR from the debris and the interferences of the substrate matrix with the ILR if present. In the present, the final decisions in fire debris analysis are based on categorical statements and it only represents the qualitative but not the quantitative value of the data. The likelihood ratio approach is one of the most widely used methods in forensic science in expressing the evidentiary value.

The purpose of this research is to introduce the likelihood ratios calculated by the Naïve Bayes approach. The data for this work was obtained by the Substrate and ILRC Databases from the National Center for Forensic Science. This project also contributed to the expansion of the Substrate Database by adding 1500 new substrate burn data records. The compounds identified from ignitable liquids and substrates were used to calculate the frequency of occurrences of the compounds in substrates and ignitable liquids. The presence or absence of the compounds was determined by the probabilities calculated by logistic regression. These frequencies of occurrences were used in the calculation of Naïve Bayes log likelihood ratios. The application, performance and validation of these models are discussed in this dissertation. These calculated log-likelihood ratios indicated that this method provides high evidentiary values in the classification of fire debris as positive for ILR in most cases but provided low evidentiary values in some other instances.

I dedicate this work to my family, friends and everyone who supported me throughout this journey

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my PI Dr. Michael E Sigman for his mentorship and support throughout my time at UCF and NCFS. It is impossible to achieve these goals without his help and guidance and I am very fortunate to be selected as a student in his research group. Also, I would like to thank National Institute of Justice for providing the funds for the research. I am very thankful to Mrs. Mary Williams who helped me with all my questions and supporting me during my time at NCFS. I would like to express my heartfelt appreciation to my dissertation committee Dr. Candice Bridge, Dr. Matthew Baudelet, Dr. Andres Campiglia and Dr. Liqiang Ni.

It is difficult to stay away from home for four years without visiting them and thanks to my friends at NCFS I did not miss home excessively. I would like to thank my colleagues and friends at NCFS, Dr. Dana-Marie Dennis, Dr. Jessica Chappell, Dr. Mark Maric, Dr. Mauro Martinez, Alyssa Allen, Richard Coulson, Jessica Kindell, Danielle Green, Bryan McCollough, Emily Lennert, Yasmine Maustafa, Michelle Cobally, Molly Terry, Quashanna Janae, Jessica Sprague, Brooke Burmington, David Funes, Kandyss Najjar, Kaitland Jones, Frances Whitehead, Nicholas Thurn and Taylor Wood for being there for me. Also, I would like to thank Mrs. Judith Stout and Mrs. Erika Remley. Finally, my mom, dad and my sister who encouraged me to cross oceans to achieve my goals.

TABLE OF CONTENT

LIST OF FIGURES ix

LIST OF TABLES xv

CHAPTER 1: DIFFERENT CHEMOMETRIC MODELS USED IN FIRE DEBRIS ANALYSIS 1

1.1. Introduction 1

1.2. Previous Chemometric Studies on Fire Debris Analysis 3

1.3. Outline of Chapters 6

CHAPTER 2: BURN METHODS AND PRODUCTS FORMED IN THE PYROLYSIS OF SUBSTRATES 7

2.1. Pyrolysis of Substrates 7

2.1.1. Pyrolysis Mechanisms 7

2.2. Laboratory Substrate Burns 11

2.2.1. Sample Preparation 11

2.2.2. Burn Methods 11

2.2.3. Sample Extraction (Passive Headspace Technique) 14

2.2.4. The Types of Substrates 15

2.2.5. Instrumental Parameters 16

2.3. Variations of the Three Burn Methods 17

2.3.1. Comparison of the Three Burn Methods 17

2.4. Pyrolysis and Combustion Products Formed in Different Types of Substrates 26

2.4.1. Plastics 26

2.4.2. Paper Products 29

2.4.3. Apparel 37

2.4.4. Automobile 41

2.4.5. Miscellaneous 44

2.4.6. Flooring 46

2.4.7. Building Materials 48

CHAPTER 3: LOGISTIC REGRESSION ANALYSIS 52

3.1. Analysis of the Burn Substrates and Ignitable Liquids 52

3.1.1. Calculation of Frequency of Occurrence of Compounds by Identified Five Major Peaks in Substrates and Ignitable Liquids Databases 52

3.1.2. Automated Mass Spectral Deconvolution and Identification System (AMDIS) 53

3.1.3. Data Analysis of Substrates and Ignitable Liquids 54

3.2. Logistic Regression Analysis of Substrates and Ignitable Liquids Data 55

3.2.1. Logistic Regression 55

3.2.2. Good-Turing Estimation 61

CHAPTER 4: CALCULATION OF LIKEILHOOD RATIOS USING NAÏVE BAYES 64

4.1. Application of Naïve Bayes Classifier 64

4.1.1. Naïve Bayes 64

4.1.2. Calculation of Likelihood Ratios using Naïve Bayes 65
4.2. Cross-validation and Calibration of Log-Likelihood Ratios 67
4.2.1. Cross-validation (CV) 67
4.2.2. Calibration of LLRs using Logistic Regression 81
CHAPTER 5: RESULTS OF THE LOGISTIC REGRESSION ANALYSIS 87
5.1. The Results of Five Major Compounds Analysis of Ignitable Liquids and Substrates 87
5.2. Compounds Identified Only in SUB, IL and Compounds Identified in Both SUB and IL 93
5.3. Compounds Identified in ASTM E1618-14 Ignitable Liquid Classes 106
CHAPTER 6: VALIDATION OF THE METHODS 129
6.1. Validation of the Naïve Bayes LLR Calculation Method 129
6.2. Classification of Fire Debris Data without Frequency Adjustments 134
6.3. Validation of Fire Debris Samples Using Frequency Adjustments 136
6.3.1. Laplace Estimation 137
6.3.2. Calculation of Naïve Bayes Log-Likelihood Ratios 138
6.3.3. Fire Debris LLR Projection on Pure Substrate and Ignitable Liquids Data 141
6.4. Likelihood Ratio Calibration using Logistic Regression 153
6.4.1. Calibration of Likelihood Ratios Calculated using All Compounds 154
6.4.2. Calibration of the Likelihood Ratios Calculated using Compounds in IL 157
CHAPTER 7: CONCLUSION AND FUTURE WORK 161

7.1. Conclusion 161

7.2. Future Work 163

APPENDIX A: DATA TABLES REQUIRED FOR CHAPTER 5 164

APPENDIX B: PUBLICATIONS AND COPY RIGHTS 176

REFERENCES 178

LIST OF FIGURES

Figure 1: Random scission mechanism of polyethylene[25] 8

Figure 2: Alkadiene, alkane and alkenes in the chromatogram of burned plastic wrap 9

Figure 3: Side group scission of PVC and formation of aromatic products[25] 10

Figure 4: Monomer reversion of polymethylmethacrylate[25] 10

Figure 5**:** Modified Destructive Distillation Method burn setup 12

Figure 6: Direct Heat burn setup 13

Figure 7**:** Indirect Heat Method burn setup 13

Figure 8: Total ion chromatograms of a) 2 min MDDM b) 2 min IH and c) Overlay of MDDM and IH of English Toffee carpet 18

Figure 9: Total ion chromatograms of a) 2 min MDDM b) 2 min IH 20

Figure 10: Total ion chromatogram of 2 min DH of English toffee carpet 21

Figure 11: Total ion chromatograms of a) 1 min MDDM b) 2 min MDDM c) 5 min MDDM of Sand Dune carpet 22

Figure 12: a) 2 min MDDM b) 5 min MDDM Hickory wood 25

Figure 13: 2 min MDDM total ion chromatogram of a plastic drinking water bottle 26

Figure 14: 2 min MDDM total ion chromatogram of an empty detergent container 27

Figure 15: 2 min MDDM total ion chromatogram of a Ziploc plastic container 28

Figure 16: 1 min MDDM total ion chromatogram of a styrofoam cup 29

Figure 17: 2 min MDDM total ion chromatogram of old newspaper 30

Figure 18: Total ion chromatogram of 2 min MDDM new newspaper 31

Figure 19: Total ion chromatogram of 2 min MDDM old magazine 32

Figure 20: Total ion chromatogram of 1 min DH old magazine 32

Figure 21: Total ion chromatogram of 2 min IH old magazine 33
Figure 22: Total ion chromatogram of 2 min MDDM new magazine 34
Figure 23: Total ion chromatogram of 1 min MDDM dixie cup 35
Figure 24: a) 1 min MDDM, b) 1 min DH, 1 min IH and d) unburned total ion chromatograms of carbonless paper 36
Figure 25: Expanded view of the TIC region of 9.5 – 13.5 min of 1 min MDDM carbonless copy paper 37
Figure 26: Total ion chromatogram of 1 min MDDM leather jacket 38
Figure 27: Total ion chromatogram of 1 min MDDM cotton shirt 39
Figure 28: Total ion chromatograms of 1 min a) MDDM b) DH and IH of women casual shoe 40
Figure 29: Total ion chromatogram of 2 min MDDM worn tire tread 42
Figure 30: Total ion chromatograms of 2 min a) MDDM b) DH and c) IH of dashboard 43
Figure 31: Total ion chromatogram of a) 2min and b) 5 min MDDM of railroad ties 45
Figure 32: Total ion chromatogram of 1 min MDDM olefin carpet 47
Figure 33: Total ion chromatogram of 2 min MDDM vinyl sheet 48
Figure 34: Total ion chromatogram of 2 min MDDM burn poly vinyl chloride (PVC) pipe 49
Figure 35: Total ion chromatogram of 2 min MDDM heavy duty construction adhesive 50
Figure 36: Total ion chromatogram of 2 min MDDM roof shingles 51
Figure 37: Probability cutoff determination using sensitivity (True Positive Rate) and specificity (True Negative Rate) a) ignitable liquids b) substrates (sensitivity and specificity are plotted in red and blue respectively) 58
Figure 38: Logistic regression curves a) ignitable liquids b) substrates 58
Figure 39: Performance of logistic regression models a) Ignitable liquids b) Substrates 60

Figure 40: The best fit linear model obtained for 2 min MDDM substrates 62

Figure 41: The best fit linear model obtained for ignitable liquids 63

Figure 42: Section of a data frame which includes the presence or absence of compounds 65

Figure 43: Sets of compounds used to calculate the likelihood ratios. 68

Figure 44: ROC curve obtained for calculated LLRs using compounds present in SUB and IL . 70

Figure 45: DET plot obtained for calculated LLRs using compounds present in SUB and IL 71

Figure 46: ECE plot obtained for calculated log-likelihood ratios 73

Figure 47: Representation of the discriminating power of calculated LLRs using tippet a tippet plot 74

Figure 48: Distribution of calculated LLRs a) IL b) SUB 75

Figure 49: ROC curve obtained for calculated LLRs using common compounds in SUB and IL76

Figure 50: (a) DET plot and (b) ECE plot obtained for calculated LLRs..... 77

Figure 51: (a) Tippet plot and (b) Histogram obtained for calculated LLRs..... 78

Figure 52: ROC curve obtained for calculated LLRs using compounds in IL 79

Figure 53: a) DET plot and b) ECE plot obtained for calculated LLRs 80

Figure 54: (a) Tippet plot and (b) Histogram obtained for calculated LLRs..... 81

Figure 55: (a) ROC plot (b) ECE plot and (c) Histogram obtained for the calibrated LLRs using all compounds in SUB and IL..... 83

Figure 56: a) ROC plot b) ECE plot and c) Histogram obtained for the calibrated LLRs using common compounds in SUB and IL..... 84

Figure 57: a) ROC plot b) ECE plot and c) Histogram obtained for the calibrated LLRs using compounds present in IL..... 86

Figure 58: Distribution of compound types in ignitable liquids and substrates 122

Figure 59: Compound types seen in aromatic and gasoline classes 124
Figure 60: Compound types identified in the petroleum distillate classes 125
Figure 61: Compound types identified in naphthenic paraffinic and iso-paraffinic ignitable liquid classes. 126
Figure 62: Compound types distribution in miscellaneous, oxygenated and normal alkane ignitable liquid classes 127
Figure 63: Projection of calculated LLR (method A) for fire debris samples to the ROC curve (red: samples with ILR, blue: samples without ILR) 130
Figure 64: Projection of calculated LLR (method B) for fire debris samples to the ROC curve (red: samples with ILR, blue: samples without ILR) 131
Figure 65: Projection of calculated LLR (method C) for fire debris samples to the ROC curve (red: samples with ILR, blue: samples without ILR) 132
Figure 66: ROC plots obtained for the calculated LLRs for pure SUB and IL by equal distribution frequency adjustments using a) all compounds b) compounds in IL c) compounds only in both SUB and IL 144
Figure 67: ROC curve generated from the calculated LLRs for 128 samples using compounds in IL (with 95% Confidence interval) 145
Figure 68: ROC plots obtained for the calculated LLRs for pure SUB and IL by Florida Fire Marshall data distribution frequency adjustments using a) all compounds b) compounds in IL c) compounds in both SUB and IL only. 149
Figure 69: ROC curve generated from the calculated LLRs for 54 samples using compounds in IL (with 95% Confidence interval) 150

Figure 70: ROC plots obtained for the calculated LLRs for pure SUB and IL by SUB and IL database data (at NCFS) distribution frequency adjustments using a) all compounds b) compounds in IL c) compounds in both SUB and IL only. 152
Figure 71: ROC curve generated from the calculated LLRs for 101 samples using compounds in IL (with 95% Confidence interval) 153
Figure 72: ROC plots obtained for the likelihood ratios calculated for known ground truth fire debris samples a) uncalibrated b) calibrated 154
Figure 73: The ECE plots obtained for the calculated log-likelihood ratios a) Uncalibrated b) Calibrated 155
Figure 74: The distribution of log likelihood ratios a) uncalibrated b) calibrated 156
Figure 75: Tippet plots obtained for calculated log-likelihood ratios a) uncalibrated b) calibrated 156
Figure 76: a) ROC plot b) ECE plot obtained for the log-likelihood ratios before calibration .. 157
Figure 77: a) Tippet plot b) histogram obtained for the log-likelihood ratios before calibration158
Figure 78: a) ROC b) ECE c) tippet and d) histogram of the calibrated likelihood ratios calculated by compounds in IL 159

LIST OF TABLES

Table 1: Compounds identified in both IL and SUB 88

Table 2: Compounds identified only in ignitable liquids 89

Table 3: Compounds identified only in substrates 92

Table 4: The compounds only identified in Substrates 93

Table 5: The compounds identified only in ignitable liquids 96

Table 6: Compounds seen in both SUB and IL 97

Table 7: Substrate types that can be seen in target compounds of GAS, MPD and HPD 102

Table 8 Frequency of occurrences of compounds in ASTM E1618-14[5] IL classes 108

Table 9: Compound types in the standard mass spectral library 121

Table 10: Total Number of peaks identified in each IL class 123

Table 11: Details of the laboratory-prepared fire debris samples 129

Table 12: Uncalibrated and calibrated log-likelihood ratios (LLR) obtained by the 3 methods described above 133

Table 13: Summary of the classification of fire debris samples 134

Table 14: Classification of fire debris based on IL class by the calculated LLRs using compounds present in IL 135

Table 15: The ratios between IL classes and SUB in the 3 distributions mentioned above 137

Table 16: Summary of the analysis of LLRs calculated by all compounds in SUB and IL 139

Table 17: Summary of the analysis of LLRs calculated by all compounds in IL 139

Table 18: Summary of the analysis of LLRs calculated by compounds in both SUB and IL only 140

Table 19: The number of samples selected for the calculation of likelihood ratios from each distribution 141

Table 20: Correct classification of SUB and SUB/IL mixture samples (equal distribution)..... 143

Table 21: Correct classification of SUB and SUB/IL mixture samples (Florida Fire Marshall Data distribution) 148

Table 22: Correct classification of SUB and SUB/IL mixture samples (SUB and IL Database distribution)..... 151

Table 23: Compounds and compound type 165

Table 24: Data obtained for compound type charts in Chapter 5 173

CHAPTER 1: DIFFERENT CHEMOMETRIC MODELS USED IN FIRE DEBRIS ANALYSIS

1.1. Introduction

The origin of fire can be accidental or intentional. An accidental or incendiary fire is one of the major problems in the United States and across the world. Unfortunately, it is not only a cause of property damages but also a major catastrophic event of death and permanent disability of living species. Fire investigation is one of the most important and challenging fields in forensic science since the evidence in the scene is destroyed by the fire. One of the major challenges in fire investigation is to identify whether a fire was caused by an accident or was a case of arson.

Fire debris analysis is defined as the identification of ignitable liquid residue (ILR) from fire debris samples collected at a fire scene[1]. Identifying major compounds of ignitable liquids in fire debris is one of the many tools used to determine the presence of ignitable liquid residue from a collected debris sample. Identification of these major compounds is a challenge because in most incidents, the analyst will not be able to detect any trace amount of ILR from the sample due to the vaporization of the liquid or the combination of pyrolysis and combustion products of substrates with ILR. Therefore, analyzing compounds in fire debris is not an easy task. Gas Chromatography – Mass Spectrometry (GC-MS) is the widely used technique in the analysis and is considered the gold standard in forensic science for fire debris analysis[2-4].

According to the American Society for Testing and Materials (ASTM) standard method E1618, ignitable liquids are classified under 7 classes: gasoline (GAS), petroleum distillate (PD), isoparaffinic (ISO), aromatic (AR), naphthenic paraffinic (NP), normal-alkanes (NA),

oxygenated solvents (OXY) and miscellaneous (MISC). These classes are differentiated by their chemical characteristics. In ASTM E1618-14[5], all ignitable liquid classes except gasoline are divided into subclasses of light, medium and heavy based on their carbon range. Some ignitable liquid classes (GAS, PD) have specific patterns which are characteristic to each class. Each class can be distinguished by their alkane, cycloalkane, aromatic and condensed ring aromatic profiles. The other important identification criteria for an ignitable liquid class is the presence of target compounds. In ASTM E1618-14[5], target compounds of gasoline, medium petroleum distillate and heavy petroleum distillate are presented in Tables 3, 4 and 5 respectively. Identification and evaluation of the presence of target compounds within fire debris is performed based on "total ion chromatograms" (TIC) and "extracted ion profiles" (EIP)[2-3].

At present, reporting and testimony in fire debris analysis are based on categorical statements which is based on the class determination of ignitable liquid residue using ASTM E1618-14[5]. In most instances, these statements are subjected to bias and do not reflect the quantitative value of the data. These categorical statements can be replaced by probabilistic statements, which contain a number, value or probability to reflect the quality of the data and express the strength of the evidence. In some fields of forensic science, the strength of evidence is reported as a likelihood ratio, an expression of the evidential value and log likelihood ratio, the weight of the evidence. The research reported in this dissertation applied logistic regression to identify the compounds in burned substrates and ignitable liquids and Naïve Bayes method was used to calculate likelihood ratios for reporting evidence in fire debris analysis.

A part of this work was an extension of the Substrate Database[6] (created and maintained by the National Center for Forensic Science) by adding of 1350 records corresponding to burned

samples and 150 records for unburned samples. Each substrate has 9 burned and 1 unburned sample. The samples were burned using 3 different methods: Modified Destructive Distillation Method (MDDM)[7], Direct Heat (DH) and Indirect Heat (IH). The major compounds of the total ion chromatograms (TIC) of these substrate samples and neat ignitable liquid samples present in the Ignitable Liquid Reference Collection Database[8] were identified using Automated Mass Spectral Deconvolution and Identification System (AMDIS) software.

1.2. Previous Chemometric Studies on Fire Debris Analysis

Chemometrics is defined as "the chemical discipline that uses mathematical, statistical, and other methods employing formal logic (a) to design or select optimal measurement procedures and experiments, and (b) to provide maximum relevant chemical information by analyzing chemical data"[9].

In the forensic science discipline, there have been a limited number of chemometric studies conducted until recently. The most commonly applied chemometric methods are:

i. Discriminant Analysis (DA) which can be divided into,
   a. linear (LDA)
   b. quadratic (QDA)
ii. Partial least-squares discriminant analysis (PLS-DA)
iii. Support vector machines (SVM)
iv. Naïve Bayes classifier (NBC)
v. Artificial neural networks (ANN)
vi. Principal Component Analysis (PCA)
vii. Cluster analysis (CA)

Some of these chemometric methods have already been applied in the field of fire debris analysis. One of the previous works was performed by Tan *et al.* used principal component analysis (PCA) to study the effects of pyrolysis products of substrates on ignitable liquids classification and developed a soft independent model classification analogy (SIMCA) to evaluate the variations in fire debris samples and to classify the class of the ignitable liquids correctly[10]. Sinkov *et al.* applied SIMCA and partial least square discriminant analysis (PLS-DA) to classify arson samples based on the ignitable liquid content in the samples[11].

Principal component analysis (PCA) and artificial neural networks (ANN) were used to classify premium and regular gasoline using gas chromatography and mass spectral data by Doble *et al.*[12]. Sandercock *et.al* performed PCA and LDA to differentiate samples of unevaporated gasoline using trace polar and poly aromatic hydrocarbons (PAH) compounds[13]. In another work, QDA and LDA were used to predict the fire debris samples as positive or negative for the presence of ignitable liquid residue (ILR)[14]. Sigman *et al.* applied SVM, LDA, QDA and kNN models to calculate likelihood ratios (LR) for fire debris samples[15]. Samples for that work were prepared by mixing data computationally from the ignitable liquid and substrate databases of the National Center for Forensic Science[6, 8]. Another study was conducted, to assess the evidentiary value of fire debris samples based on the models generated from the random draws of substrate and ignitable liquids database[6, 8] records of the National Center of Forensic Science. In this study likelihood ratios were calculated using one-level Gaussian kernel density models and multivariate means[16].

Analysis of the 5 major compounds identified in each neat Ignitable Liquids Reference Collection Database (ILRC)[6, 8] and burned samples of Substrate Database were done as an initial study for this work[17]. The results of this study will also be discussed in this dissertation. The

work reported in this dissertation, logistic regression analysis was performed for the compounds identified in substrates and neat ignitable liquids to calculate the probability of a presence of a compound in the respective substrate or ignitable liquid sample.

Logistic regression analysis was previously used in forensic speaker recognition[18] and identification of race using human skeletons in anthropology[19]. Naïve Bayes and other Bayesian networks have previously been applied in the criminal profiling of wild fires[20] but not in the field of fire debris analysis. In this work, logistic regression is used to identify the compounds present in ignitable liquids and burned substrates based on retention times combined with mass spectral data. Using this information, a Naïve Bayes method was applied to calculate likelihood ratios for substrates and ignitable liquids. This method was validated using 16 laboratory generated fire debris samples.

These methods were also validated using a large number of fire debris samples (405). In this, the frequencies of the compounds present in substrates and ignitable liquids were adjusted using 3 different population distributions of substrate and ignitable liquid class contributions. The data to obtain these distributions were obtained from Florida fire marshal data, NCFS databases of substrate and ignitable liquids and equal distributions of substrate and ignitable liquid classes.

1.3. Outline of Chapters

Chapter 2 of this dissertation discusses the different pyrolysis mechanisms which occur in substrates followed by the experimental procedures (burn methods and extraction of the samples) and the instrumental method and, there will be a discussion on comparison of effects in total ion chromatograms of selected substrates resulted by different burn methods.

Chapter 3 provides a detailed description of the data analysis using Automated Mass Spectral and Deconvolution System (AMDIS)[21]. Logistic regression analysis is the statistical method that was used to analyze the compounds as present or absent in substrates and ignitable liquids. This chapter will discuss the evaluation of the logistic regression models by ROC analysis, calculation of frequency occurrences of compounds that were present in substrates and ignitable liquids and finally the application of Good-Turing estimation[22].

Chapter 4 discusses the calculation of the Naïve Bayes likelihood ratios for fire debris samples using the frequency of occurrences of compounds in substrates and ignitable liquids and cross validation performed for these calculated values. These calculated likelihood ratios were calibrated using logistic regression, which is discussed in detail.

The logistic regression analysis of compounds present in substrates and ignitable liquids are discussed in Chapter 5 followed by the validation chapter (Chapter 6) of the methods used to calculate the log likelihood ratios of fire debris samples and the application of Laplace estimation will be discussed and finally, conclusion and future work will be discussed in chapter 7.

CHAPTER 2: BURN METHODS AND PRODUCTS FORMED IN THE PYROLYSIS OF SUBSTRATES

This chapter contains two sections. The first section discusses the pyrolysis mechanism of substrates; random scission, side group scission and monomer reversion and the experimental procedures and the details of the instrumental method. The samples were prepared using three different burn methods; Modified Destructive Distillation Method (MDDM), Direct Heat (DH) and Indirect Heat (IH) methods. These samples were extracted following American Society Testing and Materials standard E1412 (ASTM E1412 – 12)[23]. The second section of this chapter discusses the products formed by the different burn methods of various substrates.

2.1. Pyrolysis of Substrates

2.1.1. Pyrolysis Mechanisms

Pyrolysis is defined as the process, by which solids or liquids undergo degradation of their chemicals into lighter weight volatile molecules under heat without the interaction of oxygen or any other oxidant[24]. The rate of pyrolysis is directly proportional to the heat provided to the material by the source. Pyrolysis of the materials occurs by three main different mechanisms of chemical degradation. They are random scission, side group scission and monomer reversion (Figure 1, 3 and 4).

2.1.1.1. Random Scission

Random scission is a process which usually occurs in pyrolysis of polyethylene (PE) and polypropylene (PP)[25]. When this process occurs in PE, the backbone of the polymer randomly breaks into segments as explained in Figure 1. This degradation initiates radical formation. Random scission produces small alkanes, alkenes and alkadienes by forming radicals. This scission results in "triplets" in the total ion chromatogram (TIC) of PE, as depicted in Figure 2. This TIC was obtained from burning a plastic wrap. In this TIC, alkene and alkane were identified as undecene and n-undecane (C_{11} region).

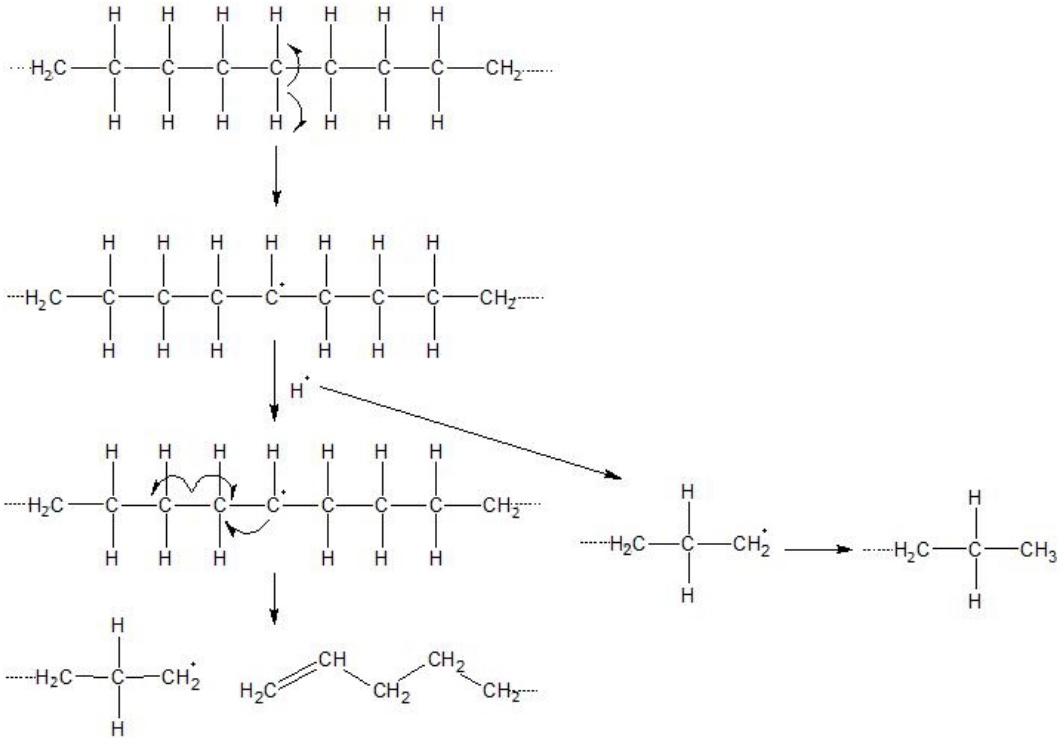


Figure 1: Random scission mechanism of polyethylene[25]

Random scission in polypropylene gives rise to branched alkanes[24]. They do not have a significant pattern in the respective TIC as in polyethylene.

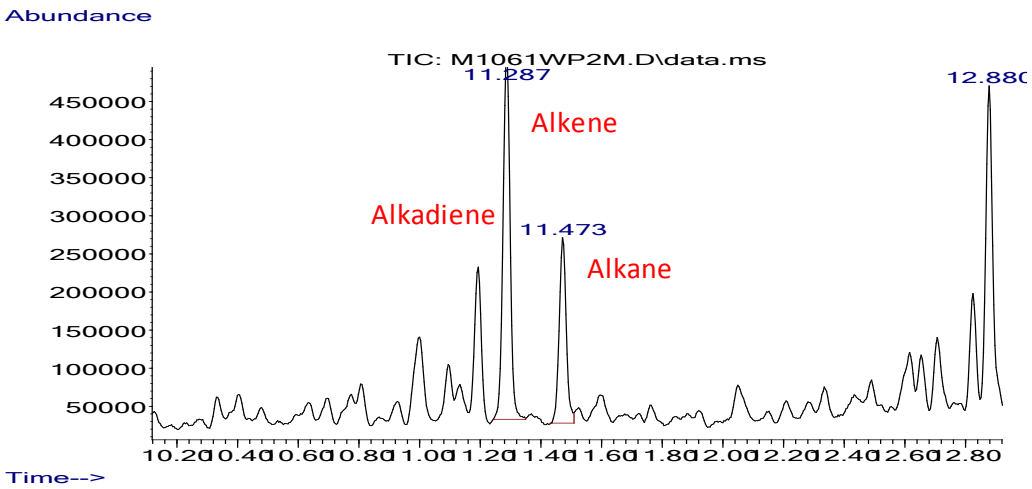


Figure 2: Alkadiene, alkane and alkenes in the chromatogram of burned plastic wrap

2.1.1.2 Side Group Scission

Side group scission forms unsaturated linear carbon chains by cleaving the side chain of the polymer[25]. This process forms the aromatics as the final pyrolysis products. An example of this mechanism is depicted in Figure 3. This explains the formation of aromatic products in polyvinyl chloride (PVC) which was generated by randomly breaking the unsaturated carbon backbone. Some of the examples of these aromatic products are benzene, toluene, naphthalene and ethylbenzene.

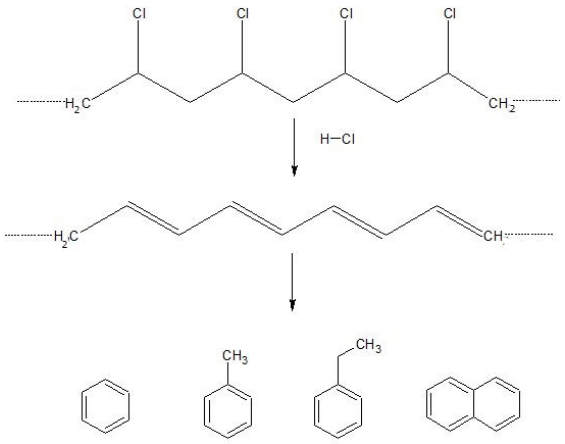


Figure 3: Side group scission of PVC and formation of aromatic products[25]

2.1.1.3. Monomer Reversion

In monomer reversion, the polymer is simply transformed back to its original version. This is explained in Figure 4. This is the monomer reversion mechanism of polymethylmethacrylate (PMMA).

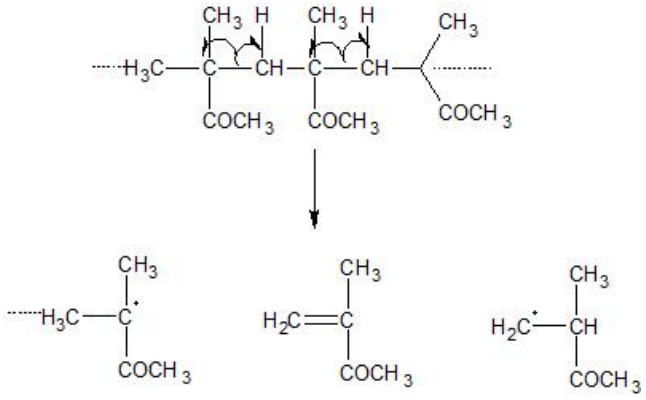


Figure 4: Monomer reversion of polymethylmethacrylate[25]

*Figure 1, 3 and 5 were adapted from the Reference 26 given in the dissertation

2.2. Laboratory Substrate Burns

2.2.1. Sample Preparation

For this study, ten samples were prepared for each substrate; 3 samples for each burn method and the unburn sample. The measured weight of all the substrate samples was above 1.0000 g. The area of the cut substrate samples was approximately 16 cm^2. All the information about each substrate sample can be found in the Substrate Database of National Center for Forensic Science (http://ilrc.ucf.edu/substrate/)[6]. The samples were burned using three burn methods, Modified Destructive Distillation Method (MDDM)[7], Direct Heat (DH) and Indirect Heat (IH) methods.

2.2.2. Burn Methods

2.2.2.1. Modified Destructive Distillation Method (MDDM)

Each sample was placed in a quart-sized can where the top surface of the substrate was contacting the bottom of the can and the can was closed with a lid. This lid was punctured to create nine holes. The diameter of a punctured hole was approximately 1 cm. The burn setup of the MDDM is presented in Figure 5. A propane torch was used as the heat source. The distance between the tip of the propane torch and the bottom of the quart can was fixed to 4 cm. The MDDM was performed for each substrate for 3 different time intervals (1, 2 and 5 mins). The total time was measured as soon as the flame touched the bottom of the can and the designated time intervals were started as the smoke appeared through the holes of the lid. After completing the burn at the required time interval, the punctured lid was replaced with a non-punctured lid.

This lightly closed can was allowed to cool down to room temperature and after that, the sample was prepared for the extraction which will be discussed later in this chapter (Section 2.2.3).

Figure 5**:** Modified Destructive Distillation Method burn setup

2.2.2.2. Direct Heat Method (DH)

The DH method was performed at 3 designated time intervals (1, 2 and 3 mins). The sample was placed on the lid where the top surface of the substrate was directly interacting with the flame of the propane torch. The measuring of the time intervals started as soon as the flame interacted with the substrate. After burning the substrate, it was covered with a clean inverted quart can. Then the substrate sample containing can was allowed to cool down to room temperature. The DH method setup is given in Figure 6. In the DH, the total time was not measured since the substrate was directly interacting with the flame.

Figure 6: Direct Heat burn setup

2.2.2.3. Indirect Heat Method (IH)

As same as in the Direct Heat method, the Indirect Heat method was also performed at 1, 2 and, 3 minutes time intervals. In this method, the sample was placed on the lid where the top surface of the substrate directly interacting with the heat. The total time for the burn was measured as soon as the flame touched the bottom of the lid. The IH method setup is shown in Figure 7. Measuring of the time intervals was begun as the smoke appeared. After completion of the burn at the required time interval, the lid was covered using an inverted quart can and was allowed to cool down to room temperature before the extraction.

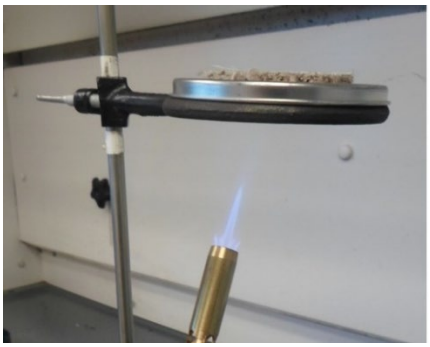


Figure 7**:** Indirect Heat Method burn setup

2.2.3. Sample Extraction (Passive Headspace Technique)

The samples were prepared for extraction following American Society Testing and Materials standard E1412 (ASTM E1412 – 12)[23] protocol. This standard is the common extraction practice that is used in the field of fire debris analysis since it is a sensitive and a non-destructive technique[23]. Once the burned sample was cooled down, the activated charcoal strip was inserted to the headspace of the can using a clean paper clip and non-scented/un-waxed dental floss to suspend the strip above the sample and then the can was tightly sealed. The area of the charcoal strip was about 100 mm^2. The burned samples were cooled down to allow the vapors to condense inside the quart can prior to inserting the charcoal strip.

The activated charcoal strip inserted can was placed in an oven for 16 – 18 hours at 66 °C. This temperature and the duration allow the lighter weight volatile compounds to adsorb onto the charcoal strip. If the temperature was increased or the duration was longer, the higher molecular weight volatile compounds would adsorb to the charcoal strip but could reduce the abundance of the lighter molecular weight compounds. This procedure was followed for all burned and unburn substrate samples. The samples were analyzed following ASTM E1618 – 14 protocol[5]. Once the can was removed from the oven and allowed to cool down to room temperature, the charcoal strip was removed from the can and inserted to a glass vial. After that, the charcoal strip was submerged completely with 0.5 mL of carbon disulfide (CS_2) and GC-MS analysis was performed for the extracted samples.

2.2.4. The Types of Substrates

The samples were selected from different categories. They were flooring, building materials, apparel, miscellaneous, automobile, furnishings, paper and plastic products. Flooring: carpet, vinyl/linoleum, hardwood, laminate, carpet padding, garage/exercise flooring, and engineered. Building materials: roofing, insulation, wood, drywall, siding, concrete/masonry, adhesives, ceiling, composite decking, moldings/trim, particle/fiberboard, PVC pipe and pre-treated wood. Apparel: new clothing, new footwear, worn clothing, worn footwear and accessories. Miscellaneous: Railroad ties, household materials, rope and packing materials. Furnishing: bedding, upholstery, cushions, window treatments, chair/couch, dresser, table, bed and accessories. Paper products: cardboard, newspaper, magazines, paper (thermal, copy, fax, ruled, resume (cotton), letterhead). Detailed information of the substrate materials and the total ion chromatograms can be found on the Substrate Database of National Center for Forensic Science[6].

2.2.5. Instrumental Parameters

All substrate samples were analyzed using Gas – Chromatography/Mass – Spectrometry (GC-MS). The gas chromatograph was an Agilent 7890A with a G45567A series autosampler with a Merlin septumless injector which was connected to a 5977E mass spectrometer. Split injection method was used to introduce the sample to the instrument. In this method, 1 µL of the sample was split to 50:1 ratio and injected to the instrument at 250 °C. During each 30 mins run the sample was held at 50 °C for 3 mins and ramped up to 280 °C at a rate of 10 °C/min^{-1} and the hold time at the end was 4 mins. The chromatographic column was a 0.2 µm internal diameter, 24.36 m, 0.5 µm film thickness HP-1 methyl siloxane column operated with a He carrier gas flow of 34 cm s^{-1} linear velocity. The scanning range of the mass analyzer was 30 – 350 m/z. The quad temperature of the mass spectrometer was 150°C and the source temperature was 230°C.

2.3. Variations of the Three Burn Methods

2.3.1. Comparison of the Three Burn Methods

One of the most important observations was that the formation and the abundance of the pyrolysis products of the total ion chromatogram (TIC) depend on the burn method and the designated time intervals. This also depends on the substrate. Out of all 3 burn methods, MDDM was considered the best burn method since it captured more pyrolysis products inside the can. A comparison of chromatograms in MDDM with IH and DH methods are done using TICs of burned polyethylene terephthalate 100% (English toffee/brown color) carpet and Natural Maple hardwood flooring.

2.3.1.1. Comparison of MDDM and Indirect Heat Method

The total ion chromatograms of the 2 min MDDM and IH of the English Toffee carpet are presented in Figure 8a and 8b. An overlay of these chromatograms is presented in Figure 8c. The 2 min MDDM total ion chromatogram (TIC) indicates the presence of more compounds than the Indirect Heat method TIC. In the 2 min MDDM, the five most abundant peaks were biphenyl (15.420 min), benzophenone (18.432 min), benzoic acid (12.379 min), vinyl benzoate (11.733 min) and 2,2,4-trimethyl-1,3-pentanediol diisobutyrate (TXIB, 18.237 min).

(a)

(b)

(c)

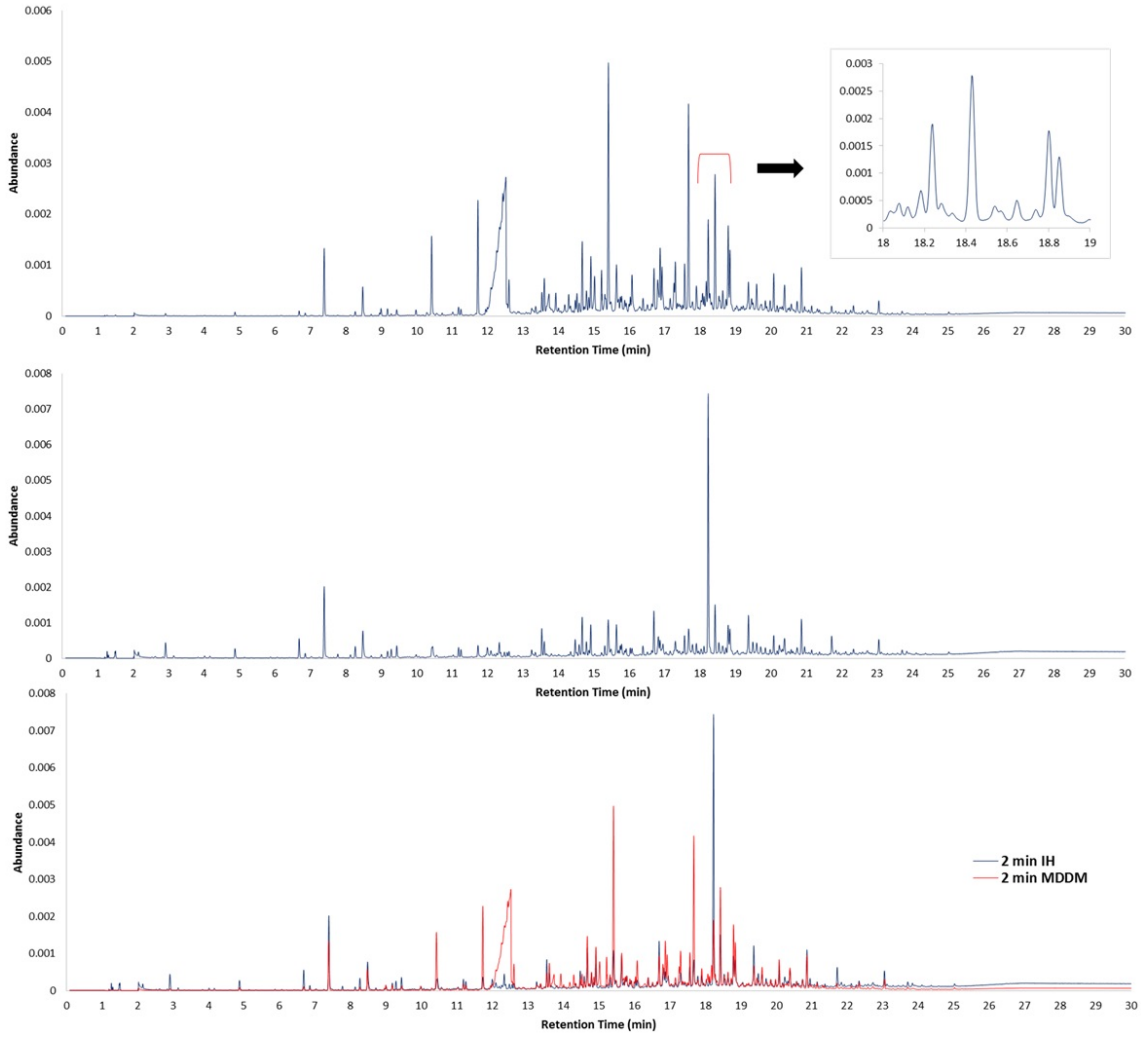


Figure 8: Total ion chromatograms of a) 2 min MDDM b) 2 min IH and c) Overlay of MDDM and IH of English Toffee carpet

When compared to the MDDM, the total number of peaks and their relative abundance in the Indirect Heat method were low. One of the possible reasons for this was the loss of products caused by the IH burn method. In this method, the majority of the volatile compounds were released directly to the open environment unlike in the MDDM. The most abundant peaks of IH were TXIB and styrene. TXIB is generally used as a plasticizer in products[26]. This trend changed

in some engineered hardwood flooring materials. In Natural Maple engineered flooring, the 2 min IH method provided more products than the 2 min MDDM. The chromatograms of these are provided in Figure 9a and 9b respectively.

Engineered hardwood consists of two main layers; a top layer which is real hardwood and a core which is made out of multiple layers of plywood and a durable plank or high-density fiberboard (HDF)[27]. Due to these multiple layers, engineered flooring requires more heat to penetrate through the substrate to form more pyrolysis or combustion products from each layer.

In the 2 min IH method (Figure 9b), the five most abundant peaks were 2-furaldehyde (5.815 min), 2-methoxyphenol (10.866 min), 2,6-dimethoxyphenol (14.721 min), 2-methoxy-4-methylphenol (12.585 min) and 4-ethyl-2-methoxyphenol (13.901 min). In the MDDM (Figure 9a), 2-furaldehyde (5.820 min) had the highest abundance followed by n-eicosane (22.657 min), n-nonadecane (21.666 min), 2-methoxyphenol (10.879 min) and furfuryl alcohol (6.436 min).

In the pyrolysis products of MDDM in hardwood, the long chain alkanes (n-eicosane and n-nonadecane) have a higher abundance than the oxygenated products. Formation of more oxygenated products in IH could be an indication that the substrate interacted with more O_2 than MDDM since it was on the lid itself and not contained. In addition, the relative abundance of the products was also increased in the IH method burn. One possible reason for this could be the presence of more O_2 increased the combustion of the flooring material, hence increased the abundance of the oxygenated products.

(a)

(b)

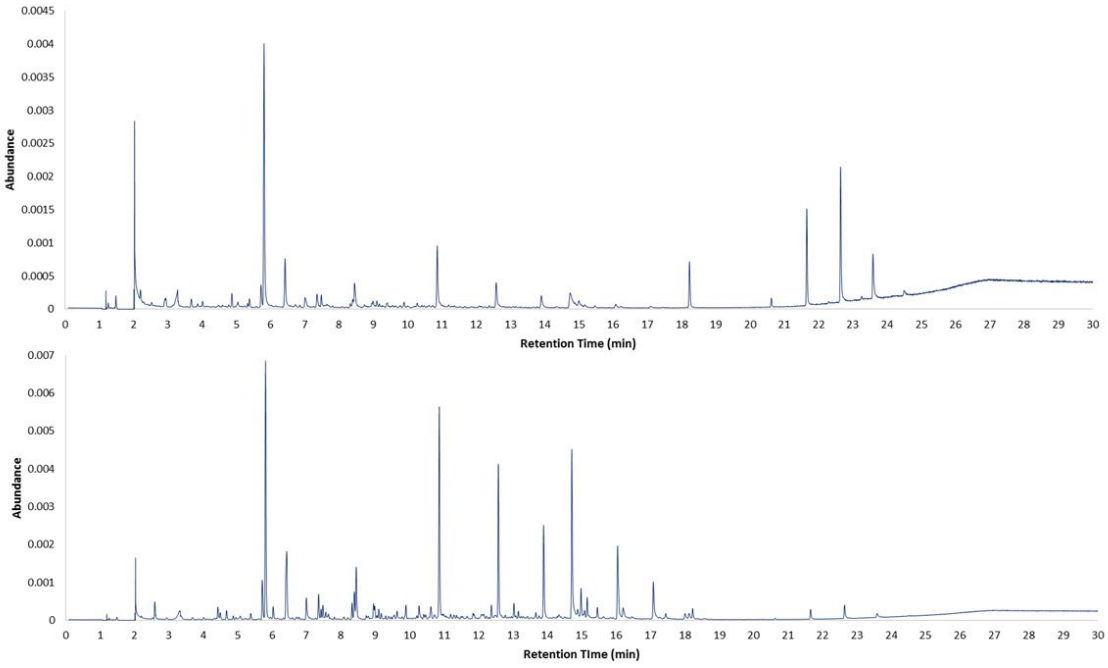


Figure 9: Total ion chromatograms of a) 2 min MDDM b) 2 min IH

2.3.1.2. Comparison of MDDM and Direct Heat Method

Generally, the presence of the total amount of products observed in the total ion chromatograms in DH method was low when compared to MDDM. This is explained using the 2 min Direct Heat burn of English toffee carpet which was made out of polyester and depicted in Figure 10. As seen in the IH method, the relative abundance of the compounds was decreased in the DH method. The most abundant peaks of this burn were TXIB (18.22 min), styrene (7.395 min) and 2,4-dimethyl-1-heptene (6.686 min). As same as in the IH method, in DH method, the relative abundance of TXIB was high compared to MDDM.

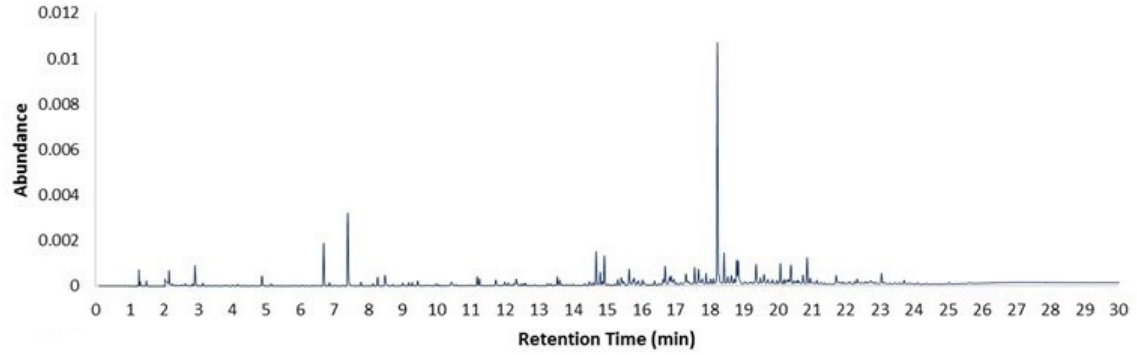


Figure 10: Total ion chromatogram of 2 min DH of English toffee carpet

One of the possible reasons for higher abundance in TXIB in DH and IH could be the interaction of the substrate with more O_2 which increases the combustion as described above. Since this process was performed on the lid, the light-weight molecular compounds were likely to be removed, but the molecular weight of TXIB was high, hence it was more likely to be remained inside the can.

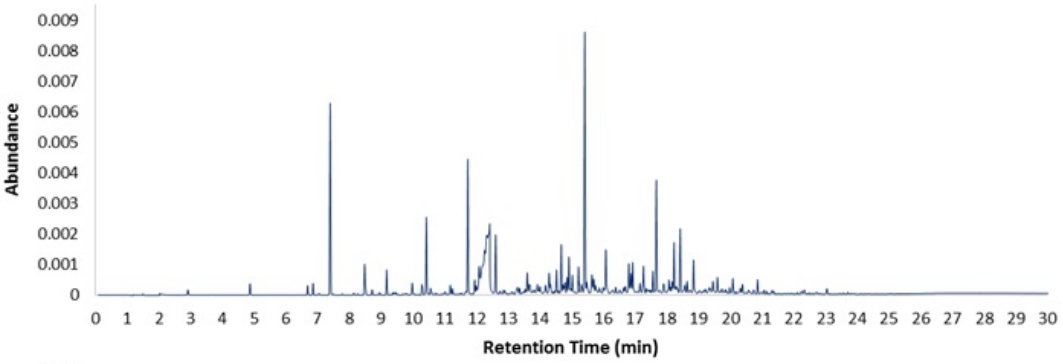
2.3.1.3. Differences of MDDM within the Time Intervals

Differences of the time intervals in the MDDM are explained using the TICs of burned 100% dyed (sand dune/brown color) polyester carpet. The chromatograms of MDDM 1, 2 and 5 min are presented in Figure 11a, 11b and 11c respectively. In the 1 min MDDM chromatogram, 5 major peaks were identified. They were, (relative abundance highest to lowest) biphenyl (15.420 min), styrene (7.37 min), vinyl benzoate (11.373 min), acetophenone (10.429 min) and benzophenone (10.429 min).

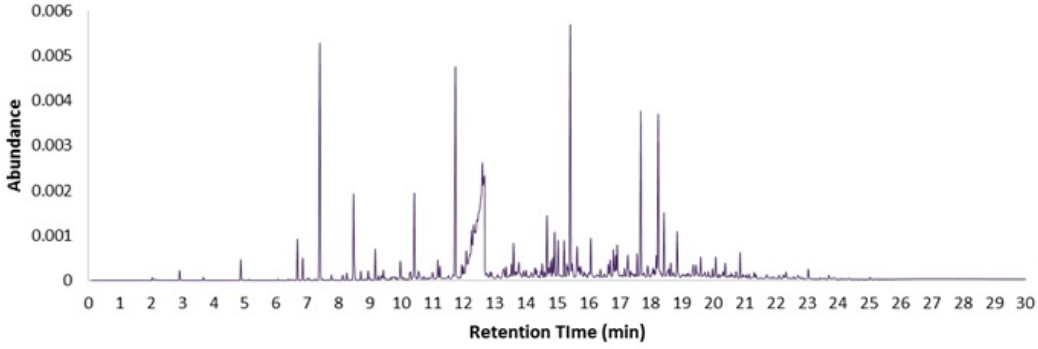
The peak at retention time 17.677 min was not identified since the standard mass spectral library did not have this compound. In general, the abundance of compounds increased in the 2 min MDDM when compared to the 1 min MDDM. The reason for this could be that when the

time interval was increased, the substrate was exposed to more heat and increased the rate of pyrolysis. This might impacted to raise the abundance of the products. The 5 most abundant peaks of this sample (2 min MDDM) were biphenyl (15.426 min), styrene (7.402 min), vinyl benzoate (11.745 min), TXIB (18.244 min) and benzaldehyde (8.490 min).

(a)



(b)



(c)

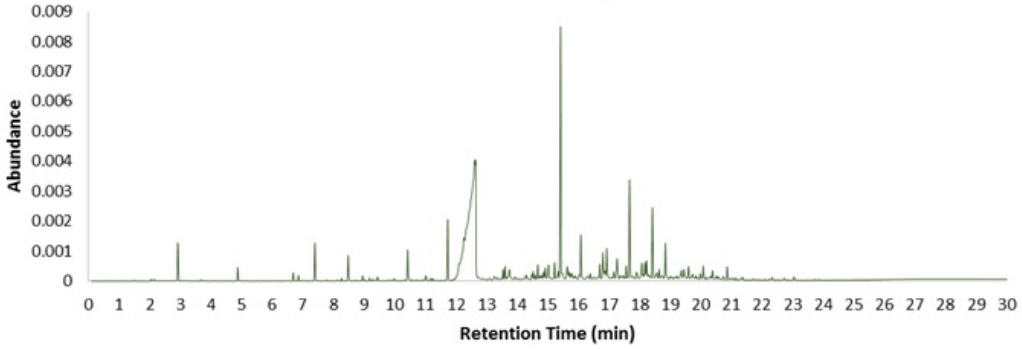


Figure 11: Total ion chromatograms of a) 1 min MDDM b) 2 min MDDM c) 5 min MDDM of Sand Dune carpet

In 5 minute MDDM, the relative abundance of the products were decreased, except biphenyl. There were two possible reasons for this; one reason was the substrate was extensively burnt and could not produce pyrolysis and combustion products further and the other reason may have been that when the time interval was increased, the formed products had a more tendency to get removed from the quart can, therefore less number of products were condensed inside the can. In 5 min MDDM identified 4 major compounds were biphenyl (15.421 min), benzoic acid (12.504 min), benzophenone (18.427 min), vinyl benzoate (11.733 min). The relative abundance of styrene (7.395 min) decreased drastically in 5 min MDDM than 1 or 2 min MDDM.

The difference of the abundance of the TICs of 1, 2 and 5 min burn intervals also depended on the type of the substrate used. This will be discussed using hickory wood. Total ion chromatograms of 2 and 5 min burned hickory wood are given in Figure 12a and 12b respectively. In these chromatographic profiles, the abundance of the compounds was increased as the designated time interval increased.

2.3.1.4. Burning of Wood

Wood is a composition of cellulose, hemicellulose and lignin. The outer layer of wood is mainly composed of cellulose and inner layers are a combination of cellulose, hemicellulose and lignin. When the top layer of the wood (cellulose wall) is exposed to heat, it destroys the chemical structure of cellulose, which produces light volatile pyrolysis products. These light volatile products react with O_2 and produce combustion products of wood[28].

Pyrolysis of wood undergoes two pathways depending on the environmental conditions such as temperature, O_2 concentration or fire retardants. At temperatures below 300 ˚C, cellulose and lignin chemically break down to form carbonyls, carboxyls and hydroperoxides and free radicals. When these products react with O_2, exothermic combustion of the substrate occurs. The heat generated from this process in the vapor phase is then transferred back to the wood. This process increases the pyrolysis rate by raising the temperature of the solid material[29].

At 300 ˚C, cellulose undergoes depolymerization by transglycosylation to form 1,6-anhydro-β-D-glucopyranose and 1,6-anhydro-β-D-glucofuranose. These compounds are then converted into lighter molecular weight products. At temperatures above 300 ˚C, the formation of tar increases whereas char formation decreases[29].

In the MDDM of hickory wood, the relative abundance of the majority of the products increased with the designated burning time interval. As described above, combustion of the wood occurred when these volatile pyrolysis products reacted with O_2, therefore it could raise the temperature of the inner layers of the wood. This process increased the rate of pyrolysis of the inner layers of the wood. As the time interval increased, the heat exposure of the substrate

was also increased. In 1 min MDDM, there were no products observed in the chromatogram. The pyrolysis products can be seen in 2 and 5 min MDDM (Figure 12a and 12b).

(a)

(b)

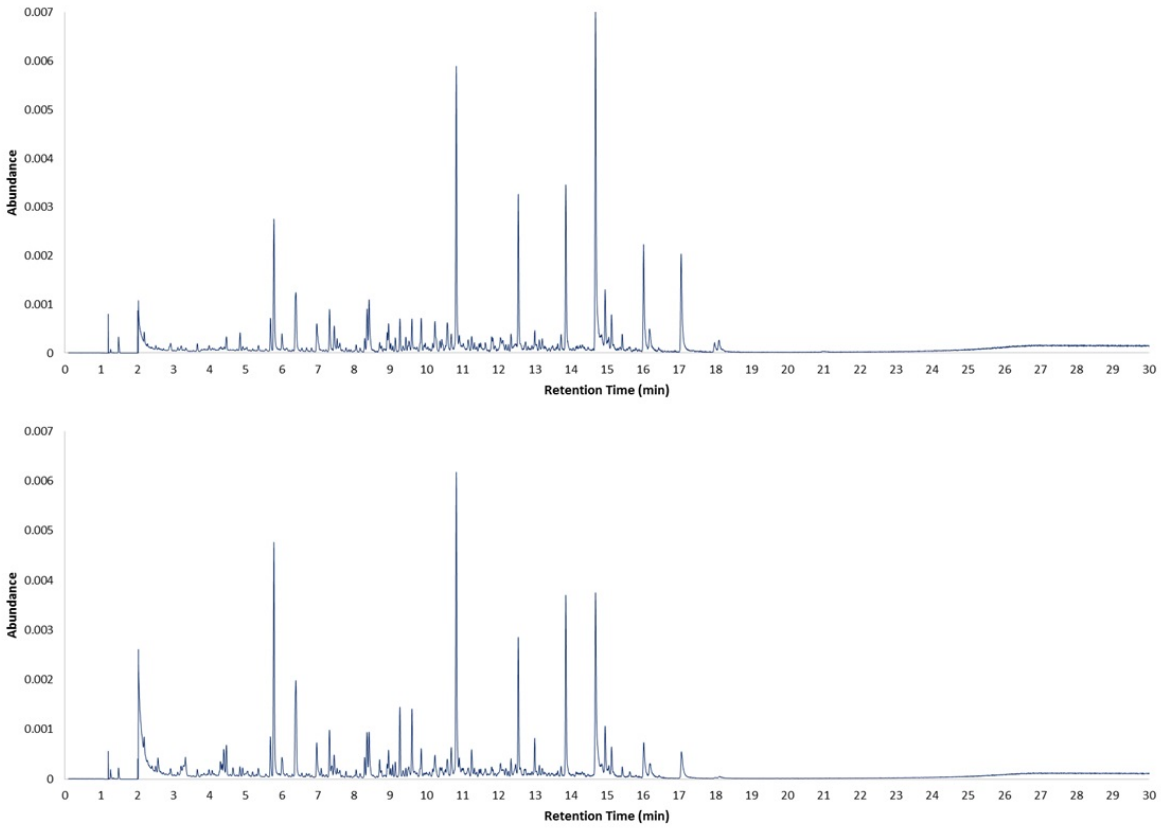


Figure 12: a) 2 min MDDM b) 5 min MDDM Hickory wood

The five most abundant peaks in 2 min MDDM are 2,6-dimethoxy phenol (14.686 min), 2-methoxy phenol (10.831 min), 4-ethyl-2-methoxyphenol (13.864 min), creosol (12.549 min) and 2-furaldehyde (5.783 min). In the 5 min MDDM, the five major abundant peaks were 2-methoxy phenol (10.831 min), 2-furaldehyde (5.783 min), 2,6-dimethoxy phenol (14.687 min), 4-ethyl-2-methoxyphenol (13.864 min) and creosol (12.549 min).

2.4. Pyrolysis and Combustion Products Formed in Different Types of Substrates

2.4.1. Plastics

Plastic types which were burned for this work can be categorized into PETE (Polyethylene terephthalate), HDPE (High-Density Polyethylene), PE (Polyethylene), PP (polypropylene), PS (Polystyrene).

2.4.1.1. Polyethylene Terephthalate (PETE or PET)

Polyethylene terephthalate is a polymer synthesized by terephthalic acid and ethylene glycol. A disposable drinking water bottle (500 mL) was burned to analyze the compounds of PET. The pyrolysis products were only seen in the 2 min MDDM burn (Figure 13). The major identified peaks of the TIC were vinyl benzoate (11.716 min) and benzene (2.905 min).

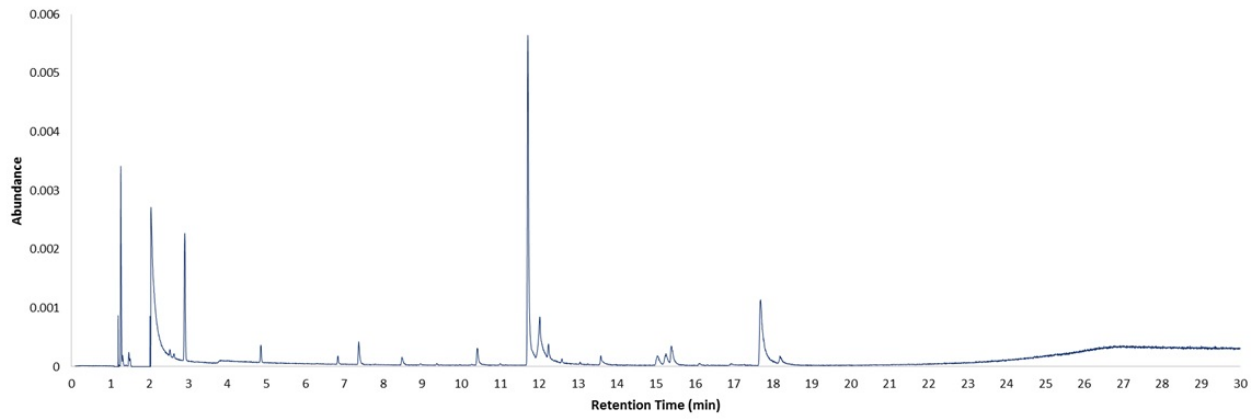


Figure 13: 2 min MDDM total ion chromatogram of a plastic drinking water bottle

2.4.1.2. High-Density Polyethylene (HDPE)

A laundry detergent container made out of HDPE was pyrolyzed. Pyrolysis products of HDPE have a very significant pattern in the chromatogram as discussed at the beginning of this chapter. The 2 min MDDM total ion chromatogram (TIC) provided the highest abundance of the products when compared to the other methods. The TIC is provided in Figure 14. The five major identified peaks of this sample were 1-dodecene (12.8793 min), 1-undecene (11.2885 min), 1-decene (9.5514 min), 1-pentadecene (17.0235 min) and limonene (10.1939 min). Limonene peak was more likely to be obtained from the detergent itself and not as a pyrolysis product of the material.

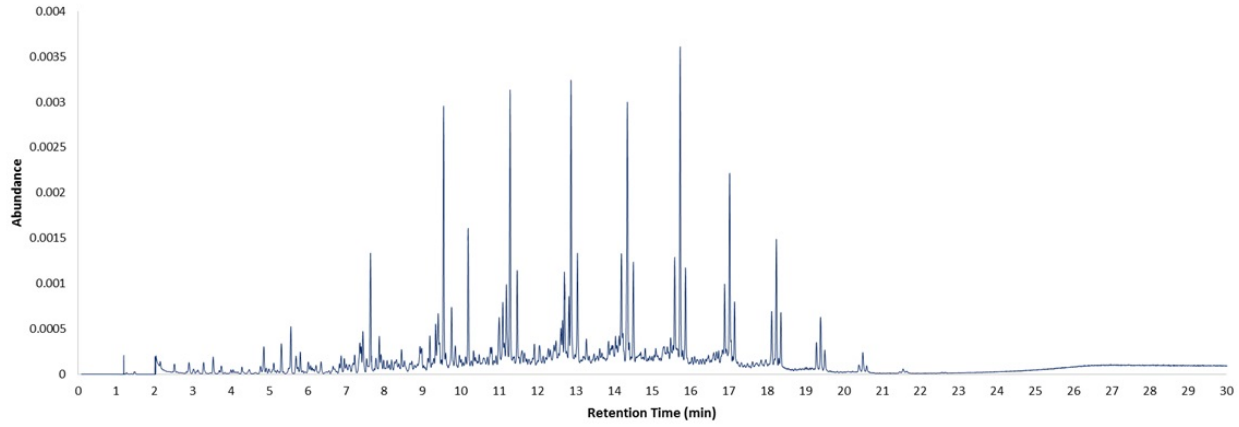


Figure 14: 2 min MDDM total ion chromatogram of an empty detergent container

2.4.1.3. Polypropylene

A typical sample for polypropylene was a Ziploc plastic container. As same in other types of plastics, the 2 min MDDM produced the highest amount of compounds. The five major identified peaks in this chromatogram were 2,4,6,8-tetramethyl-1-undecene (isotactic) (14.656 min), 2,4,6,8-tetramethyl-1-undecene (syndiotactic) (14.897 min), 2,4-dimethyl-1-heptene (6.666 min), 2,4,6,8-tetramethyl-1-undecene (heterotactic) (14.774 min), and 2,4,6-trimethyl-1-nonene (meso) (11.162 min). This TIC is given in Figure 15.

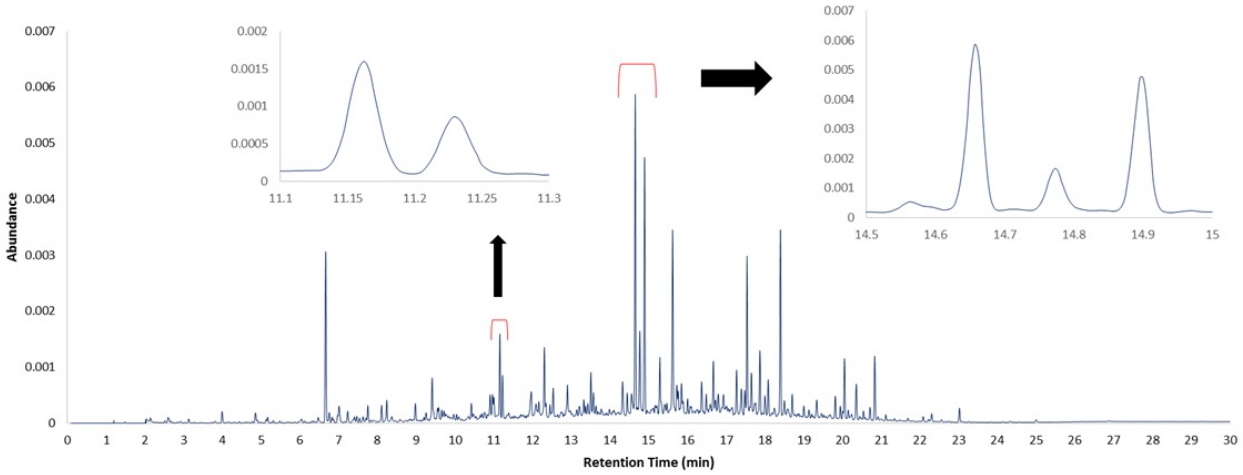


Figure 15: 2 min MDDM total ion chromatogram of a Ziploc plastic container

2.4.1.4. Polystyrene

A styrofoam cup is a typical example of a sample of polystyrene. In this, 1 min MDDM showed the highest abundance in compounds (Figure 16). Styrene (7.410 min) had the highest abundance in this TIC followed by 1,3-diiphenyl-1-butene (styrene dimer) (19.585 min), bibenzyl (17.243 min) and benzaldehyde (8.466 min).

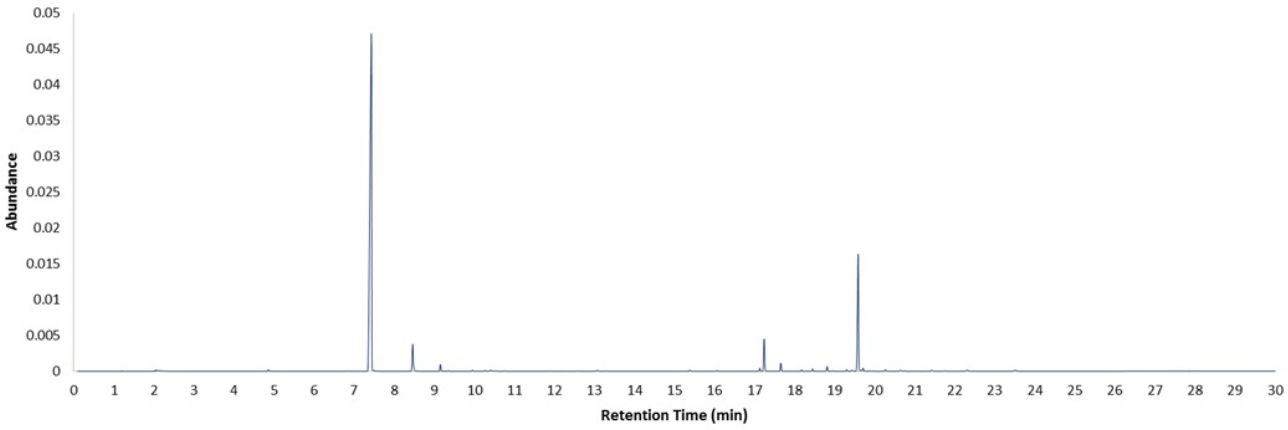


Figure 16: 1 min MDDM total ion chromatogram of a styrofoam cup

2.4.2. Paper Products

Paper products used in this work were newspapers, copy paper, magazines, letterhead paper, cotton linen paper, thermal paper and carbonless paper. The common compounds that can be seen in all paper products were furfural, 2-methoxy phenol, creosol and 5-methyl furfural. When the burn times increased in all methods, the substrate was burned to ash.

2.4.2.1. Newspaper

For this study, an old and a new newspaper were burned to observe if the condition of the material had an effect on the pyrolysis products of the substrate. Only 1 and 2 min MDDM methods provided a higher number of products when both old and new newspaper were burned. Comparison of 2 min MDDM of the old and new newspapers are presented in Figure 17 and 18.

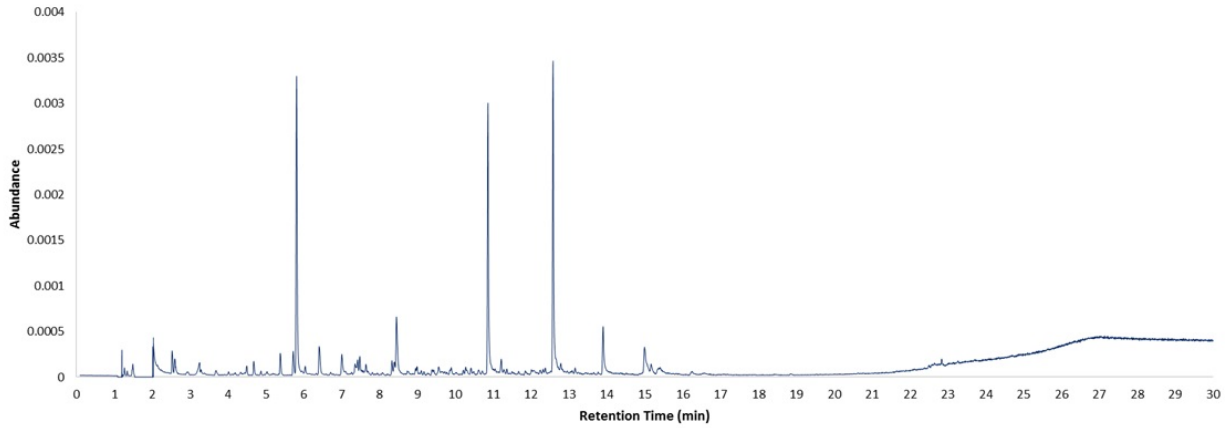


Figure 17: 2 min MDDM total ion chromatogram of old newspaper

In this chromatogram, the identified major peaks were furfural (5.811 min), 2-methoxy phenol (10.864 min) and creosol (12.584 min).

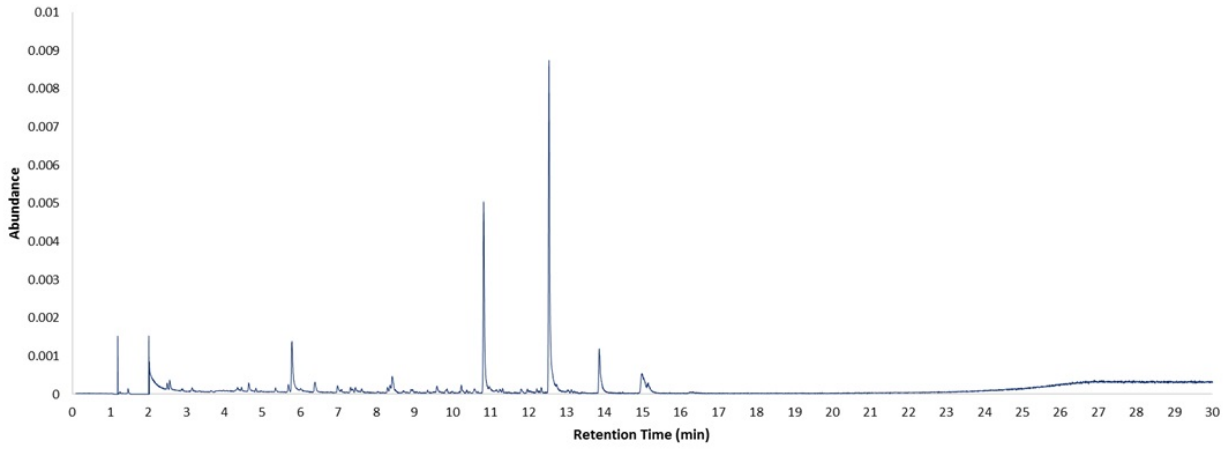


Figure 18: Total ion chromatogram of 2 min MDDM new newspaper

In the burned new newspaper, the abundance of furfural was drastically decreased. However, 2-methoxy phenol (10.839 min) and creosol (12.565 min) have high abundance in the chromatogram. The abundance of products obtained from DH and IH was very low, therefore the details of these burn methods are not discussed here.

2.4.2.2. Magazines

One of the main differences in the products formed in the old magazine was the abundant of styrene which was not observed in the new magazine. There was a difference in the products formed in MDDM, DH and IH of the old magazine. This was visible in the 18 – 21 min range in the total ion chromatograms. These differences in these TICs are depicted in Figure 19, 20 and 21 respectively.

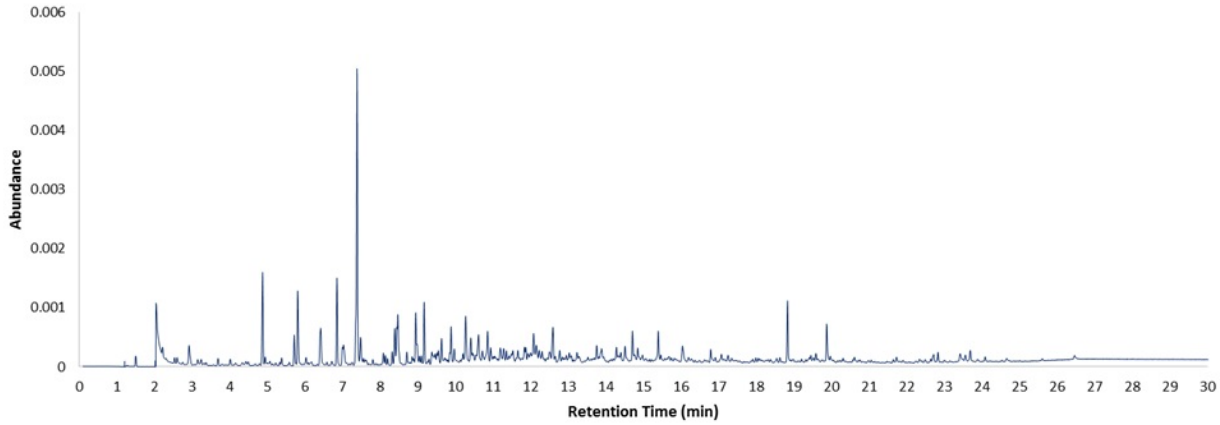


Figure 19: Total ion chromatogram of 2 min MDDM old magazine

The identified highest abundant peaks of this chromatogram were styrene (7.377 min), toluene (4.867 min), ethylbenzene (6.847 min) and furfural (5.807 min).

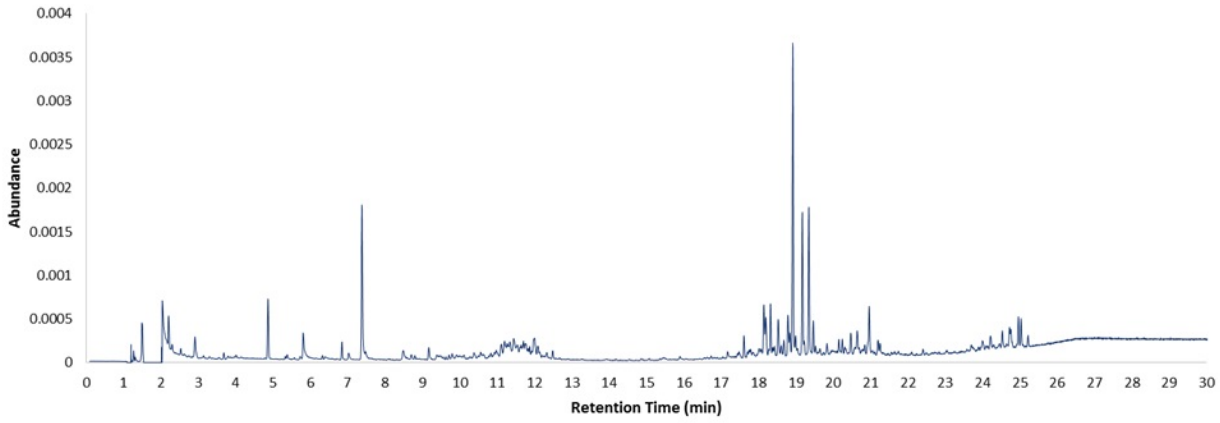


Figure 20: Total ion chromatogram of 1 min DH old magazine

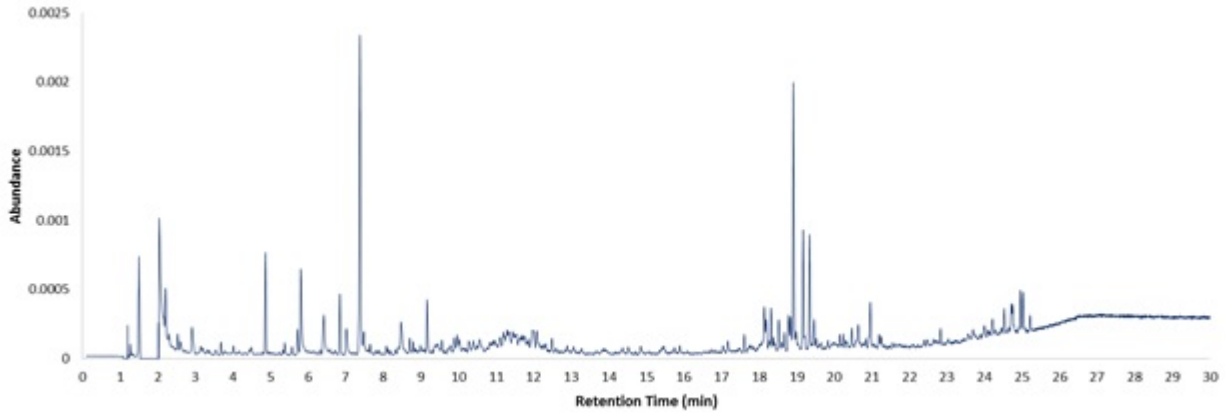


Figure 21: Total ion chromatogram of 2 min IH old magazine

The compounds between 18 – 21 min in DH and were not identified. The difference between MDDM and other two burn methods could be the limitation of O_2 in MDDM. Therefore, the products in this range might have been produced from combustion rather than from the pyrolysis of the material. The most abundant peak in the new magazine was 2-methoxy phenol which was different from the old magazine. The chromatogram for 2 min MDDM burned new magazine is depicted in Figure 22.

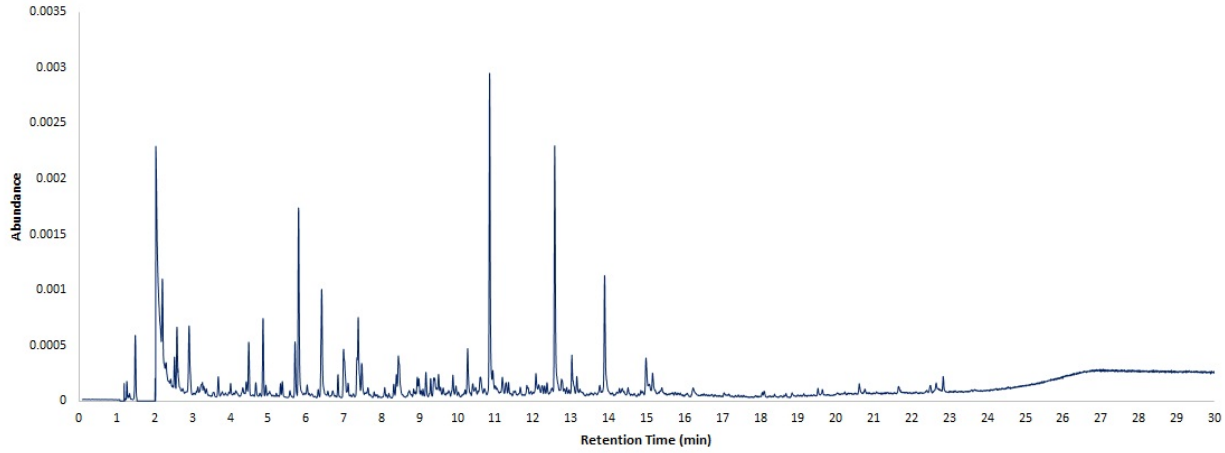


Figure 22: Total ion chromatogram of 2 min MDDM new magazine

The major identified peaks of this sample were 2-methoxyphenol (10.856 min), creosol (12.581 min), furfural (5.811 min), 4-ethyl-2-methoxy phenol (13.900 min), and furfuryl alcohol (6.419 min). Direct heat and indirect heat methods for this substrate did not provide any pyrolysis or combustion products since the material was burned completely.

2.4.2.3. Dixie cup

The pyrolysis products formed in the dixie cup were different from other paper products. Furfural at 5.803 min was the major peak identified in the TIC which is shown in Figure 23. The other identified peaks were alkenes. However, the abundance of these peaks was very low compared to furfural.

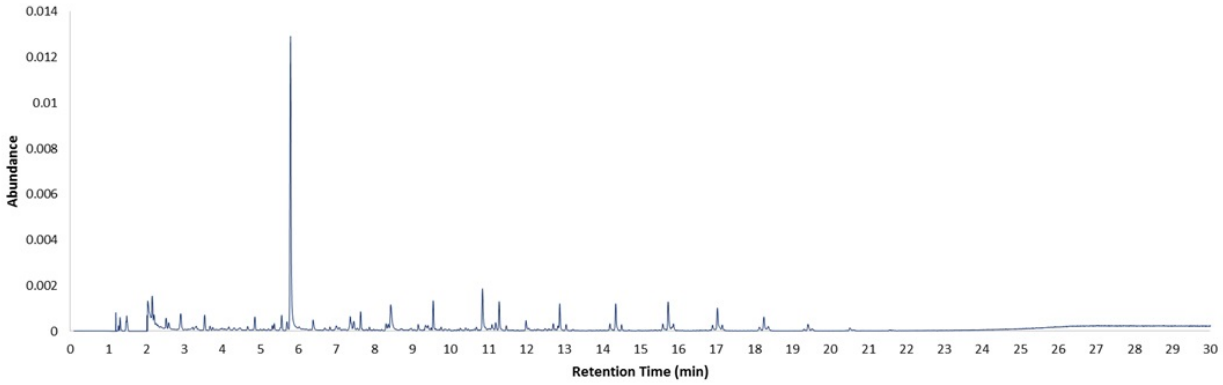


Figure 23: Total ion chromatogram of 1 min MDDM dixie cup

The identified peaks in this chromatogram were 1-tetradecene (15.743 min), 1-pentadecene (17.043 min), 1-dodecene (12.882 min) and 5-methylfurfural (8.447 min).

2.4.2.4. Carbonless Paper

The carbonless paper consists of mainly two layers. The top layer which is undercoated with microencapsulated dye precursor and a reagent layer. When the pressure of the pen is applied on the paper, the undercoated microcapsules break and release the dye precursor. This precursor reacts with the reagent layer to give the final colored product. Typical reagent layers include clays, organic materials or zinc silicylates[30]. One of the interesting observations in the burned and unburned carbonless paper was the significant iso-paraffinic ignitable liquid pattern. Pattern observation in the unburned indicated that the specific pattern was not a result of the pyrolysis of the material. These chromatograms are presented in Figure 24a, 24b, 24c and 24d.

(a)

(b)

(c)

(d)

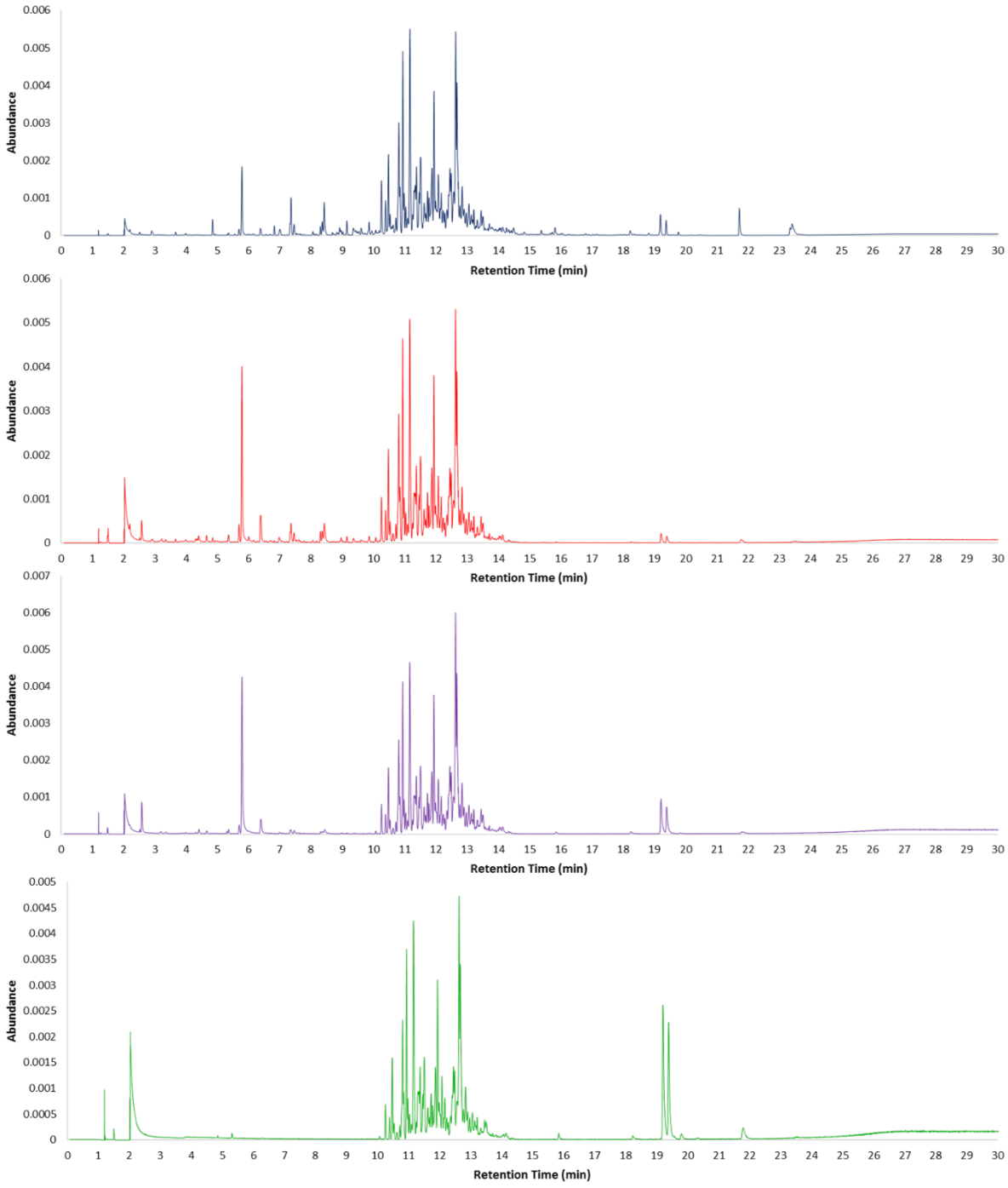


Figure 24: a) 1 min MDDM, b) 1 min DH, 1 min IH and d) unburned total ion chromatograms of carbonless paper

2-Furaldehyde (5.785 min), styrene (7.361 min) and 5-methyl furfural (8.422 min) were only found in the burned samples. The zoomed view of the chromatogram region between 9.5 – 13.5 min is provided in Figure 25. In this, the only identified major compounds were 3-methyl-5-propylnonane (11.172 min) and 2-methyl phenol at 10.246 min.

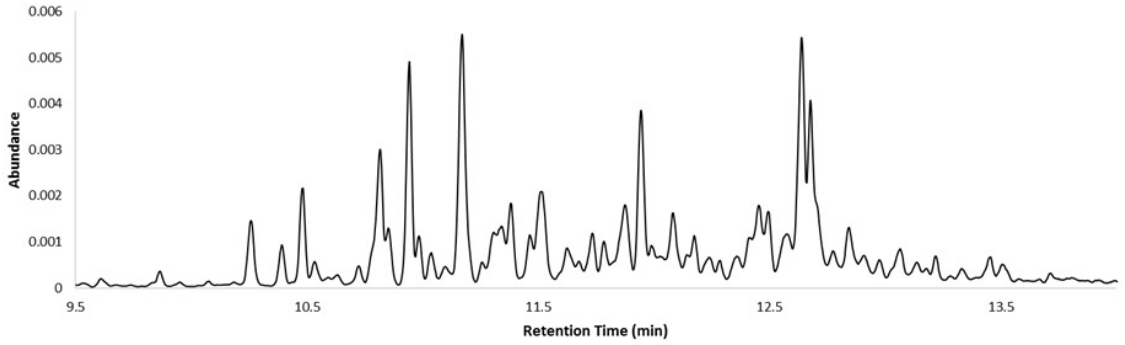


Figure 25: Expanded view of the TIC region of 9.5 – 13.5 min of 1 min MDDM carbonless copy paper

2.4.3. Apparel

Different types of apparel were burned under this category. Some of the examples are leather jacket, cotton shirt, casual shoe (women), new footwear and rain boots. The pyrolysis and combustion products obtained from these substrates are discussed in this section using the total ion chromatograms of the MDDM burns of the materials.

2.4.3.1. Leather Jacket

The highest abundance peak identified in the most of the burns of leather jacket was toluene. The overall abundance of the compounds in the TIC were higher in all MDDM methods relative to the other two methods. Total ion chromatogram of 1 min MDDM burn is given in Figure 26. The identified five major peaks were toluene (4.870 min), benzonitrile (8.805 min), 2-ethyl-1-hexanol (10.002 min), benzene (2.912 min) and 2-ethyl-1-hexene (5.539 min).

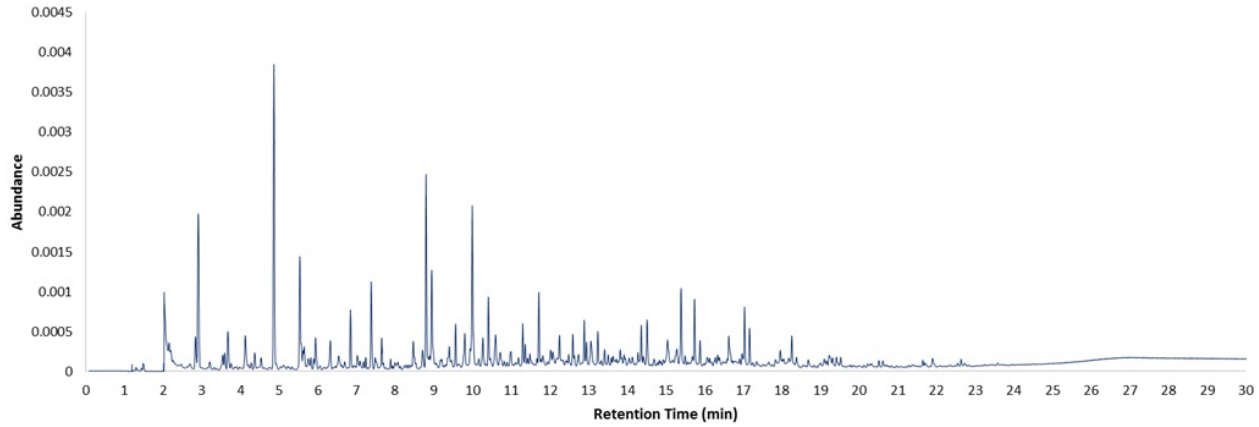


Figure 26: Total ion chromatogram of 1 min MDDM leather jacket

2.4.3.2. Cotton Shirt

Paper products and cotton shirt produced common compounds in the pyrolysis of the materials. These compounds were furfural and 5-methyl furfural. Total ion chromatogram of the 1 min MDDM of the cotton shirt is provided in Figure 37.

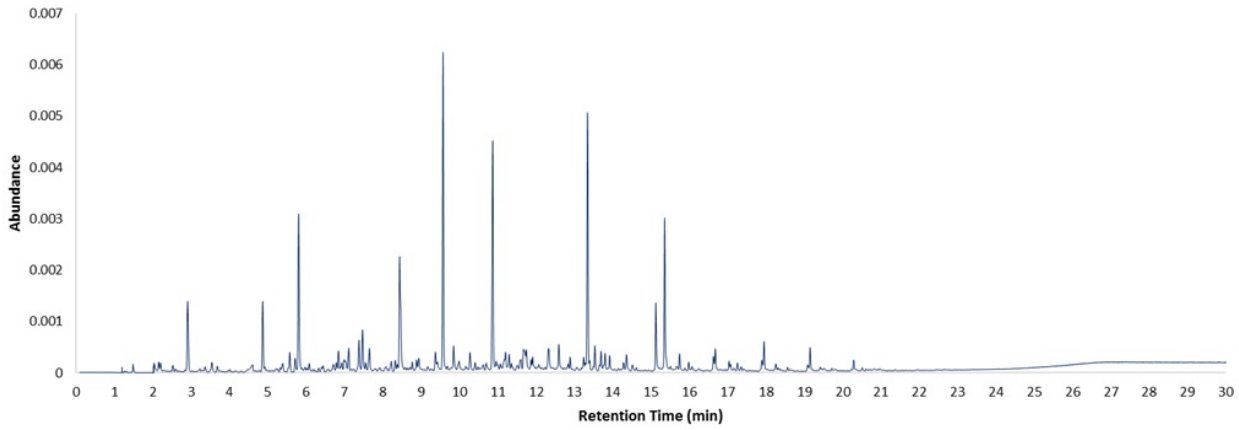


Figure 27: Total ion chromatogram of 1 min MDDM cotton shirt

The major identified peaks of the chromatogram were benzyl chloride (9.575 min), furfural (5.808 min), 5-methyl furfural (8.455 min), benzene (2.911 min) and toluene (4.867 min).

2.4.3.3. Women Casual Shoe

The burned sections of the shoe consisted of the top surface and the sole. As the burn time increased in MDDM, the number of pyrolysis products also increased. The abundance of the compounds in chromatograms produced from DH and IH were drastically different from the MDDM since the loss of products formed in pyrolysis and combustion. The five most abundant peaks in 1 min MDDM of the shoe were styrene (7.374 min), limonene (10.200 min), 1-tetradecene (15.735 min), 1-tridecene (14.358 min) and 1-dodecene (12.881 min). The 1 min MDDM, DH and IH chromatograms are presented in Figure 28a, 28b and 28c respectively.

(a)

(b)

(c)

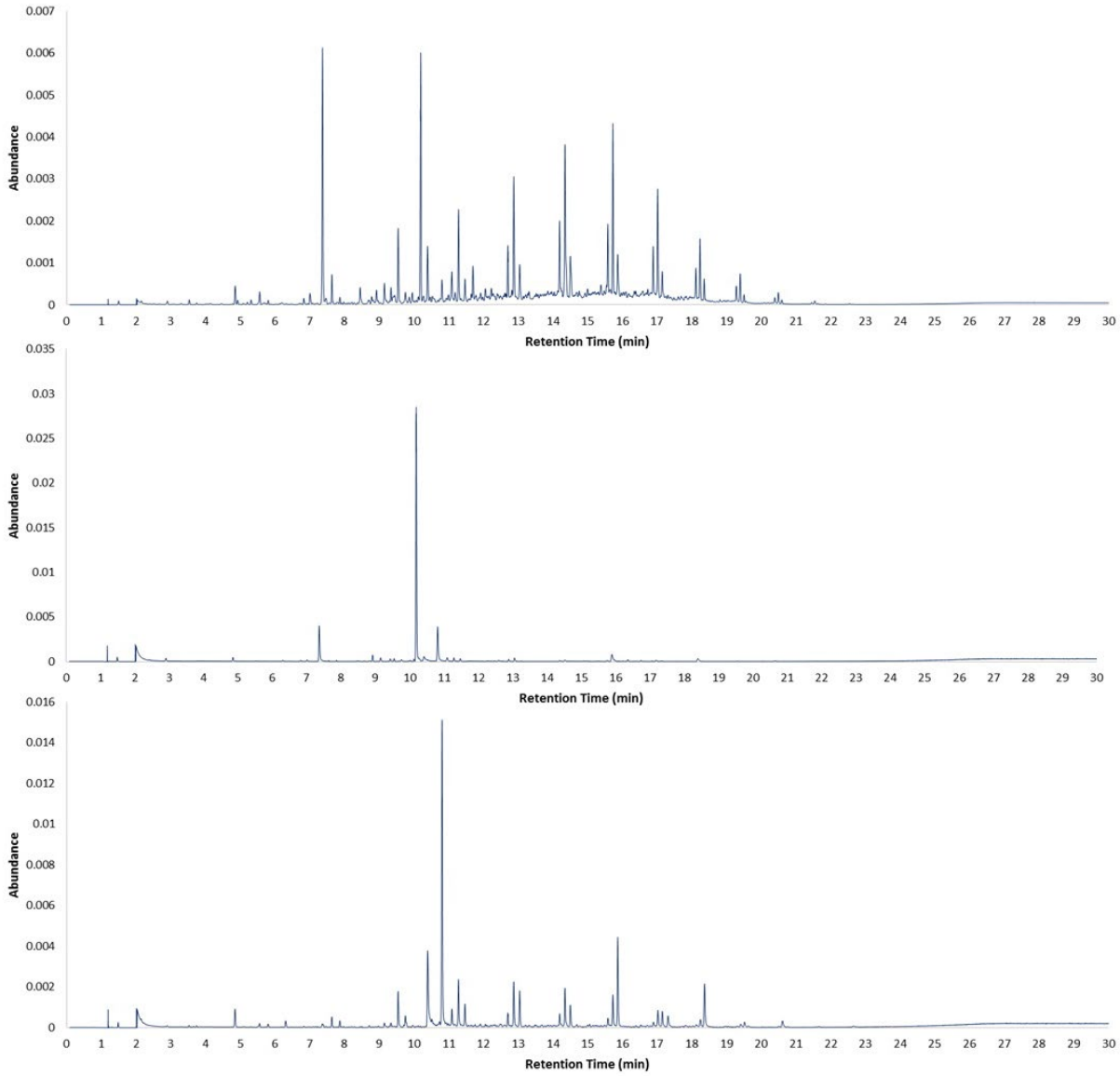


Figure 28: Total ion chromatograms of 1 min a) MDDM b) DH and IH of women casual shoe

In DH and IH burned chromatograms, the most prominent peak was limonene. In IH, the pattern observed in MDDM could be seen but in low abundance. In MDDM, as seen in polyethylene the triplet pattern was clearly observed.

2.4.4. Automobile

The chromatograms of the automobile parts described in this section were worn tire tread, car seat, car mat, dashboard panel and steering wheel panel. The pattern observed in the burned car mat total ion chromatogram was similar to that of a polyester carpet. The patterns observed in the total ion chromatograms obtained for MDDM, DH and IH of the dashboard were different. This will be discussed later in this section.

2.4.4.1. Worn tire tread

The tires are mainly made out of rubber, carbon black and fillers. Mostly the rubber used in tires is a blend of natural and synthetic, which was derived from petroleum-based derivations[31]. The five major compounds that were seen on the total ion chromatogram of 2 min MDDM burned tire (Figure 29) were limonene (10.202 min), styrene (7.373 min), biphenyl (15.385 min), naphthalene (12.586 min) and benzothiozole (13.027 min).

Pyrolysis of rubber forms pyrolysis oil which contains predominant aromatic and terpene products. Mainly benzene, toluenes, styrene, indene and limonene[32]. This explains the presence of a major limonene peak and other aromatic components in the total ion chromatogram.

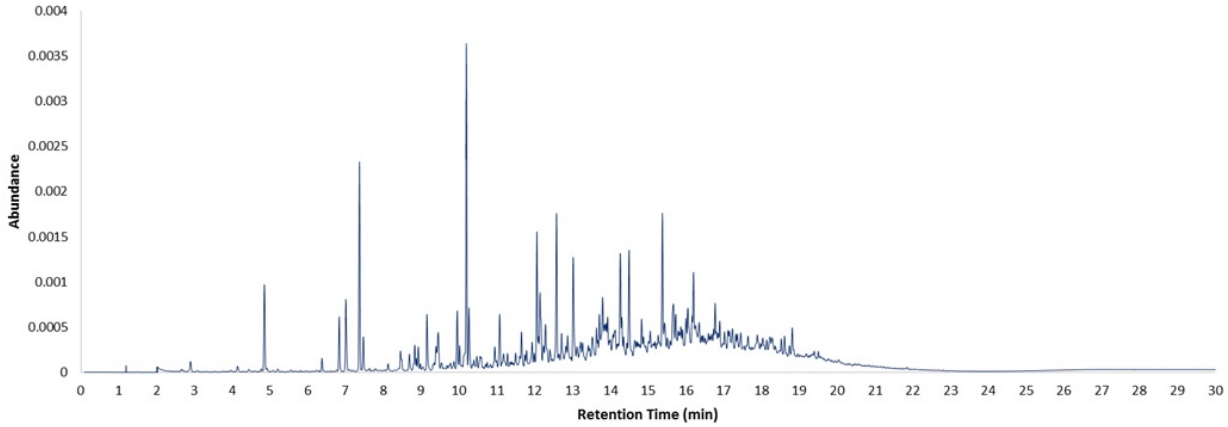


Figure 29: Total ion chromatogram of 2 min MDDM worn tire tread

2.4.4.2. Dashboard

The dashboard was primarily made out of PVC blended with a block polymer which was made of acrylonitrile, butadiene and styrene. The five most abundant peaks in 2 min MDDM were 2,4-dimethyl-1-heptene (6.646 min), 2,4,6,8-tetramethyl-1-undecene (heterotactic) (14.632 min), 2,4,6,8-tetramethyl-1-undecene (isotactic) (14.872 min), 2,4,6-trimethyl-1-nonene (racemic form) (11.143 min) and styrene (7.359 min). The relative abundance of these compounds varies in the different methods. The total ion chromatograms of 2 min MDDM, DH and IH are presented in Figure 30a, 30b and 30c respectively.

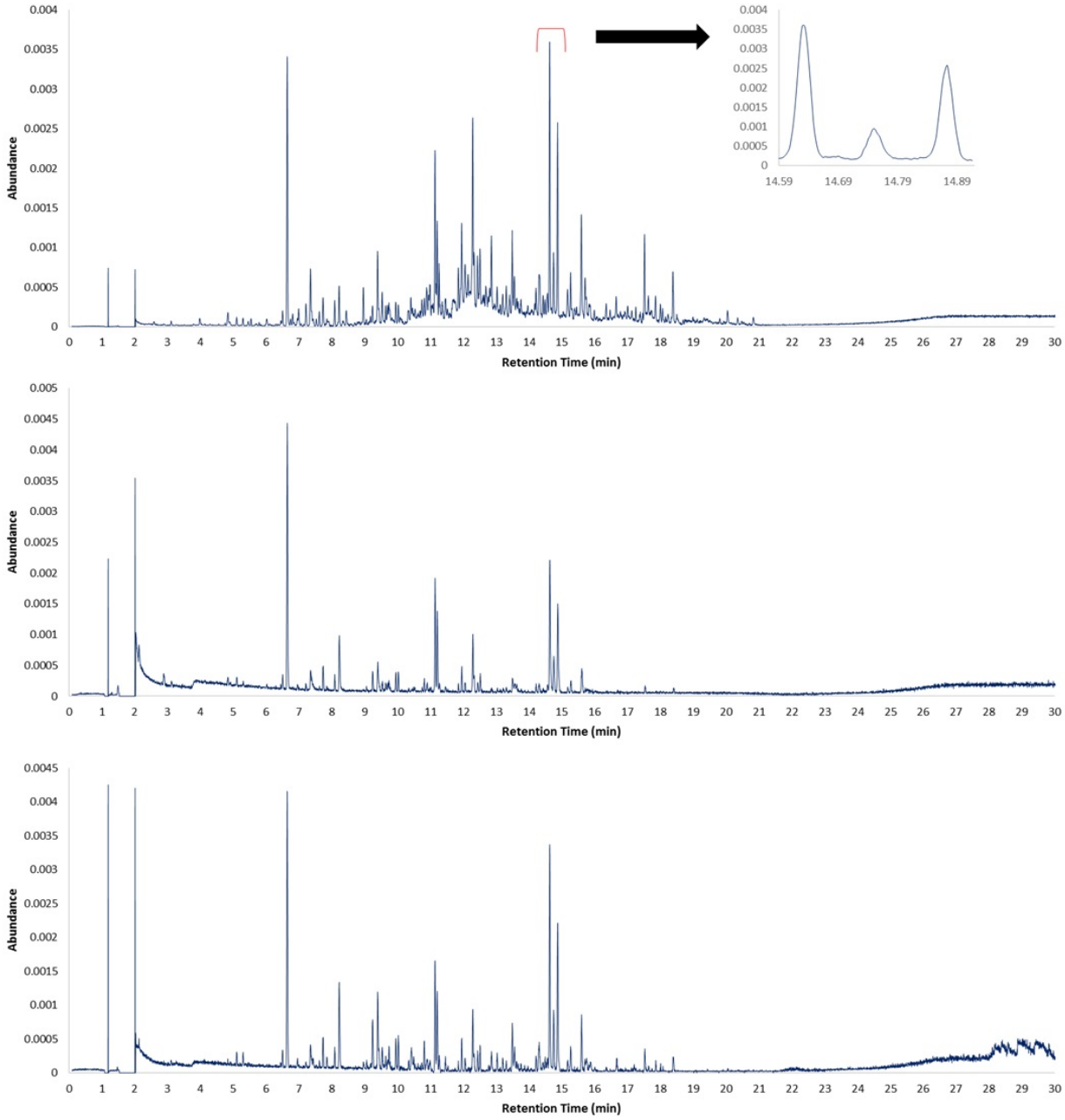


Figure 30: Total ion chromatograms of 2 min a) MDDM b) DH and c) IH of dashboard

As seen in previous burns, the total number of the compounds were higher in MDDM method than DH or IH. However, 2,4-dimethyl-1-heptene has a higher abundance in DH method than in MDDM.

2.4.5. Miscellaneous

In the miscellaneous section, 12 types of various products were included. They were, a cotton towel, plastic drop cloth, shop towel, duct tape, plastic clothesline, manila rope, jute rope, railroad tie, cell phone case, bubble wrap, film and packaging foam and yoga mat. But in this section, only the products obtained from railroad tie will be discussed.

2.4.5.1. Railroad Tie

Railroad ties are made out of wood and treated with creosote[33]. Creosote is a composition of polycyclic aromatic hydrocarbons (PAH), phenolic compounds and heterocyclics[34]. Due to this reason, the pyrolysis products of railroad ties contained many carcinogens such as anthracene[35]. The significant difference of the burn methods lies in the 5 min MDDM. All the other chromatograms obtained from all the burn methods were nearly similar to each other. Therefore, in this, the product obtained for 5 min MDDM and 2 min MDDM are discussed. This is presented in Figure 31.

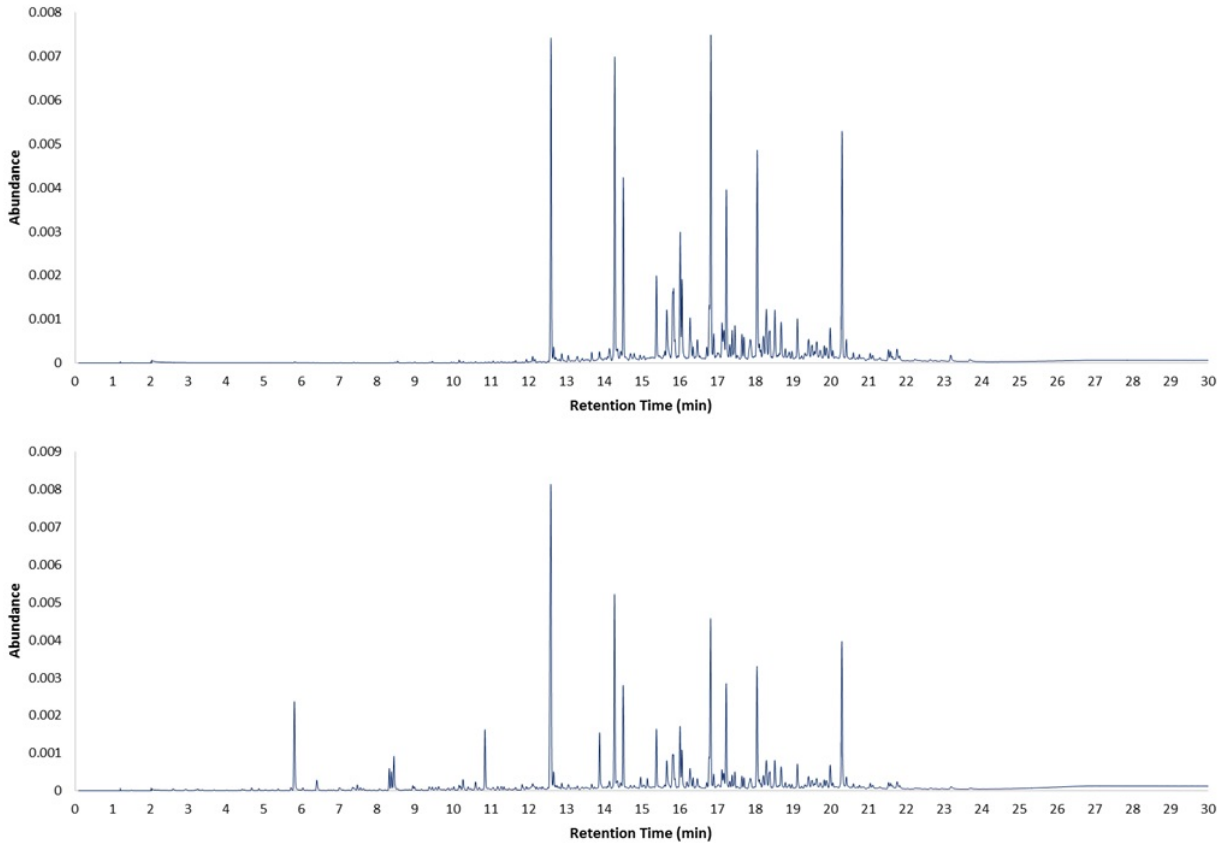


Figure 31: Total ion chromatogram of a) 2min and b) 5 min MDDM of railroad ties

The identified peaks in the 2 min MDDM chromatogram were acenaphthene (16.838 min), naphthalene (12.604 min), 2-methylnaphthalene (14.36 min), anthracene (20.313 min) and fluorene (18.068 min). In 5 min MDDM, in addition to the compounds identified in 2 min MDDM, 2-fufural (5.811 min), 5-methylfurfural (8.448 min) and 2-methoxyphenol (10.860 min) were identified.

2.4.6. Flooring

Different types of flooring materials were burned to obtain the total ion chromatograms. They were carpet, carpet padding, vinyl, engineered, laminate and hardwood flooring. In this section, some of the examples from each section will be discussed.

2.4.6.1. Olefin Carpet

This carpet was made of propylene, bulk continuous filament (BCF) propylene fibers. The products obtained from the pyrolysis of this and combustion of the carpet are illustrated using the total ion chromatogram of 1 min MDDM burn of the carpet. This is presented in Figure 32.

Most of the compounds present in olefin carpets are also similar to that of polyester carpets. The identified top five compounds of this were 2,4,6,8-tetramethyl-1-undecene (isotactic) (14.684 min), TXIB (18.239 min), styrene (7.398 min), biphenyl (15.420 min) and naphthalene (12.610 min).

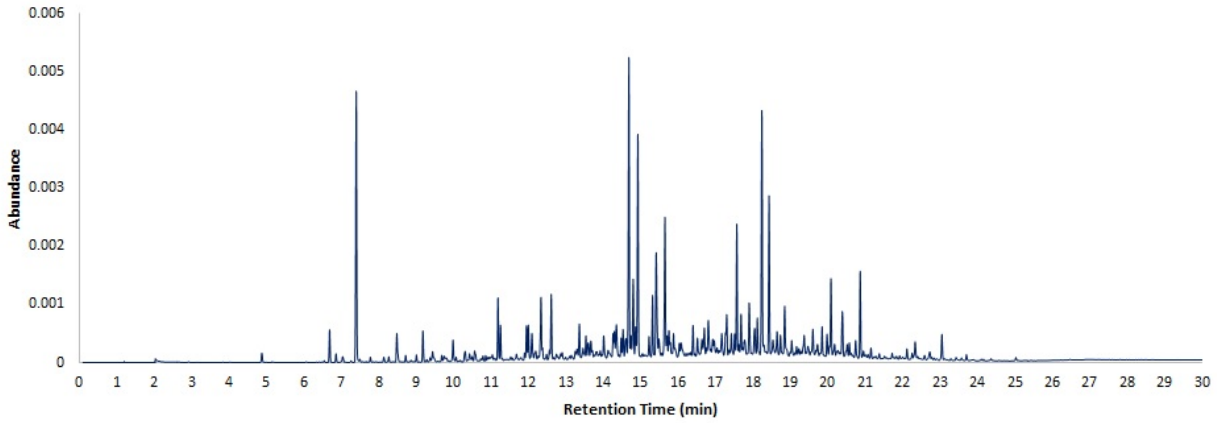


Figure 32: Total ion chromatogram of 1 min MDDM olefin carpet

2.4.6.2. Vinyl Sheet

Vinyl sheet was made out of polyvinyl chloride (PVC). Polyvinyl chloride undergoes side group scission in the pyrolysis process which gives rise to many aromatic products as mentioned in the previous section. In this, 2 min MDDM produced many pyrolysis products when compared with other burn methods. The total ion chromatogram of 2 min MDDM is given in Figure 33.

Identified two major peaks from the chromatogram were TXIB at 18.247 min and 2-chloroethylbenzoate at 15.522 min. In all the other burns, only TXIB peak was present in the respective chromatograms.

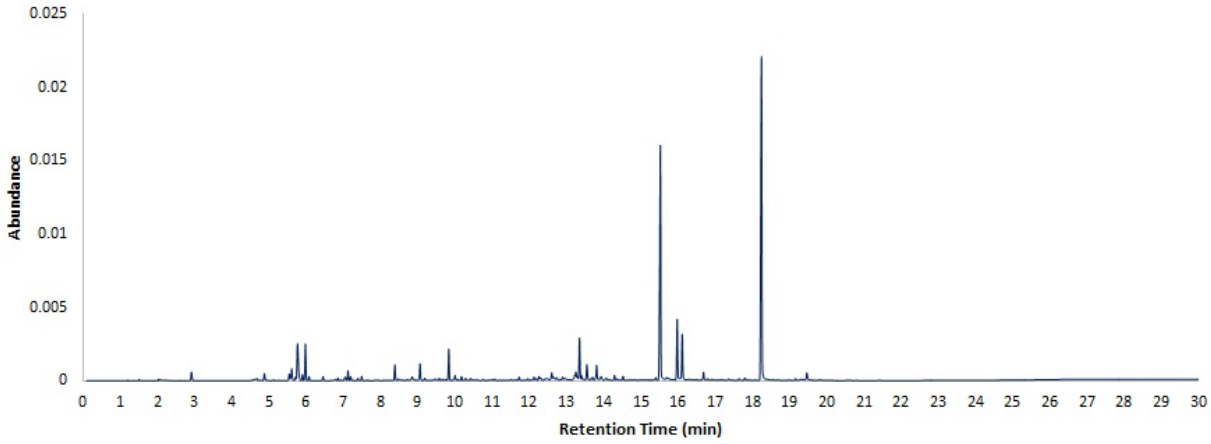


Figure 33: Total ion chromatogram of 2 min MDDM vinyl sheet

2.4.7. Building Materials

As mentioned above, the used building materials for the burns were roofing, insulation, wood, drywall, siding, concrete/masonry, adhesives, ceiling, composite decking, moldings/trim, particle/fiberboard, PVC pipe and pre-treated wood. In this section, only the 2 min MDDM chromatograms of roof shingles, liquid nail heavy duty construction adhesive and PVC pipe are discussed.

2.4.7.1. Poly Vinyl Chloride (PVC) Pipe

As mentioned earlier in this chapter, PVC undergoes side group scission to produce aromatic compounds. In the 2 min MDDM chromatogram of PVC pipe this was clearly observed. The major identified peaks in this chromatogram were benzene (2.910 min), toluene (4.864 min), naphthalene (12.593 min), o-xylene (7.481 min) and 1-methylnaphthalene at 14.279 min. The chromatogram of the 2 min burn of PVC pipe is given in Figure 34.

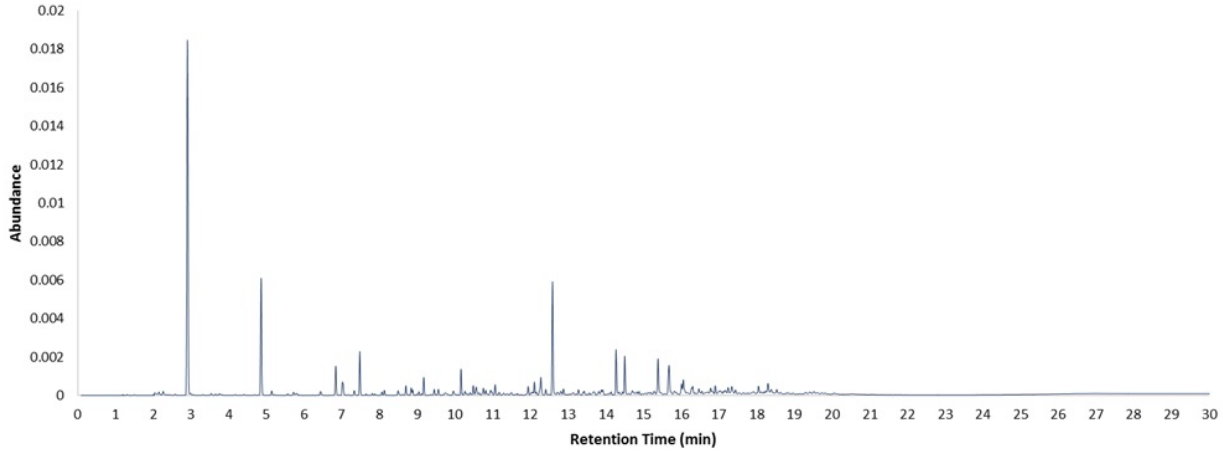


Figure 34: Total ion chromatogram of 2 min MDDM burn poly vinyl chloride (PVC) pipe

2.4.7.2. Adhesives

The main ingredients of this adhesive were kaolin, light petroleum distillates, limestone, cyclohexane, n-hexane and titanium dioxide. The five major identified peaks of this chromatogram were indene (10.255 min), styrene (7.359 min), 2,6-ditertbutyl-4-methylphenol (17.133 min), alpha-methylstyrene (9.146 min) and naphthalene at 12.65 min. The chromatogram of the 2 min MDDM of this adhesive is provided in Figure 35.

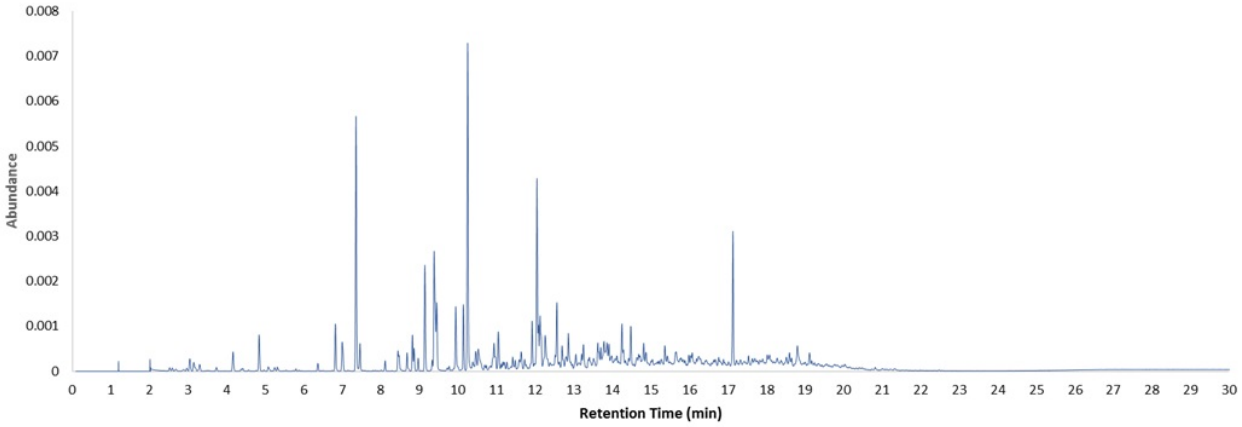


Figure 35: Total ion chromatogram of 2 min MDDM heavy duty construction adhesive

2.4.7.3. Roof Shingles

Roof shingles burned in this produced a pattern which was similar to high-density poly ethylene (HDPE). This can be seen between 8 to 20 min span in the chromatogram. The major compounds identified in this were 1-decene (9.556 min), 1-tridecene (14.366 min), n-tridecane (14.524 min), 1-nonene (7.655 min) and n-tetradecane at 15.893 min. The chromatogram of the roof shingle is depicted in Figure 36.

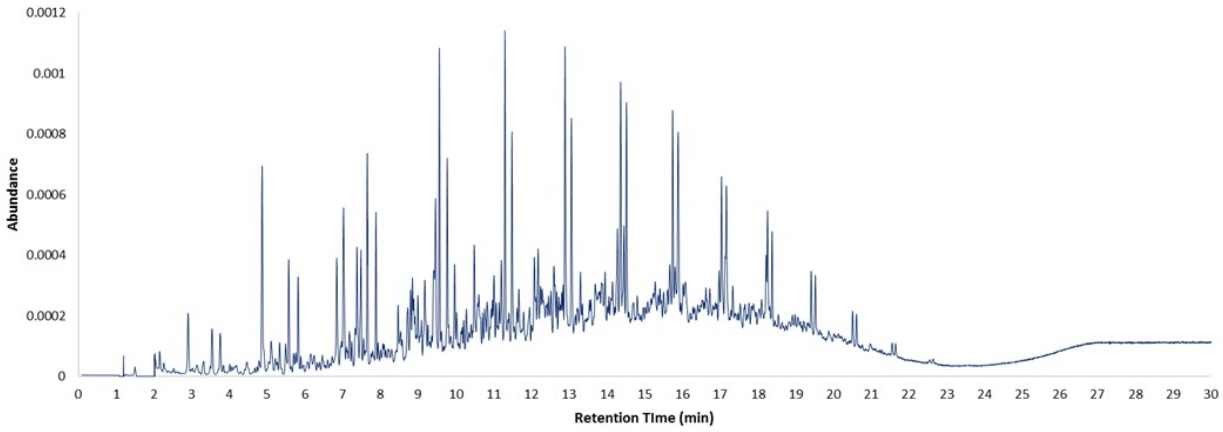


Figure 36: Total ion chromatogram of 2 min MDDM roof shingles

The data analysis and the calculation of the frequency of occurrences of compounds in substrate and ignitable liquids will be discussed in Chapter 3.

CHAPTER 3: LOGISTIC REGRESSION ANALYSIS

In this chapter, the application of Automated Mass Spectral Deconvolution and Identification System (AMDIS) in identifying compounds in the total ion chromatograms of substrates and ignitable liquids is discussed. In a later section of this chapter, the logistic regression analysis of the compounds identified in substrates and ignitable liquids, utilization of receiver operating characteristic (ROC) curves in data analysis and the application of Good – Turing estimation in calculations will be discussed.

3.1. Analysis of the Burn Substrates and Ignitable Liquids

3.1.1. Calculation of Frequency of Occurrence of Compounds by Identified Five Major Peaks in Substrates and Ignitable Liquids Databases

As an initial study, an analysis of 5 major peaks identified in substrate and ILRC Databases of National Center for Forensic Science was performed[17]. In the preliminary study 647 ignitable liquids and 106 burned substrates were used. These substrates were prepared by 2 minute MDDM method as discussed in Chapter 2. The criteria of the identification of the peaks in these chromatograms was a high-quality spectral match and the retention time difference of ±0.05 min to a standard library compound. Two independent analysts from the ILRC/Substrates Database committee analyzed the compounds and a third analyst verified the identification of the compounds. The frequency of occurrences of the compounds of IL and SUB were calculated using these identified 5 major peaks. The results of this study will be discussed in Chapter 5.

3.1.2. Automated Mass Spectral Deconvolution and Identification System (AMDIS)

Gas-chromatograph and mass spectral (GC-MS) data of 660 neat ignitable liquid and 1500 substrate samples were analyzed using AMDIS software[21]. The main purpose of this software is to reduce the false positive identification of compounds and increase the reliability of the identified compounds in the total ion chromatograms. In the identification of compounds, this software extracts the individual component spectrum from the GC-MS data files.

The spectrum extraction for this method was based on the model peak method of Dromey *et al.*[36]. A major problem of this model peak method is the inability to extract weak signals. In AMDIS, this was solved by the explicit consideration of signal to noise ratio throughout the analysis. Signal-to-noise-ratio can be considered as a parameter to measure the sensitivity of the GC-MS instrument. It is simply the ratio between the height of the chromatographic peak and the height of the noise[37]. The data analysis in AMDIS proceeds through four steps: noise analysis, component perception, spectrum deconvolution and compound identification. Noise analysis in AMDIS is done by noise factor calculation (N_f) for GC-MS data.

Component perception identifies each individual component in the chromatogram and determines the model peak shape of the relevant component. Spectral deconvolution extracts the better spectrum by fitting the extracted ion chromatograms to the model profile. Finally, compounds in the TIC are identified using the mass spectral data in the standard mass spectral library[38].

3.1.3. Data Analysis of Substrates and Ignitable Liquids

In the analysis settings in AMDIS, the minimum match factor (Net) was set to 80 and the type of analysis was set to "use retention times". The minimum match factor is a parameter which measures the mass spectral similarity between the target compound in the chromatogram and the standard library compound[21]. The medium threshold was selected for the analysis and the data file format was set to the common data format (CDF) for this work.

Resolution, sensitivity and peak shape requirement parameters were set to medium. The substrate and ignitable liquid samples were analyzed using a standard mass spectral library (created by the National Center for Forensic Science) which contained the mass spectral and retention time data of 293 compounds. All substrate and ignitable liquids samples were analyzed using batch analysis. The batch analysis takes only net factor into consideration and every peak identified in each chromatogram contains 3 hits based on the net factor.

The identified compounds of substrates and ignitable liquids sample chromatograms were then subjected to logistic regression analysis. Logistic regression analysis is a parametric statistical technique and was performed using a parameter derived using the retention time difference between the standard library compound and the identified peak. Logistic regression analysis of the data is explained in detail in the next section.

3.2. Logistic Regression Analysis of Substrates and Ignitable Liquids Data

3.2.1. Logistic Regression

Fire debris analysis is one of the most challenging areas in forensic science. One of the reasons for this is that most of the compounds that are found in ignitable liquids are also produced in pyrolysis and combustion of substrates. Therefore, there is a difficulty in differentiating whether an ignitable liquid was used to start a fire or not. The other important factor is the evaporation of the liquid from the fire debris, which reduces the possibility of identifying ignitable liquid residue in a sample.

The use of statistical methods is one way of approaching this challenging problem. In this section, the calculation of probability in substrates and ignitable liquids using logistic regression is discussed and the calculation of likelihood ratios using Naïve Bayes approach will be discussed in Chapter 4. Naïve Bayes and logistic regression can be defined as classifiers. A classifier is a mathematical function which transforms unlabeled information to labeled using given datasets[39]. Naïve Bayes is considered the simplest form of the Bayesian equation and assumes conditional independence[40]. Logistic regression models can be explained as a mathematical relationship between the predictor variable and a categorical response variable[41]. Logistic regression is used to estimate the probabilities based on the categorical responses given to predictor variables.

For logistic regression of the data, two separate models were created, one for substrates and one for ignitable liquids. The substrate model was generated using 42 substrate samples from the Substrate Database of the National Center for Forensic Science[42]. This model contained data from the Modified Destructive Distillation Method (MDDM) 2 minute burns and unburned

samples. The ignitable liquids logistic regression model was generated using 42 samples of neat ignitable liquids from the Ignitable Liquids Reference Collection (ILRC) Database of the National Center for Forensic Science[43].

AMDIS generates a batch report with 3 hits for each identified peak in the chromatogram. According to the identification of peaks of selected substrates and ignitable liquid samples in the databases, the compounds identified from AMDIS were assigned 1 or 0 which indicated the correct or incorrect identification respectively. A new parameter, S_{RT} was introduced for logistic regression analysis as the predictor variable. This parameter was calculated by Equation 3.1.

$$S_{RT} = \frac{1}{(1+|\Delta RT|)} \quad \textbf{(3.1)}$$

In this equation, ΔRT is the retention time difference between the peak of the sample chromatogram and the standard library compound. Logistic regression calculates the probability of the presence of a compound using S_{RT} parameter. In R, the logistic regression model fitting is performed using a generalized linear model (glm). In a glm model, the response variable is followed by an exponential family distribution, which is a non-linear function[44]. R uses glm function and the family binomial. In logistic regression model fitting, family = "binomial" indicates that it converts the logit function, which is $\log\left[\frac{p}{(1-p)}\right]$ to logistic function. The glm function in R fits the data using maximum likelihood (MLE) estimation. Equation 3.2 is the logit equation used in this analysis, β_0 and β_1 represent the intercept and the coefficient of S_{RT} parameter and P is the probability of the presence of a compound in ignitable liquids (IL) or substrates (SUB). The logistic form of Equation 3.2 is presented in Equation 3.3.

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 S_{RT} \quad \textbf{(3.2)}$$

$$P = \frac{1}{(1+e^{-(\beta_0+\beta_1 S_{RT})})} \quad \textbf{(3.3)}$$

For the identification of compounds, maximum probability threshold cutoff points were determined by plotting sensitivity (true positive rate) and specificity (true negative rate) for substrates and ignitable liquids separately. To determine the TPR and FPR, known ground truth data was used from the IL and SUB databases as mentioned earlier. At this cutoff point, sensitivity and specificity were maximized and fewer number of false positives were identified[45], but at the same time, all the true positives were not identified. The threshold points for substrates and ignitable liquids are given in Figure 37a and 37b respectively. Ignitable liquids and substrates have a probability cutoff of 0.82 and 0.91 respectively. The logistic regression curves with the cutoff points obtained for substrates and ignitable liquid models are illustrated in Figure 38a and 38b respectively.

At the 0.82 cutoff in ignitable liquids, the minimum S_{RT} value was 0.97 and |ΔRT| was 0.034 min, whereas in substrates, at the 0.91 cutoff, the minimum S_{RT} value and |ΔRT| were 0.93 and 0.079 min respectively. The performance of these logistic regression models was evaluated by receiver operating characteristic curves (ROC)[46]. The ROC curves generated for the ignitable liquids and substrate models are given in Figure 39a and 39b respectively.

(a) (b)

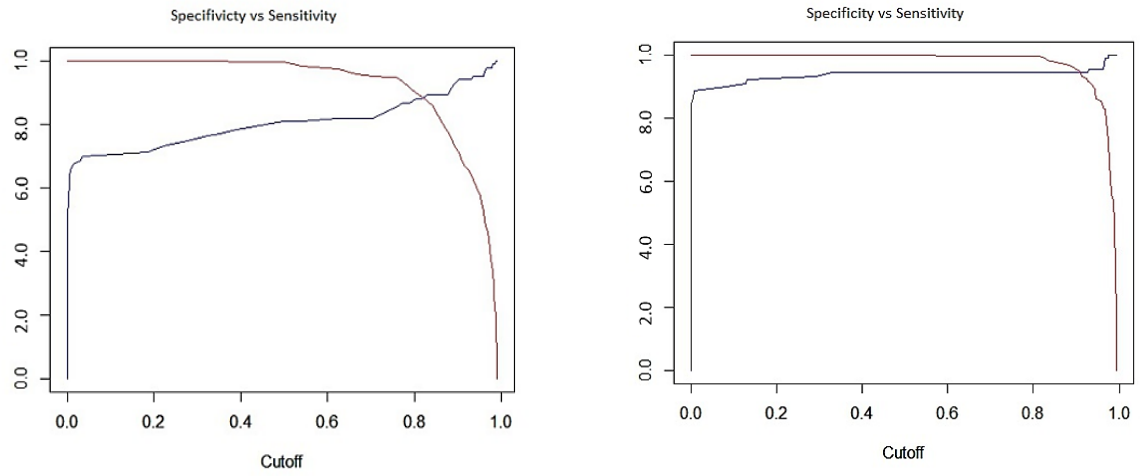


Figure 37: Probability cutoff determination using sensitivity (True Positive Rate) and specificity (True Negative Rate) a) ignitable liquids b) substrates (sensitivity and specificity are plotted in red and blue respectively)

(a) (b)

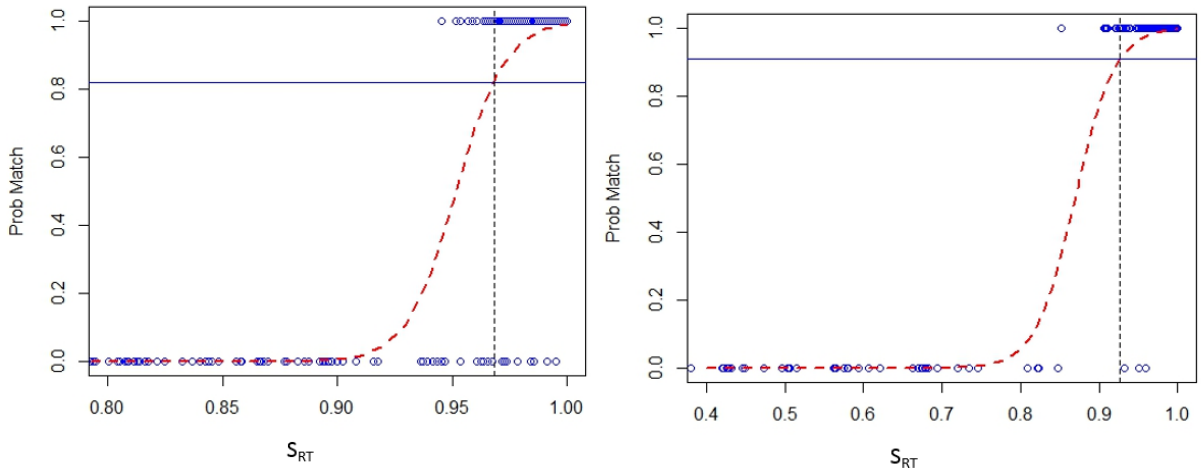


Figure 38: Logistic regression curves a) ignitable liquids b) substrates

The ROC curve is a plot between the true positive rate (TPR) and the false positive rate (FPR). True positive rate is the ratio between the true positives and total positives whereas false positive rate is the ratio between classified false positives and the total negatives. The area under the curve (AUC) of a ROC curve measures the quality of the model. AUC is calculated using the formula in Equation 3.4[47]. In this formula, TP, FP, P and N are true positives, false positives, total positives and total negatives respectively.

$$A_{ROC} = \int_0^1 \frac{TP}{P}\, d\,\frac{FP}{N} = \frac{1}{PN}\int_0^N TP\ dFP \qquad \textbf{(3.4)}$$

If the AUC = 1, then it is considered as a perfect classifier whereas AUC = 0.5 considered as a random classifier. The expected area under the curve of the classifiers in practice should be close to 1. The area under the curve also defines the probability that a randomly chosen positive variable will have a higher score than a randomly chosen negative variable.

In the ignitable liquids model, the area under the curve (AUC) of the ROC curve is 0.95, whereas, in the substrate model, the AUC is 0.99. The AUC of 0.95 in the ignitable liquid model indicates that the IL logistic regression model has a probability of 0.95 that a randomly chosen positive compound to have a higher score than a randomly chosen negative compound in ignitable liquids. The substrate model has an AUC of 0.99 which indicates that the SUB logistic regression model has a probability of 0.99 that a randomly chosen positive compound from a substrate to have a higher score than a randomly chosen negative compound[45].

In a ROC plot, the likelihood ratio (LR) can be calculated by three ways. These procedures are; the tangent of a specific point on the curve, the slope between the origin and the specific point on the curve and the slope between two specific points of the ROC curve[48]. The

method used to calculate the LRs in these models was the second procedure. In the ignitable liquids model, the calculated LR+, which is the slope of the origin to 0.82 probability threshold was 7.5 whereas in substrates model, the calculated LR+ is 17.27 (the slope of the origin to 0.91). Likelihood ratios at these cutoff points are calculated using Equation 3.5.

(a) (b)

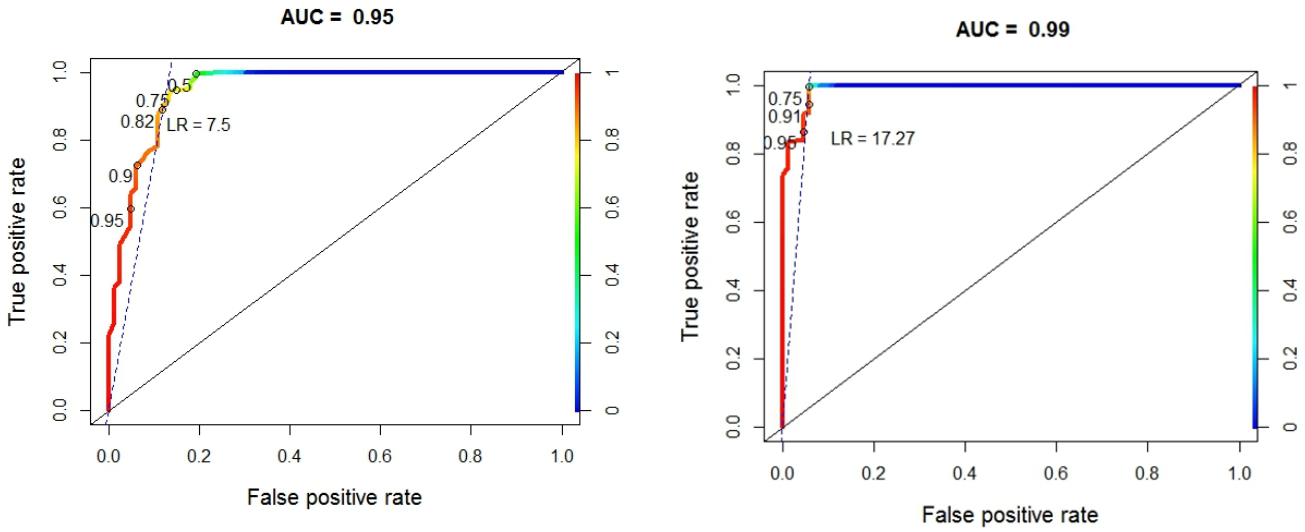


Figure 39: Performance of logistic regression models a) Ignitable liquids b) Substrates

.

$$LR = \frac{True\ Positive\ Rate\ (TPR)}{False\ Positive\ Rate\ (FPR)} = \frac{Sensitivity}{1-Specificity} \qquad \textbf{(3.5)}$$

These probability cutoff points for the calculations were determined by Figure 37a and 37b. In the literature[45], it was explained that at these cutoff points all the true positives are not identified but the identification of false positives is reduced. This was considered as an important factor in the analysis of data for this work.

Based on the presence and absence of compounds in substrates and ignitable liquids, the frequency of occurrences for SUB and IL were calculated for each compound in the standard library using Equation 3.6.

$$Frequency\ of\ occurrence\ of\ Compound\ A = \frac{Number\ of\ Samples\ contain\ compound\ A}{Total\ number\ of\ samples} \quad \textbf{(3.6)}$$

3.2.2. Good-Turing Estimation

When calculating the frequency of occurrences of compounds present in substrates and ignitable liquids, there were some chemical compounds not observed in both substrates and ignitable liquids. This is not an indication that these compounds are absent in SUB and IL in the general population. This frequency (or the probability) of unseen species (not seen in the sample but there is a probability that they exist in the general population) in the samples can be estimated by Good-Turing frequency estimation technique[22].

The initial step in the Good-Turing estimation is the calculation of the frequency of frequencies of appearance for compounds observed in the sample. If the frequency is *r* then the frequency of frequencies is considered N_r. The next important step is the smoothing of the frequency of frequencies, which is performed using Equation 3.7[22].

$$\log Z_r = log\left(\frac{2N_r}{r'' - r'}\right) \quad \textbf{(3.7)}$$

In this Equation, $\log Z_r$ is the log count, r' and r'' are values immediately adjacent to *r*.

Then log count ($\log Z_r$) and log frequency ($\log r$) are plotted to obtain the best fit linear model. The best fit Good-Turing plots obtained for the 2 minute MDDM substrate burns and ignitable liquids are presented in Figure 40 and 41 respectively.

Finally, the total probability of unseen species is calculated using Equation 3.8.

$$Total\ probability\ of\ unseen\ species = \frac{N_0}{N} \quad \textbf{(3.8)}$$

In this equation, N_0 is the frequency of 0 frequencies and N is the total number of compounds in the standard mass spectral library.

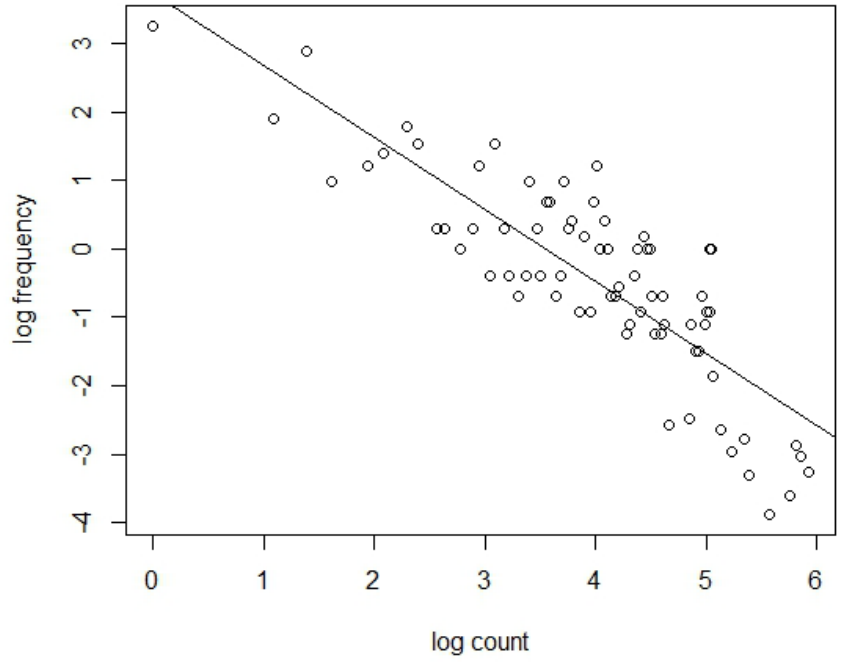


Figure 40: The best fit linear model obtained for 2 min MDDM substrates

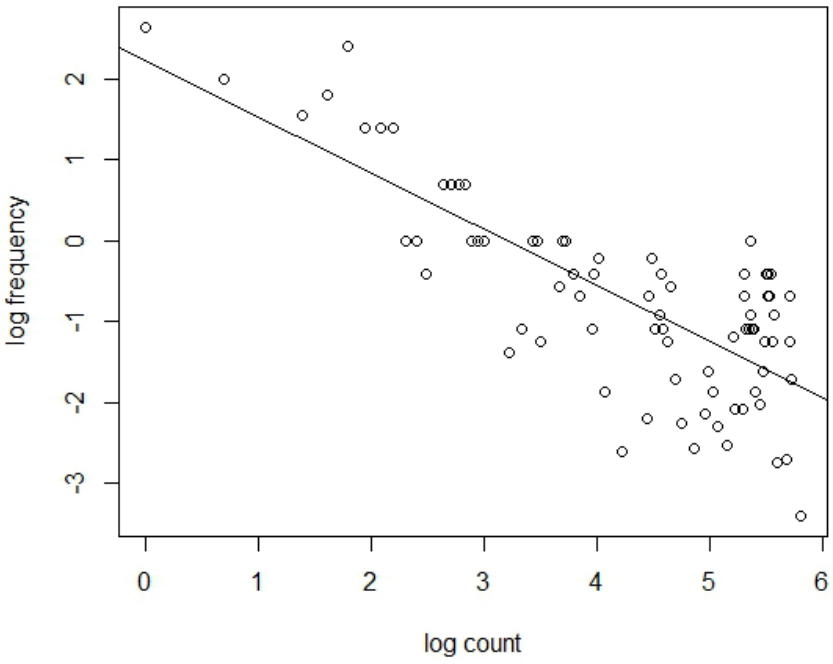


Figure 41: The best fit linear model obtained for ignitable liquids

Laplace estimation of frequencies was also used for these calculations. This will be discussed later in Chapter 6.

CHAPTER 4: CALCULATION OF LIKEILHOOD RATIOS USING NAÏVE BAYES

This chapter discusses the calculation of likelihood ratios (LR) using a Naïve Bayes approach. The calculated LRs are then converted into log-likelihood ratios (LLR). Cross-validation was performed to evaluate the calculated LLRs for ignitable liquids and substrates and the preferred method used in this study was 10-fold stratified cross-validation. The calculated LLRs for test data were calibrated using logistic regression.

4.1. Application of Naïve Bayes Classifier

4.1.1. Naïve Bayes

A classifier in general is defined as a mathematical function which transforms unlabeled information to labeled using a given data set[39]. Naïve Bayes classifier is the simplest form of the Bayesian equation and assumes conditional independence[41]. This assumption reduces the parameters from the original data when modeling the probability of X given Y, P(X|Y). In this study, a Naïve Bayes classifier was used to classify fire debris samples as either positive or negative for the reference of an ignitable liquid residue (ILR).

The conditional independence assumption in probability calculations can be illustrated in Equation 4.1. In this equation, for a given sample containing ignitable liquids, X is the evidence (the probability of compounds observed in the sample) and Y is the proposition where the sample is positive for ignitable liquids (IL). The conditional independence assumption allows each variable to be counted separately.

$$P(X_1 \ldots . X_d|Y) = \prod_{i=1}^{d} P(X_i|Y) \quad \textbf{(4.1)}$$

4.1.2. Calculation of Likelihood Ratios using Naïve Bayes

The presence and absence of major compounds in substrates (SUB) and ignitable liquid (IL) samples are indicated in a data frame using 1 and 0 respectively. The presence or absence of each compound was determined using a R code written in-house and based on the previously discussed theory. A partial screen shot of the substrate data frame is given in Figure 42. Compounds are labeled using the Chemical Abstracts Service (CAS) number.

FileName	67_56_1	75_07_0	64_17_5	75_05_8	78_78_4	67_64_1	107_02_8	123_38_6	67_63_0
1	0	0	0	0	0	0	0	0	0
101	0	0	0	0	0	0	0	0	0
103	0	0	0	0	0	0	0	0	0
105	0	0	0	0	0	0	0	0	0
107	0	0	0	0	0	0	0	0	0
109	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
111	0	0	0	0	0	0	0	0	0
113	0	0	0	0	0	0	0	0	0
115	0	0	0	0	0	0	0	0	0
117	0	0	0	0	0	0	0	0	0
119	0	0	0	0	0	0	0	0	0
121	0	0	0	0	0	0	0	0	0
123	0	0	0	0	0	0	0	0	0
125	0	0	0	0	0	0	0	0	0
127	0	0	0	0	0	0	0	0	0
129	0	0	0	0	0	0	0	0	0
M1002TB5	1	0	0	0	1	1	0	0	0
M1010AS1	0	0	0	0	0	1	0	0	1
M1011AS2	0	0	0	0	0	1	0	0	1
M1012AS5	0	1	0	0	0	1	0	0	0
M1020BS1	0	0	0	0	0	0	0	0	0
M1021BS2	0	0	0	0	0	0	0	0	0
M1022BS5	0	1	0	0	0	1	0	0	0
M1030PR1	0	0	0	0	1	1	0	0	0
M1031PR2	0	1	0	0	0	1	1	0	0
M1032PR5	0	1	0	0	0	0	0	0	0
M1040GB1	0	1	0	0	0	0	1	0	0
M1041GB2	0	0	0	0	0	1	1	0	0
M1042GB5	0	1	0	0	0	1	0	0	0
M1050KT1	0	0	0	0	1	0	0	0	0
M1051KT2	0	0	0	0	0	1	1	0	0

Figure 42: Section of a data frame which includes the presence or absence of compounds

Naïve Bayes likelihood ratios were calculated using the equation given below (Equation 4.2). In this equation, the numerator is the product ($\prod_k P(C_k|H_{IL})$) of the probabilities of observing the compounds present given an ignitable liquid whereas the denominator is the

product ($\prod_k P(C_k|H_{SUB})$) of the probabilities of compounds being present given that the sample comes from pyrolysis of a substrate.

$$LR = \frac{\prod_k P(C_k|H_{IL})}{\prod_k P(C_k|H_{SUB})} \quad \textbf{(4.2)}$$

If a compound was present in a given sample, it was multiplied with the frequency of occurrence of that compound in the SUB and IL. An example for this calculation is provided below.

Example 1: The following compounds were identified in an unknown sample.

Compound	Present/ Absent	Frequency in SUB	Frequency in IL
Toluene	1	0.722	0.361
ethyl benzene	1	0.266	0.199
naphthalene	1	0.503	0.245
2, 4-dimethylhexane	1	1.42E-05	0.164

The calculated numerator for this sample is given below:

$1 \times 0.3608 \times 1 \times 0.199 \times 1 \times 0.245 \times 1 \times 0.164 = 0.00288 = 0.003$

The calculated denominator:

$1 \times 0.722 \times 1 \times 0.266 \times 1 \times 0.503 \times 1 \times 1.42 \times 10^{-5} = 1.37 \times 10^{-6}$

Therefore, the calculated likelihood ratio for this sample $= 0.003/1.37 \times 10^{-6} = 2189.78 = 2.19 \times 10^3$, this positive likelihood ratio indicates that this sample contains ignitable liquid residue (ILR). This calculated LR is then converted to log likelihood ratio (LLR). The calculated LLR for this is 3.34. If the calculated LR < 1, then the calculated LLR will be negative. No matter how large or small the calculated LR or LLR is, the size of the prior odds, will determine the odds of the sample belonging to the IL or SUB class (i.e. the posterior odds). In this study,

calculated likelihood ratios for SUB and IL were converted into log-likelihood ratios (LLR) and were subjected to cross-validation which will be discussed in the next section.

4.2. Cross-validation and Calibration of Log-Likelihood Ratios

4.2.1. Cross-validation (CV)

The performance of any evidence evaluation method is measured by simulating a number of hypothetical events. This provides an idea of how this method will perform when applied to the real casework data. This is an empirical approach because it is based on the observations from experimental set-ups[49]. This procedure is called validation.

Validation of a procedure is divided into two categories. They are method development and validation stages[50]. This chapter is focused on the method development and validation stages are explained in Chapter 6. In this study 10-fold stratified cross validation was performed as a part of the method development. Each fold of the CV places 90% of the data (SUB 90% and IL 90%) in the training set and 10% (SUB 10% and IL 10%) in the testing set. In each fold different training and test data set were selected.

For the cross-validation, 522 substrate samples created by Modified Destructive Distillation Method (MDDM) and 642 neat ignitable liquid samples were used. All SUB and IL data were divided into two sets; training and test data. Parameters for the selected LR method are calculated using the training data set. In this study, the training data set was used to calculate the frequency of occurrences of the compounds present in SUB and IL. Likelihood ratios were then calculated for the test data using this frequency of occurrences. Training and test data contain

45% and 55% of SUB and IL respectively. Cross-validation was performed using an R code developed for this method.

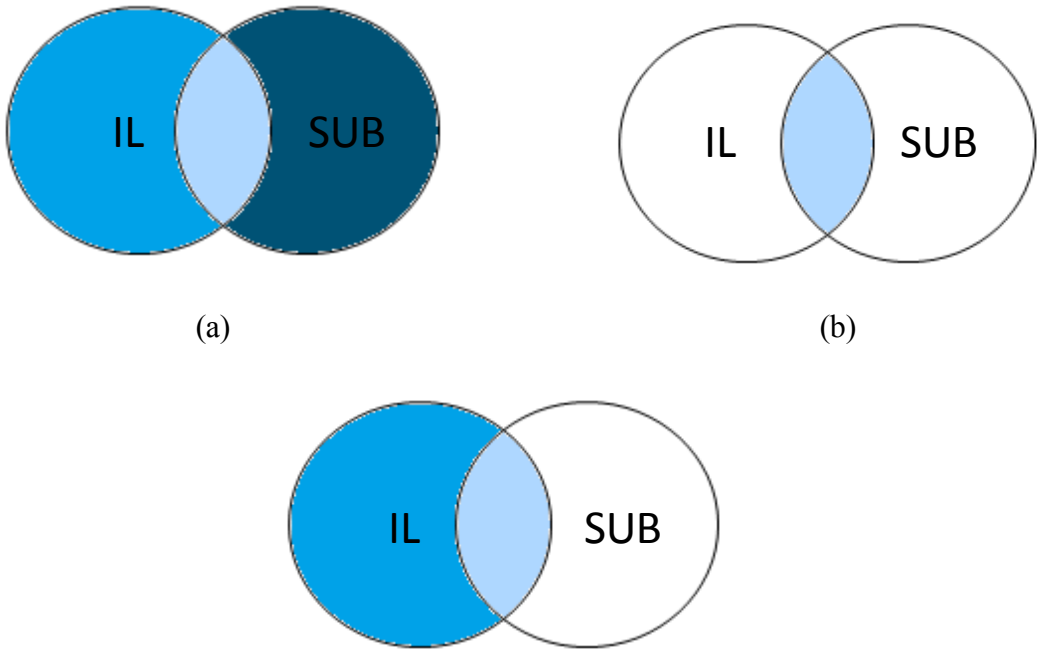
Training data: SUB – 470 (45%), IL –579 (55%)

Test data: SUB – 52 (45%), IL – 64 (55%),

Likelihood ratios were calculated using three sets of compounds. The sets of compounds were,

a) All compounds include in SUB and IL
b) Compounds common to SUB and IL
c) Compounds present in IL

These are graphically illustrated by Venn diagrams in Figures 43a, 43b and 43c respectively.



(a) (b)

(c)

Figure 43: Sets of compounds used to calculate the likelihood ratios.

4.2.1.1. Cross-validation of LLRs using All Compounds Present in SUB and IL

There was a total of 251 compounds seen in both the substrates and ignitable liquids. These 251 compounds were used to calculate the likelihood ratios for this method. There were 139 compounds common to SUB and IL, while 75 and 38 compounds can be seen only in SUB and IL respectively. The performance of the calculated LLRs was evaluated using *receiver operating characteristic curves* (ROC), *detection error trade-off* (DET), *empirical cross entropy* (ECE), *tippet* plots and *histograms*.

As mentioned earlier, a ROC curve is a plot between the true positive rate (TPR) and the false positive rate (FPR). In this plot, TPR indicates the fraction of ground truth positive samples that were correctly classified as ignitable liquids and the FPR is the fraction of ground truth positive samples for substrates that were incorrectly identified as positive for ignitable liquids. The area under the curve (AUC) of this plot is 0.99. This means that there was a 0.99 probability that a randomly chosen sample will be classified as an ignitable liquid. This ROC curve was generated from the calculated log-likelihood ratios (LLR) for test data and given in Figure 44.

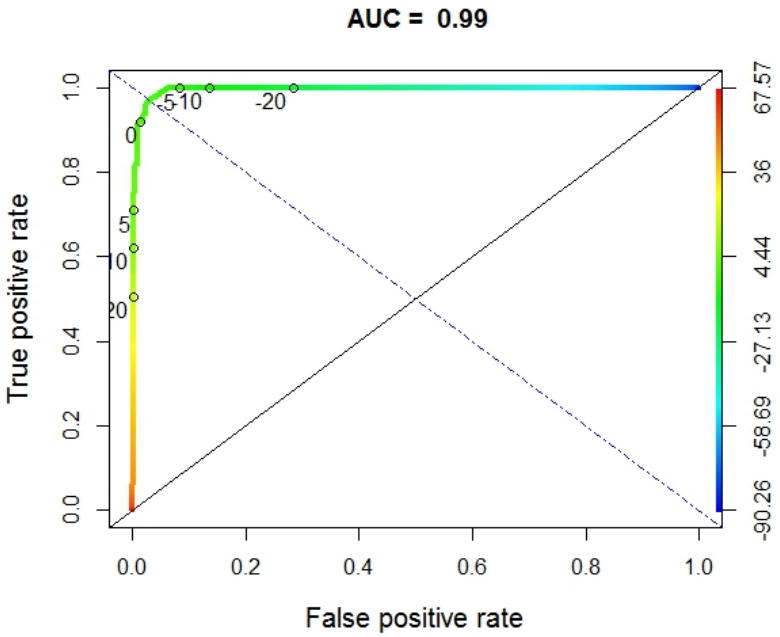


Figure 44: ROC curve obtained for calculated LLRs using compounds present in SUB and IL

Discriminating power of the calculated LLRs are illustrated using DET plots. Discriminating power can be explained as the ability to distinguish two hypotheses using the calculated LLRs. These two hypotheses are, H_1: the calculated log-likelihood ratios of ignitable liquids and H_2: the log-likelihood ratios calculated for substrates. The DET plot obtained for this method is given in Figure 45.

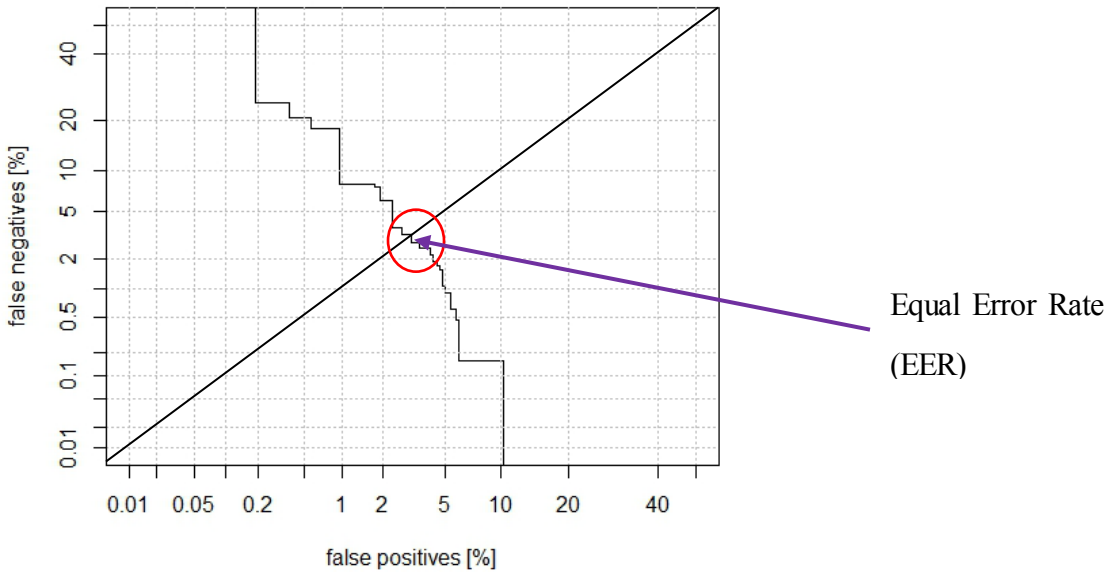


Figure 45: DET plot obtained for calculated LLRs using compounds present in SUB and IL

In a DET plot, if the curve is closer to the origin, it indicates that the calculated LLRs have a high discriminating power whereas the curve is further from the origin is an indication that the calculated LLRs have low discriminating power. In this case, the curve is closer to the origin which indicates that the calculated LLRs using this method have high discriminating power. Equal error rate (EER) is the point where the false negative percent and false positive percent are equal. This is indicated in Figure 45 and the EER for these calculated LLRs is 3%.

Even though the LLRs have a higher discriminating power, the model is not completely appropriate for evidence evaluation because in some cases tippet plots indicate that the calculated LLRs support the wrong decisions. This is related to the critical performance of the model. Consideration of the critical performance characteristics of the model is called calibration[49]. Calibration of the model assures the consistency and the reliability of the calculated likelihood ratios which lead to better decision making procedures[51]. Calibration of the calculated

LLRs can be explained using empirical cross entropy (ECE) plots[49-52]. The empirical cross-entropy function can be defined as the average of the weighted logarithmic scoring rule[49] and this represents the accuracy. These equations[49] are presented below.

$$ECE = -\frac{P(H_1|I)}{N_1}\sum_{i:H_1 is\ true} log_2 P(H_1|E_i, I) - \frac{P(H_2|I)}{N_2}\sum_{j:H_2 is\ true} log_2 P(H_2|E_i, I) \quad \textbf{(4.3)}$$

In Equation 4.3, H_1 and H_2 are the hypotheses where the ignitable liquid residue (ILR) present or absent in the sample respectively. E_i and E_j denote the evidence in each case in the validation data set and N_1 and N_2 are the numbers of samples containing ILR and number of samples without ILR. This equation can be re-written using the prior odds and is presented in Equation 4.4. **(4.4)**

$$ECE = -\frac{P(H_1|I)}{N_1}\sum_{i:H_1 is\ true} log_2\left(1 + \frac{1}{LR_i \times O(H_1|I)}\right) + \frac{P(H_2|I)}{N_2}\sum_{j:H_2 is\ true} log_2\left(1 + LR_j \times O(H_1|I)\right)$$

In this equation, $O(H_1|I) = \frac{P(H_1|I)}{P(H_2|I)}$ indicates the prior odds in favor of the hypothesis H_1 and in this case, LR_i and LR_j are the calculated log-likelihood ratios for samples with ILR and without ILR respectively.

The ECE plot obtained for the calculated LLRs is given in Figure 46. The blue dashed curve is the calibrated accuracy of LLRs using the Pool Adjacent Violators (PAV) method and the solid red curve is the experimental LLRs which explains the accuracy of the calculated LLRs. The discriminating power of LLRs was obtained by the blue dashed curve[52]. The lower the red curve, the more accurate the method of the calculation of LLRs. The dotted curve of the ECE

plot represents the neutral reference, and the accuracy curve (solid red curve) should always be lower than the neutral curve.

If the red and blue lines of the plot were adjacent to each other, then the calibration of the calculated LLRs would be better. However, in this case, the two lines are further apart, which is an indication for poor calibration of the method. The calibration of the LLRs was performed using logistic regression, which will be discussed later in the chapter.

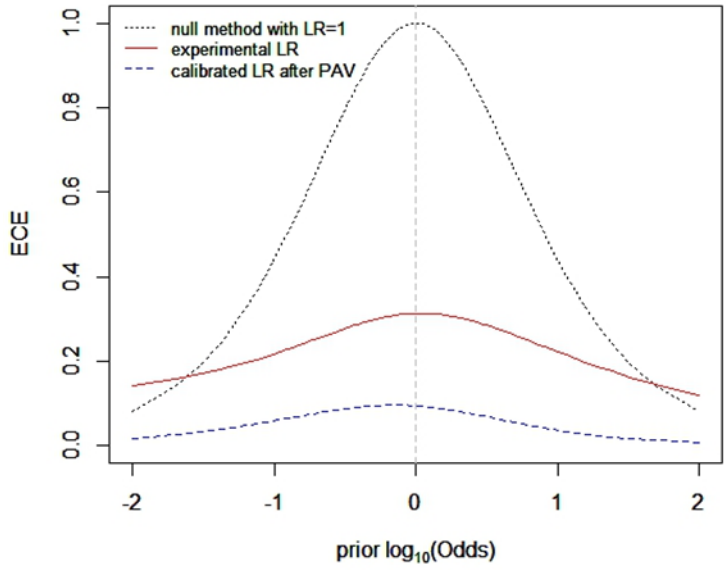


Figure 46: ECE plot obtained for calculated log-likelihood ratios

A tippet plot is another form of interpreting the discriminating power of the LLRs. In a tippet plot, the discriminating power of calculated LLRs is determined by the separation of the two curves. If the separation is large, the calculated LLRs have high discriminating power and provides strong support for the evidence while the support of evidence is less if the two curves are close to each other. The tippet plot obtained for this method is given in Figure 47.

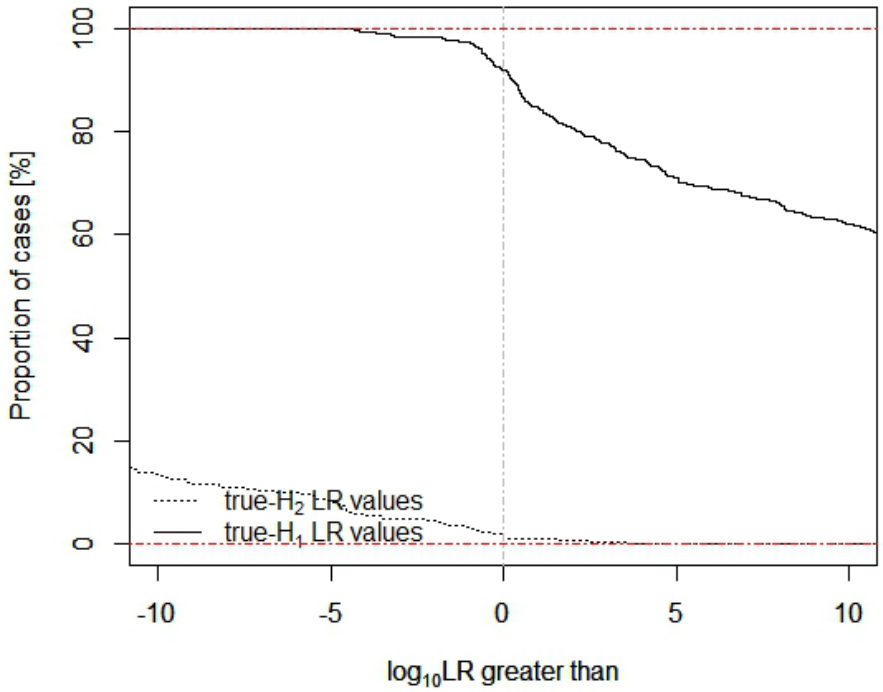


Figure 47: Representation of the discriminating power of calculated LLRs using tippet a tippet plot

A tippet plot is a graphical representation of LLR on the *x*-axis, against the proportion of cases in the *y*-axis. In the tippet plot, H_1 is the hypothesis for calculated log LLRs for substrates, whereas H_2 is the hypothesis for calculated LLRs for ignitable liquids.

Distribution of the calculated LLRs for SUB and IL can be interpreted using a histogram. The distribution of the calculated LLRs are given on the *x*-axis and the frequency of LLRs are given on the *y*-axis. This is illustrated below (Figure 48).

(a)

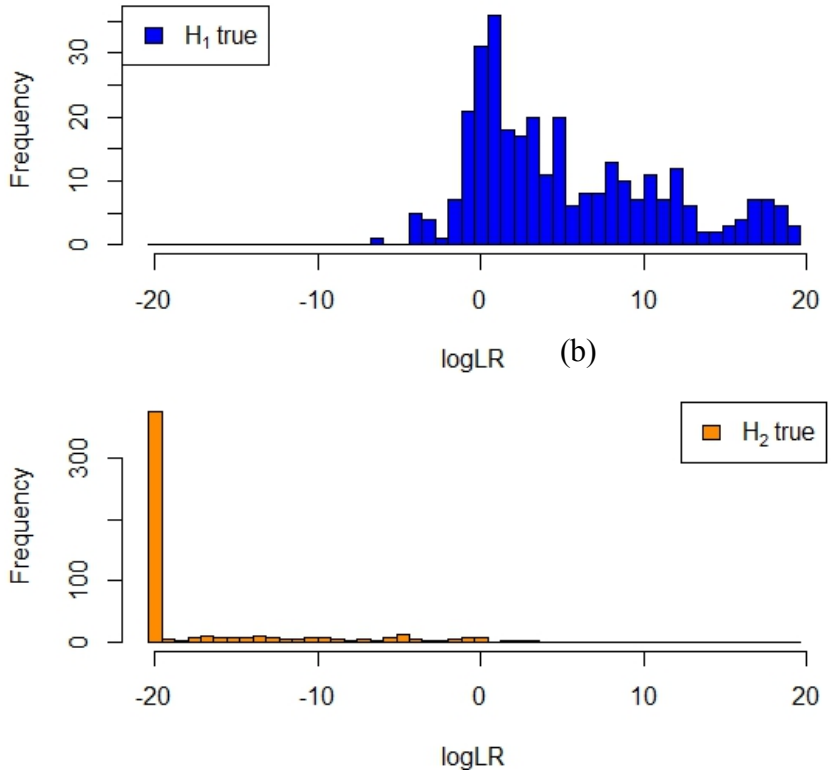


Figure 48: Distribution of calculated LLRs a) IL b) SUB

In ignitable liquids, calculated LLRs have a higher frequency in the range of 0 – 1 whereas in substrates the highest frequency of LLRs is in -20 and above. The DET, ECE, tippet and histograms were obtained by the R code provided in the book, *Statistical Analysis in Forensic Science: Evidential Value of Multivariate Physicochemical Data*[49].

4.2.1.2. Cross-validation of LLRs using Compounds Common to SUB and IL

There were 139 compounds identified in both SUB and IL. In this, cross-validation for the calculated likelihood ratios was determined using the frequency of occurrences of these compounds in SUB and IL. The performance of this method was explained using ROC, DET, ECE, tippet and histograms as described above and these are illustrated in Figure 49, 50a, 50b, 51a and 52b respectively.

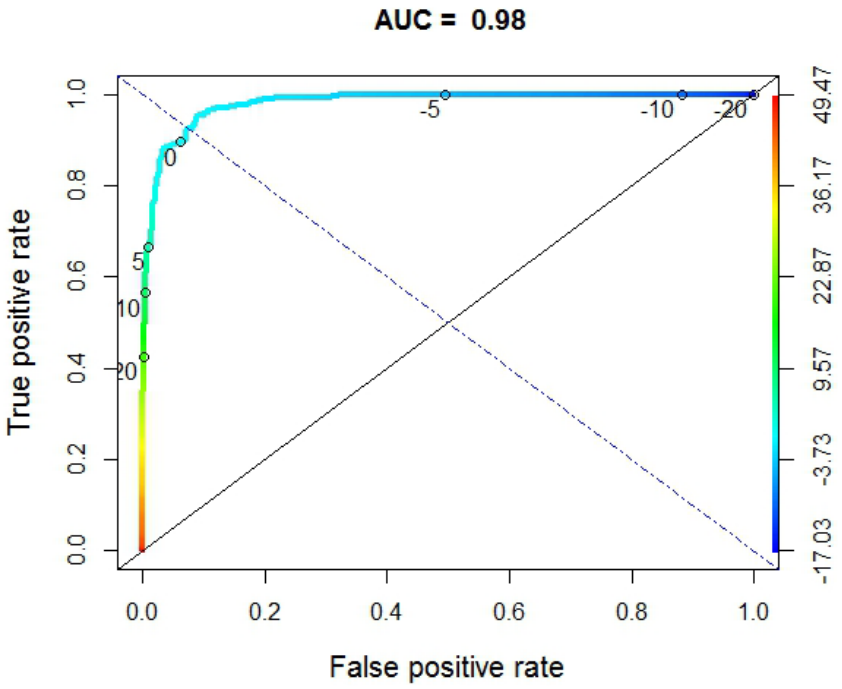


Figure 49: ROC curve obtained for calculated LLRs using common compounds in SUB and IL

The calculated AUC of this ROC plot is 0.98. This indicates that the calculated LLRs have a near perfect separation between substrates and ignitable liquids.

(a) (b)

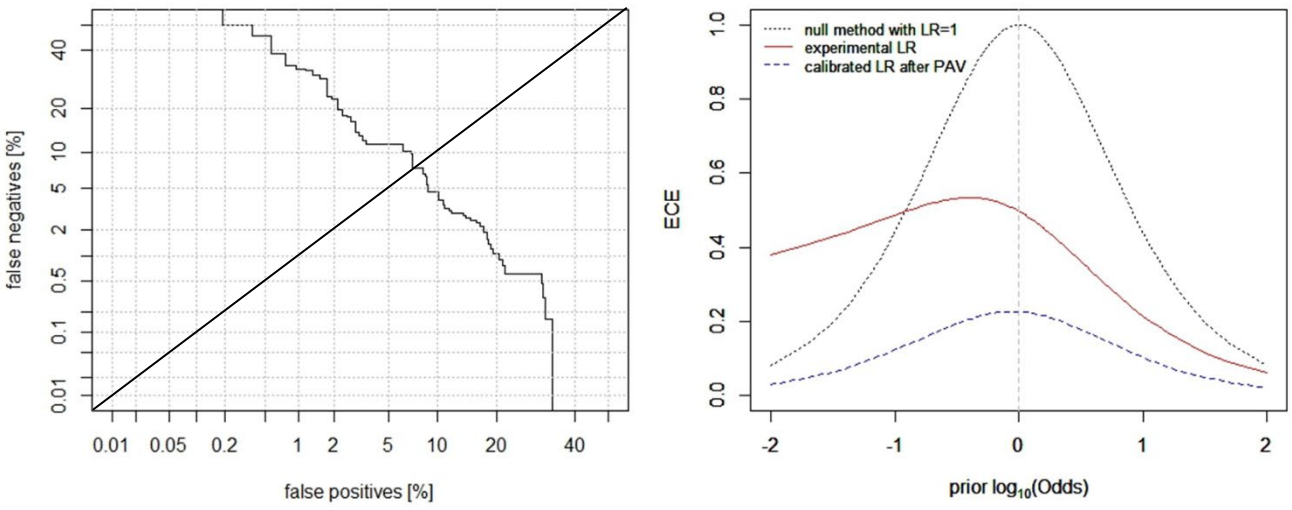


Figure 50: (a) DET plot and (b) ECE plot obtained for calculated LLRs

In the DET plot, the EER was equal to 7.5%, which was higher than the previously reported method. Also, in the ECE plot, the red line is further apart from the dashed line, which indicates that this method was poorly calibrated.

(a) (b)

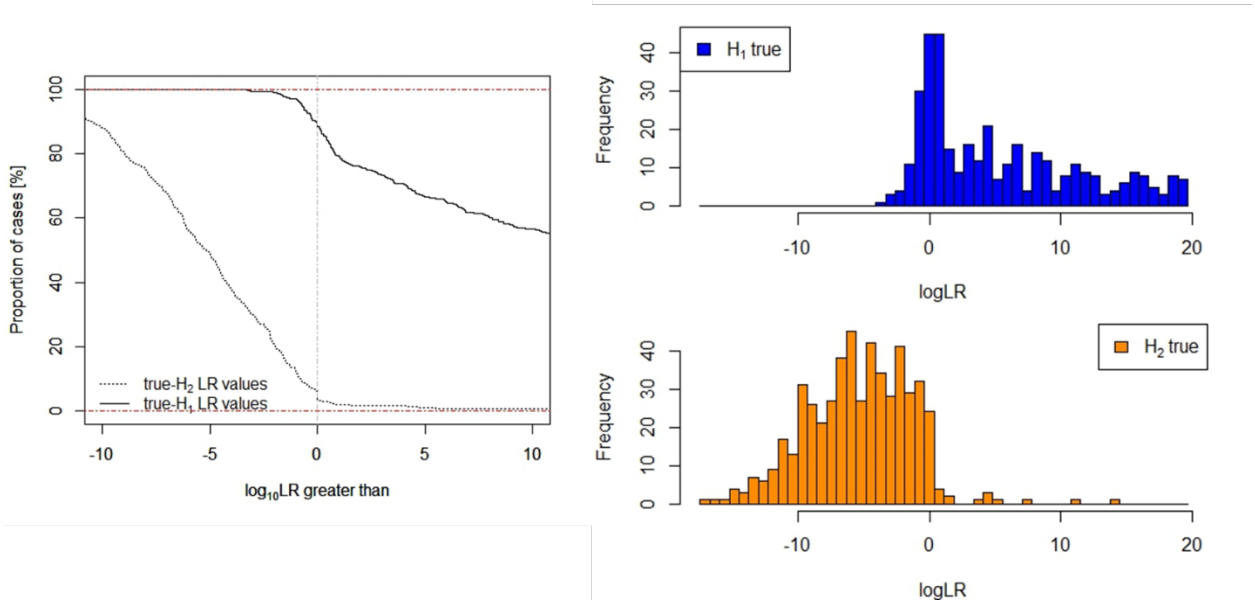


Figure 51: (a) Tippet plot and (b) Histogram obtained for calculated LLRs

Even though this method was poorly calibrated, the tippet plot (Figure 10a) indicates that the calculated LLRs support the hypothesis well. The two hypotheses were H_1: Positive likelihood ratios indicate IL, whereas H_2: Negative likelihood ratios indicate SUB. The calculated misleading evidence was less than 6%. The histogram (Figure 51 b) was a representation of the distribution in LLR in both IL and SUB. In substrates (H_2 is true), the highest occurrence of LLRs was between -4 and -6, which was different than the previous method.

4.2.1.3. Cross-validation of LLRs using Compounds Present in IL

There were 177 compounds were identified as present in ignitable liquids. These were a total combination of compounds that can be seen in both SUB and IL and compounds present only in IL. As described in the previous methods, the performance of calculated LLRs is given in ROC, DET, ECE, tippet and histograms as described above. These are given in Figure 52, 53a, 53b, 54a and 54b respectively.

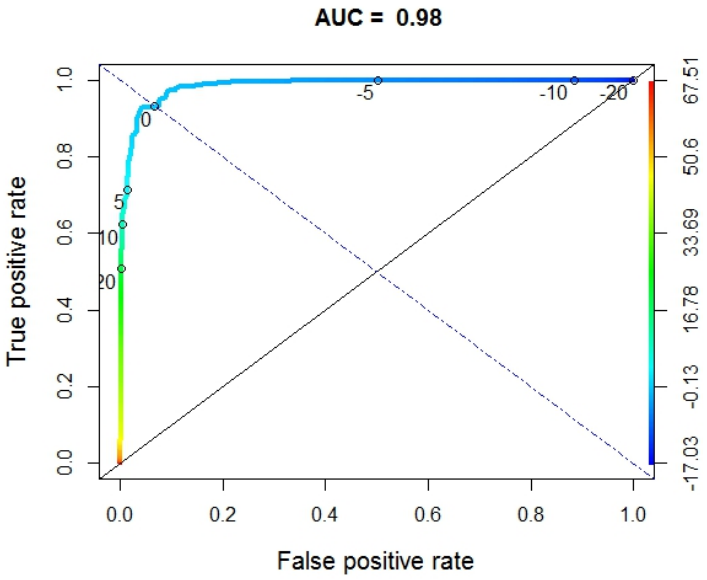


Figure 52: ROC curve obtained for calculated LLRs using compounds in IL

The calculated AUC for this method was 0.98. This indicated that the calculated LLRs using this method have a nearly perfect separation between SUB and IL. The difference between the previous two ROC plots and this one was, that the previous ROC plots indicated a bias towards ignitable liquids, which was not observed in the ROC plot obtained for calculated LLRs using compounds in IL.

(a) (b)

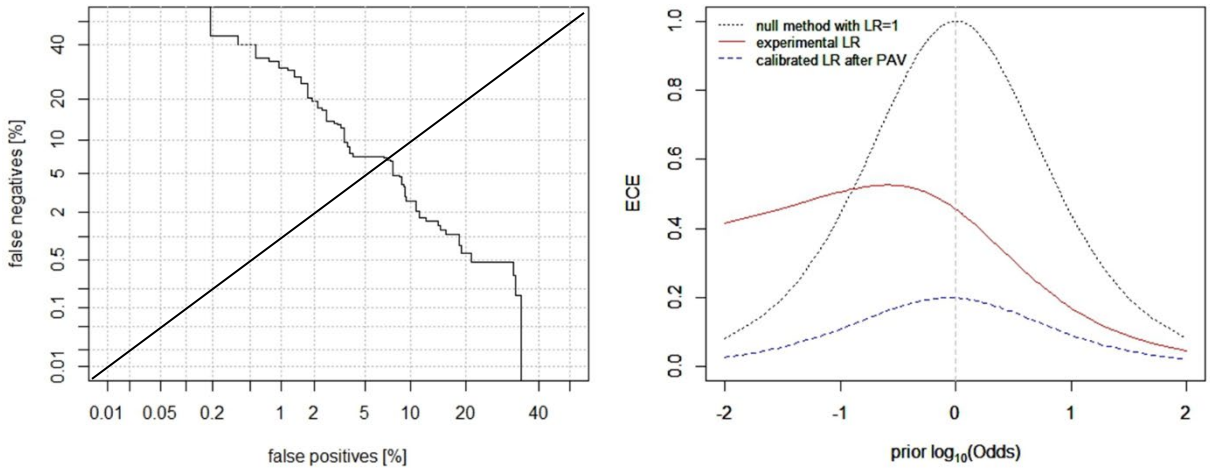


Figure 53: a) DET plot and b) ECE plot obtained for calculated LLRs

The EER obtained for this method was 7.5% (Figure 53 a). This indicated that the calculated LLRs have good discriminating power as same as in the previous method. The ECE plot (Figure 53 b) indicated that this method had poorly calibrated likelihood ratios, however, the tippet plot (Figure 54 a) showed that the misleading evidence obtained in this method was less than 5%, which explained that the calculated LLRs provide satisfactory support to the evidence. The log-likelihood ratio distribution in the histograms is similar to that of obtained in the second method in which the calculated LLRs have a higher distribution from 0 to 20 in IL and -20 to 0 in SUB (Figure 54b).

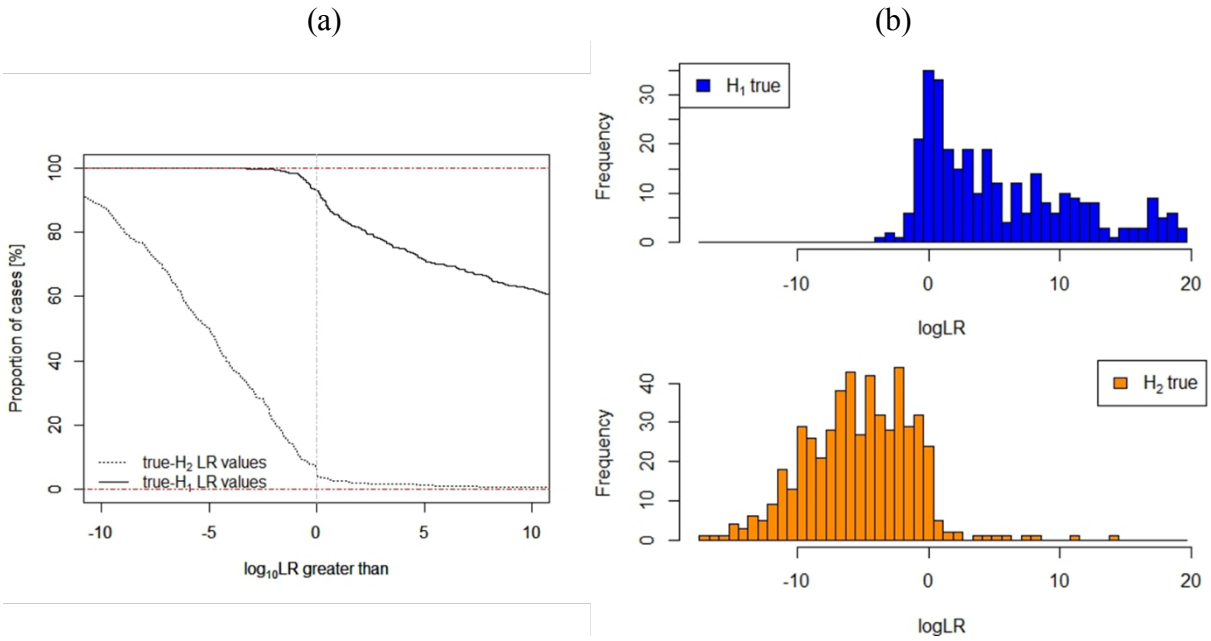


Figure 54: (a) Tippet plot and (b) Histogram obtained for calculated LLRs

4.2.2. Calibration of LLRs using Logistic Regression

In the previously discussed cross-validation methods, the log-likelihood ratios were calculated only for the test data. In the calibration process, the log-likelihood ratios were calculated for both training and test data. A logistic regression model was generated from the calculated LLRs for training data. In this model, the class (IL or SUB) was assigned 1 for calculated LLRs for IL whereas calculated LLRs of SUB were assigned the 0. The calculated LLRs for test data were calibrated in each fold in the 10-fold cross-validation process. This logistic regression model predicted the probability of the calculated LLR being an IL. This probability was calculated by the Equation 4.5. In this equation, β_0 is the intercept and β_1 is the coefficient obtained for LLR.

$$P = \frac{1}{(1+e^{-(\beta_0+\beta_1 LLR)})} \quad \textbf{(4.5)}$$

These calculated probabilities were then converted into posterior log odds. The odds form of Bayes equation for this case is given below (Equation 4.6).

$$\frac{P(H_{IL}|E)}{P(H_{SUB}|E)} = \frac{P(E|H_{IL})}{P(E|(H_{SUB})} \cdot \frac{P(H_{IL})}{P(H_{SUB})} \quad \textbf{(4.6)}$$

In this equation, $\frac{P(H_{IL}|E)}{P(H_{SUB}|E)}$ is the posterior odds, $\frac{P(E|H_{IL})}{P(E|(H_{SUB})}$ is the likelihood ratio and prior odds are presented by $\frac{P(H_{IL})}{P(H_{SUB})}$. Since we do not present a numeric value for the prior odds in the population, the calculated likelihood ratio is equal to the calculated posterior odds×1/prior odds. This fact is being considered in the calibration of LLRs. Therefore, the calibrated LLRs are calculated using Equation 4.7. In this equation, posterior log odds are equal to the LLR if prior odds = 1.

$$log_{10}\left[\frac{P(H_{IL}|E)}{P(H_{SUB}|E)} \times \frac{P(H_{SUB})}{P(H_{IL})}\right] = log_{10}\left[\frac{P(E|H_{IL})}{P(E|(H_{SUB})}\right] \quad \textbf{(4.7)}$$

4.2.2.1. Cross-validation of Calibrated LLRs Using All Compounds Present in SUB and IL

The logistic regression analysis was performed to calibrate the LLRs which were calculated by the same method as discussed above. The calculated AUC in this method was 0.99. ROC, ECE and histogram are presented in Figure 55a, 55b and 55c.

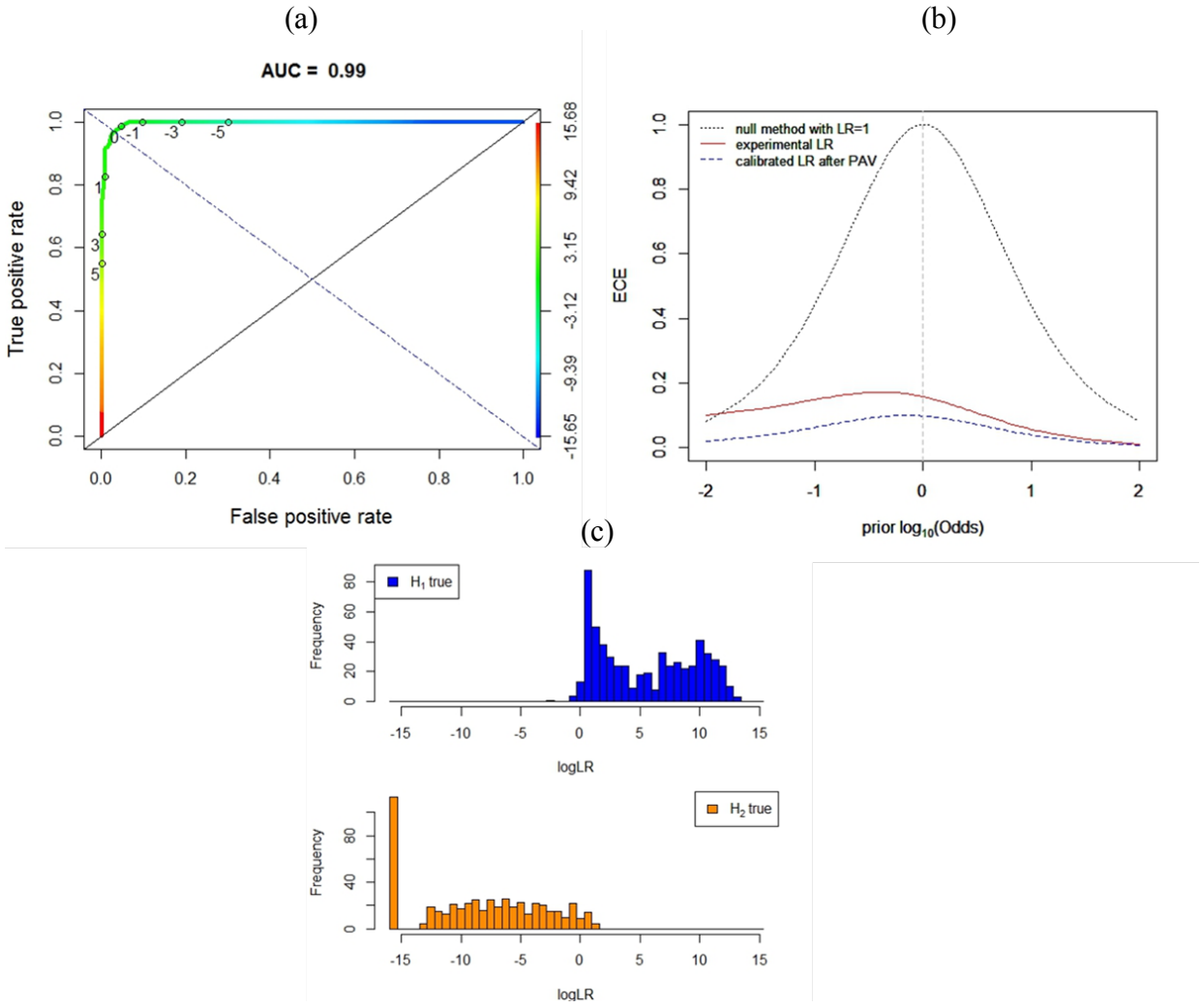


Figure 55: (a) ROC plot (b) ECE plot and (c) Histogram obtained for the calibrated LLRs using all compounds in SUB and IL

The ECE plot obtained after performing logistic regression indicated a better calibration than in Figure 46. According to this plot, the calculated LLRs for IL were better calibrated than the calculated LLRs for SUB. The LLR frequency distribution in the histogram (Figure 55c) shows that the highest frequency of LLRs was in between 1 and 2. Before the calibration, there was no significant distribution of LLRs in SUB but after calibration, the distribution changed.

4.2.2.2. Cross-validation of Calibrated LLRs Calculated using Compounds Common to SUB and IL

The AUC obtained in the ROC curve for the calibrated LLRs using this method is 0.98. The ROC curve, ECE and histograms obtained for these LLRs are given in Figure 56a, 56b and 56c respectively.

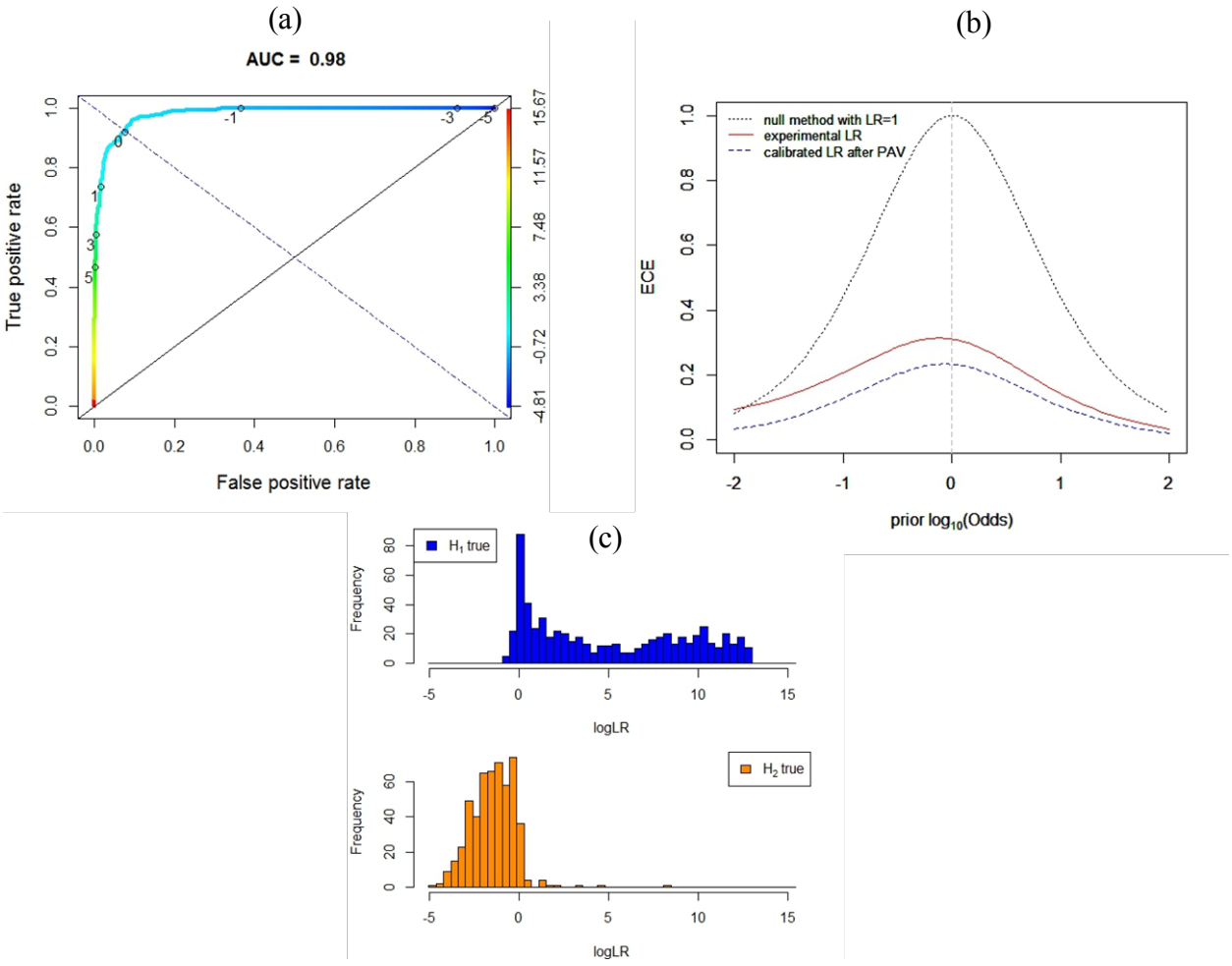


Figure 56: a) ROC plot b) ECE plot and c) Histogram obtained for the calibrated LLRs using common compounds in SUB and IL

As above, the ECE plot indicates a better calibration in the calculated LLRs for ignitable liquids. The LLR distributions in the histograms for IL and SUB were different than the previous method. In this, the majority of calculated LLRs for SUB were distributed from -5 to 0. According to this histogram, positive LLRs can be observed for SUB as well. This was an indication that the frequencies of occurrence of some compounds were higher in SUB than in IL, which can be seen in the previous method as well (all compounds in IL and SUB). Also in IL, the majority of the calculated LLRs are distributed from 0 to 13. Also, the presence of negative LLRs in IL indicates that some compounds have lower frequencies of occurrences in IL than in SUB. This was also observed in the previous methods. The frequency of occurrences of these compounds in SUB and IL will be discussed in Chapter 5.

4.2.2.3. Cross-validation of Calibrated LLRs Calculated using Compounds in IL

In this method, the calibration was performed for the LLRs calculated using the compounds present in IL. The ROC, ECE plots and histogram obtained for these LLRs are included in Figure 57a, 57b and 57c respectively. The AUC of this ROC plot was 0.98. As discussed in the previous methods, these LLRs for IL have a better calibration than the calculated LLRs for SUB.

When the calibration of these three methods was considered, a better calibration of the calculated LLRs in IL observed. Also, the distribution of the LLRs in the histogram (Figure 57c) was similar to Figure 56c. The presence of negative LLRs in IL and positive LLRs in SUB were due to the differences in the frequency of occurrences which can be found in common compounds in both SUB and IL.

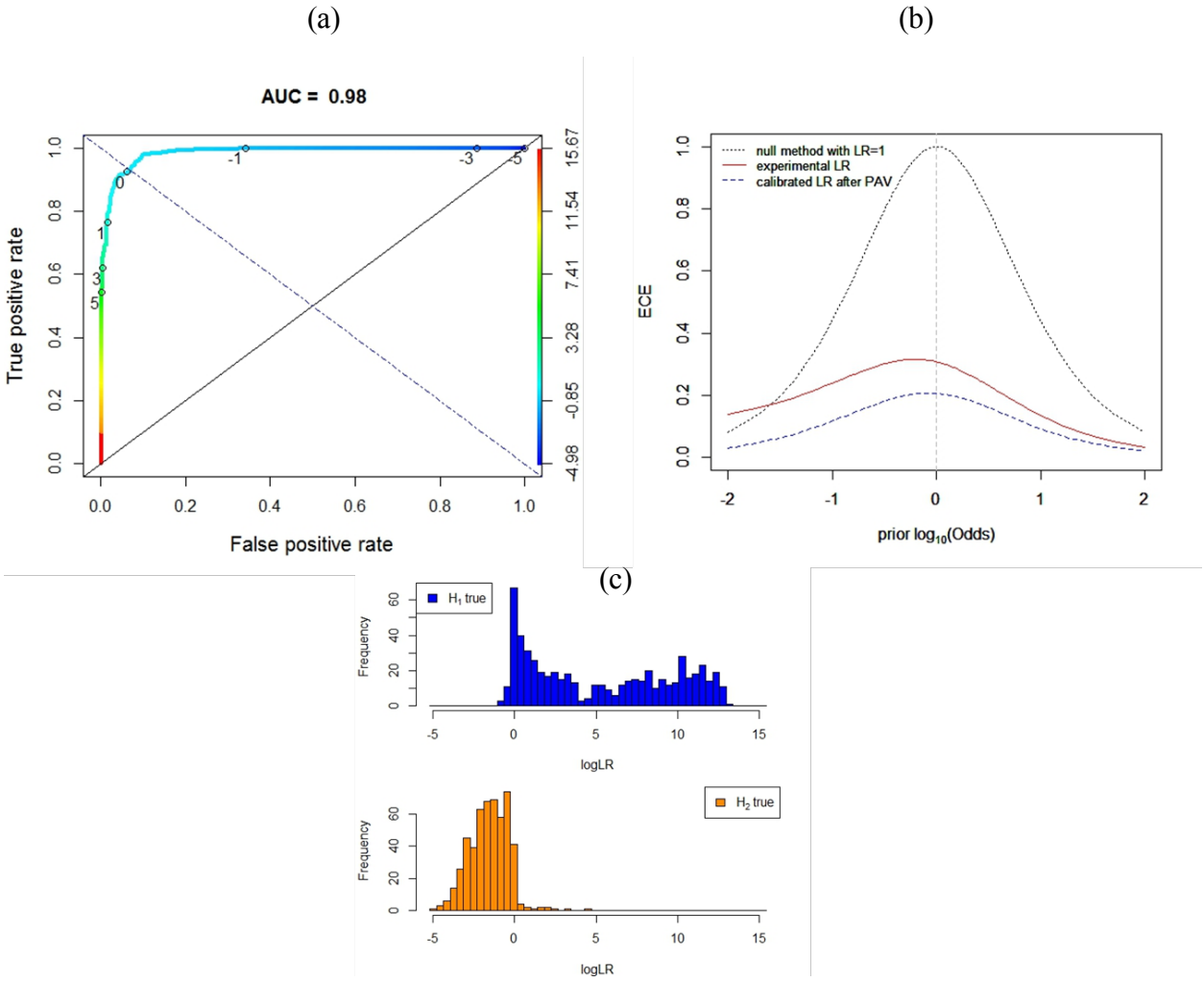


Figure 57: a) ROC plot b) ECE plot and c) Histogram obtained for the calibrated LLRs using compounds present in IL

The initial step of validation of these methods was performed using 16 laboratory generated fire debris samples. Then these methods were applied to a larger data set of fire debris samples. The results of the validation are discussed in Chapter 6.

CHAPTER 5: RESULTS OF THE LOGISTIC REGRESSION ANALYSIS

In this section, the results are included for the frequency of occurrences of the compounds identified in 5 major peaks of the substrates (SUB) and ignitable liquids (IL) in the SUB and IL databases of National Center for Forensic Science. This was published in Journal of Forensic Chemistry, Volume 5, Sept 2017. Results are also included for the logistic regression analysis of the pyrolysis and combustion products of substrates and compounds seen in ignitable liquids. Also, the compounds that can be seen in ASTM E1618-14[5] classes are provided separately. The determination of the probability cutoffs from logistic regression was discussed in Section 3.2.1. According to this cutoff determination, in substrates, the compounds which had a retention time difference (the difference between the peak of the sample chromatogram and the standard library compound) higher than 0.079 min or in ignitable liquids if the retention time difference of the compounds were higher than 0.034 min, were not counted for the calculation of frequency of occurrences of compounds. Even though this method did not identify all the compounds in substrates and ignitable liquids, it eliminated the false positive identification of compounds, which was considered as an advantage in the method.

5.1. The Results of Five Major Compounds Analysis of Ignitable Liquids and Substrates

In this analysis, 36 compounds were identified as present in both ignitable liquids and substrates. There were 102 compounds identified as present only in ignitable liquids and 47 compounds in substrates[17]. The frequencies of occurrence of these compounds are given in Table 1, 2 and 3 respectively.

Table 1: Compounds identified in both IL and SUB

Compound Name	Ignitable Liquids	Substrates
n-undecane	3.20E-01	2.80E-02
n-decane	3.01E-01	1.90E-02
n-nonane	2.49E-01	9.00E-03
n-dodecane	2.36E-01	1.90E-02
1,2,4-trimethylbenzene	1.84E-01	9.00E-03
Toluene	1.81E-01	2.26E-01
n-tridecane	1.59E-01	1.90E-02
m,p-xylene	1.39E-01	3.80E-02
n-tetradecane	1.39E-01	9.00E-03
o-xylene	6.20E-02	1.90E-02
Ethylbenzene	5.70E-02	7.50E-02
Limonene	5.70E-02	5.70E-02
Acetone	5.30E-02	1.90E-02
2-butoxyethanol	4.90E-02	4.70E-02
2,6-dimethylundecane	3.70E-02	9.00E-03
1,2,3-trimethylbenzene	3.40E-02	9.00E-03
2-methylnaphthalene	2.80E-02	2.80E-02
alpha-pinene	2.30E-02	9.40E-02
Isopropanol	2.00E-02	6.60E-02
Naphthalene	1.50E-02	1.51E-01
1-methylnaphthalene	1.10E-02	1.90E-02
beta-pinene	1.10E-02	9.00E-03
Butyl acetate	9.00E-03	9.00E-03
2-heptanone	8.00E-03	9.00E-03
Methyl ester hexadecanoic acid	8.00E-03	1.90E-02
alpha terpineol	6.00E-03	1.90E-02
2,3-dimethylnaphthalene	5.00E-03	9.00E-03
(+)-Longifolene	3.00E-03	3.80E-02
1-Butanol	3.00E-03	9.00E-03
1-methoxy-2-propyl acetate	3.00E-03	4.70E-02
3-Carene	3.00E-03	9.00E-03
1,5-dimethylnaphthalene	2.00E-03	9.00E-03
1-methoxy-2-propanol	2.00E-03	1.90E-02
Camphor	2.00E-03	9.00E-03
Diethylene glycol monoethyl ether	2.00E-03	9.00E-03
2,2,4-trimethyl-1,3-pentanediol diisobutyrate	2.00E-03	5.70E-02

Table 2: Compounds identified only in ignitable liquids

Compound Name	Frequency of Occurrences in IL
n-pentadecane	9.40E-02
m,p-ethyltoluene	7.70E-02
n-hexadecane	7.10E-02
n-octane	6.80E-02
4-methyldecane	6.50E-02
n-heptadecane	5.10E-02
n-heptane	5.10E-02
2,2,6-trimethyloctane	4.50E-02
2,2,8-trimethyldecane	4.50E-02
Methylcyclohexane	3.90E-02
2-methylundecane	3.70E-02
Methanol	3.60E-02
2,2,4-trimethylpentane	3.10E-02
2,3,4-trimethylpentane	3.10E-02
decahydro-2-methylnaphthalene	3.10E-02
n-octadecane	3.10E-02
3-methyl-5-propylnonane	2.90E-02
3-methylhexane	2.80E-02
2-methylhexane	2.60E-02
1,3,5-trimethylbenzene	2.50E-02
1,3-dimethylcyclohexane	2.50E-02
Ethylcyclohexane	2.50E-02
2-butanone	2.30E-02
4-methylnonane	2.30E-02
o-ethyltoluene	2.00E-02
3-methyldecane	1.90E-02
Ethanol	1.90E-02
1,2,4-trimethylcyclohexane	1.70E-02
1,2,3,5-tetramethylbenzene	1.50E-02
2,2,5-trimethylhexane	1.50E-02
2,4-dimethylhexane	1.50E-02
2-methylheptane	1.50E-02
Cyclohexane	1.40E-02
n-hexane	1.40E-02
9-Octadecenoic acid, methyl ester	1.20E-02
1,3,5-trimethylcyclohexane	1.10E-02
2,3,3-trimethylpentane	1.10E-02

Compound Name	Frequency of Occurrences in IL
2-methylnonane	1.10E-02
2-methyloctane	1.10E-02
4-ethyl-1,2-dimethylbenzene	1.10E-02
Methyl Isobutyl Ketone	1.10E-02
Methylene chloride	1.10E-02
Pristane	9.00E-03
(trans) Decahydronaphthalene	9.00E-03
1,2,4,5-tetramethylbenzene	9.00E-03
9, 12-Octadecadienoic acid (Z,Z)-methyl ester	9.00E-03
dipropylene glycol methyl ether isomer	9.00E-03
Phytane	9.00E-03
Propylcyclohexane	8.00E-03
1-Butoxy-2-propanol	8.00E-03
2,3-dimethylpentane	8.00E-03
2,6-dimethyloctane	8.00E-03
3-methylpentane	8.00E-03
dimethyl glutarate	8.00E-03
Nitromethane	8.00E-03
Propylbenzene	6.00E-03
1,1,3-trimethylcyclohexane	6.00E-03
2,3,5-trimethylhexane	6.00E-03
2-methylpentane	6.00E-03
3-methylnonane	6.00E-03
alpha-terpinolene	6.00E-03
Cyclohexanone	6.00E-03
Malathion	6.00E-03
methyl ester octadecanoic acid	6.00E-03
Methylcyclopentane	6.00E-03
n-nonadecane	5.00E-03
2,4-dimethylpentane	5.00E-03
2-butoxyethyl acetate	5.00E-03
2-methylbutane	5.00E-03
1-methoxy-4-(2-propenyl) benzene	5.00E-03
Dimethyl ester hexanedioic acid	5.00E-03
Ethyl 3-ethoxy-propionate	5.00E-03
Isobutyl isobutyrate	5.00E-03
n-Pentane	5.00E-03
o-chlorotoluene	3.00E-03

Compound Name	Frequency of Occurrences in IL
2,2-dimethylbutane	3.00E-03
2,2'oxybisethanol	3.00E-03
2,5-dimethylhexane	3.00E-03
alpha terpinene	3.00E-03
Diethyl Phthalate	3.00E-03
Ethyl Acetate	3.00E-03
Heptylacetate	3.00E-03
Hexylacetate	3.00E-03
Indane	3.00E-03
Isopropylbenzene	2.00E-03
(Trans)1-ethyl-4-methylcyclohexane	2.00E-03
(cis) Decahydronaphthalene	2.00E-03
1,3,5-Tris (1-methylethyl) benzene	2.00E-03
1,3-bis(1-methylethyl)benzene	2.00E-03
1-decanol	2.00E-03
1-methyl-2-Pyrrolidinone	2.00E-03
2-methyl-1-propanol	2.00E-03
3-methylheptane	2.00E-03
3-tert-butylphenol	2.00E-03
cis 1-ethyl-4-methylcyclohexane	2.00E-03
Citronellal	2.00E-03
Diazinon	2.00E-03
Diethyl ether	2.00E-03
Ethylcyclopentane	2.00E-03
Isobornyl Acetate	2.00E-03
Octylacetate	2.00E-03
Oxolane (Tetrahydrofuran)	2.00E-03

Table 3: Compounds identified only in substrates

Compound Name	Frequency of Occurrences SUB
2-Furaldehyde	3.77E-01
Styrene	2.36E-01
Benzaldehyde	2.26E-01
2-Methoxyphenol	2.17E-01
Phenol	1.89E-01
5-Methylfurfural	1.42E-01
Benzene	1.23E-01
2,6-Dimethoxyphenol	1.04E-01
4-Ethyl-2-methoxyphenol	9.40E-02
Furfuryl Alcohol	9.40E-02
Hexanal	7.50E-02
1-Pentanol	6.60E-02
2,4-dimethyl-1-heptene	5.70E-02
Acetophenone	5.70E-02
Biphenyl	5.70E-02
2-ethyl-1-hexanol	4.70E-02
Benzyl Chloride	4.70E-02
2-Pentylfuran	3.80E-02
Caprolactam	3.80E-02
1-dodecanol	2.80E-02
benzyl alcohol	2.80E-02
Phthalic acid anhydride	2.80E-02
1,3-Dichloro 2 propanol	1.90E-02
1-Tridecene	1.90E-02
1-Undecene	1.90E-02
2-methoxy-4-methylphenol	1.90E-02
alpha methylstyrene	1.90E-02
Benzophenone	1.90E-02
Bibenzyl	1.90E-02
Hexamethyl cyclotrisiloxane	1.90E-02
1,2-dichloroethane	9.00E-03
1-Decene	9.00E-03
1-Hexadecene	9.00E-03
1-Octadecene	9.00E-03
2-(2-chloroethoxy) ethanol	9.00E-03
2-(2-n-butoxyethoxy) ethyl acetate	9.00E-03
2,6-di-tert-butyl-4-methyl phenol	9.00E-03

Compound Name	Frequency of Occurrences SUB
4-phenylbutronitrile	9.00E-03
Benzoic Acid	9.00E-03
Benzonitrile	9.00E-03
bis(2-chloroethyl) ether	9.00E-03
Cedrol	9.00E-03
Cyclopentanone	9.00E-03
Dibutyl Phthalate	9.00E-03
Dimethylformamide	9.00E-03
Nonanal	9.00E-03
p-Cresol	9.00E-03

5.2. Compounds Identified Only in SUB, IL and Compounds Identified in Both SUB and IL

By logistic regression analysis, 75 compounds were identified as present only in substrates and 38 compounds were identified as present only in ignitable liquids. There were 139 compounds identified as present in both ignitable liquids and substrates. The list of compounds in these categories is given in Table 4, 5 and 6 respectively.

Table 4: The compounds only identified in Substrates

Compound	SUB Frequency of occurrences
Styrene	6.70E-01
indene	3.98E-01
furfural	3.56E-01
2-cyclopenten-1-one	3.03E-01
2-methylfuran	3.01E-01
2-methylphenol	2.95E-01
Phenol	2.85E-01
Furfuryl alcohol	2.57E-01
5-Methylfurfural	2.49E-01
creosol	1.95E-01
1-Pentadecene	1.72E-01
Cyclopentanone	1.70E-01

Compound	SUB Frequency of occurrences
Acetaldehyde	1.67E-01
2,4,6,8-tetramethyl-1-undecene (syndiotactic)(C)	1.65E-01
1,3-diphenylpropane	1.65E-01
2,4-dimethyl-1-heptene	1.57E-01
2,4,6-trimethyl-1-nonene (racemic form)(E)	1.55E-01
2,4,6,8-tetramethyl-1-undecene (isotactic)(A)	1.55E-01
2,4,6,8-tetramethyl-1-undecene (heterotactic)(B)	1.51E-01
vinyl benzoate	1.40E-01
eugenol	1.25E-01
2-ethyl-1-hexene	1.19E-01
Acetophenone	1.09E-01
1-nonene	1.05E-01
Nonanal	1.05E-01
(trans)-3-methyl-2-heptene	9.96E-02
(cis)-3-methyl-2-heptene	9.58E-02
4-vinyl-1-cyclohexene	9.58E-02
3-chloromethylheptane	9.58E-02
2,6-dimethoxyphenol	8.05E-02
methyl methacrylate	7.09E-02
TXIB	7.09E-02
Heptanal	6.32E-02
4-Ethyl-2-methoxyphenol	5.94E-02
1-methylpyrrole	5.94E-02
2-ethylhexyl benzoate	5.75E-02
Acetonitrile	4.41E-02
Acrolein	4.41E-02
Cyclohexanone	4.41E-02
benzothiazole	3.83E-02
Benzophenone	3.83E-02
2-pentylfuran	3.45E-02
Caprolactam	3.07E-02
1,3-dichloro-2-propanol	2.87E-02
bis(2-chloroethyl)ether	2.87E-02
1,2-dichloroethane	2.49E-02
1,2-diphenylpropane	2.11E-02
dimethylformamide	1.92E-02
allylbenzoate	1.92E-02

Compound	SUB Frequency of occurrences
p-tert-butylphenol	1.72E-02
Tridecanal	1.72E-02
Bibenzyl	1.72E-02
2-(2-chloroethoxy)ethanol	1.53E-02
1-butanol	1.34E-02
3-tert-butylphenol	1.34E-02
4-phenylbutronitrile	1.34E-02
1-hexadecene	1.34E-02
3-butene-1,3-diyldibenzene	1.34E-02
Benzoic acid	1.34E-02
2-phenoxyethanol	9.58E-03
Safrole	9.58E-03
Triacetin	9.58E-03
Tetradecanal	9.58E-03
2-methylpentanal	7.66E-03
1-chlorooctane	7.66E-03
Decanal	7.66E-03
1-phenoxypropan-2-ol	7.66E-03
Ethanol-2-(2-butoxyethoxy)-acetate	7.66E-03
Dibutyl phthalate	7.66E-03
Propanal	5.75E-03
2,4,4-trimethylpentane	5.75E-03
methyl salicylate	3.83E-03
methylbenzoylformate	3.83E-03
1-methoxy-4-(2-propenyl)-benzene	1.92E-03
Diethyltoluamide	1.92E-03

Table 5: The compounds identified only in ignitable liquids

Compound	IL Frequency of occurrences
2,4-dimethylhexane	1.65E-01
2,5-dimethylhexane	1.59E-01
2,3-dimethylpentane	8.57E-02
2,2,5-trimethylhexane	8.57E-02
2,3,3-trimethylpentane	8.10E-02
3-methylpentane	6.23E-02
2,2,4-trimethylpentane	2.49E-02
Alpha terpinene	1.56E-02
Tri(propylene glycol)methyl ether mixture of isomers-C	1.56E-02
n-heneicosane	1.40E-02
2,2-dimethylbutane	1.25E-02
Tri(propylene glycol) methyl ether mixture of isomers-A	1.25E-02
2-methyl pentane	1.09E-02
1-butoxy-2-propanol	9.35E-03
Di(propylene glycol)methyl ether acetate mixture of isomers-A	9.35E-03
Tri(propylene glycol)methyl ether mixture of isomers-D	9.35E-03
Tri(propylene glycol) methyl ether mixture of isomers-E	9.35E-03
Nitromethane	7.79E-03
Ethyl vanillin	7.79E-03
3-(4-tert-Butylphenyl)-2-methylpropanal	7.79E-03
9,12-Octadecadienoic acid(Z,Z)-methyl ester	7.79E-03
Methylene chloride	6.23E-03
2-methyl-1-propanol	6.23E-03
Methoxy-3-methylbutanol	6.23E-03
Methyl ester octadecanoic acid	6.23E-03
Isobutyl isobutyrate	4.67E-03
alpha methyl (trans)-cinnamaldehyde	4.67E-03
2-butoxyethyl acetate	3.12E-03
1,3-bis(1-methylethyl)benzene	3.12E-03
Diethyl Phthalate	3.12E-03
bis(2-ethylhexyl)adipate	3.12E-03
Diethyl ether	1.56E-03
2,2,8-trimethyl-decane	1.56E-03
Decamethylcyclopentasiloxane	1.56E-03
Tri(propylene glycol)methyl ether mixture of isomers-B	1.56E-03
Benzyl benzoate	1.56E-03
Malathion	1.56E-03
9-Octadecenoicacid-(Z)-methyl ester	1.56E-03

Table 6: Compounds seen in both SUB and IL

Compound	SUB	IL
Toluene	7.22E-01	3.61E-01
Benzaldehyde	6.44E-01	6.23E-03
Benzene	6.03E-01	1.07E-01
Naphthalene	5.04E-01	2.45E-01
m-xylene	4.18E-01	3.68E-01
1-methylnaphthalene	3.26E-01	1.42E-01
Acetone	3.01E-01	2.96E-02
o-xylene	2.97E-01	4.03E-01
Alpha-methylstyrene	2.91E-01	1.56E-03
2-methoxy phenol	2.82E-01	1.56E-03
m-ethyltoluene	2.76E-01	4.08E-01
p-ethyltoluene	2.74E-01	3.93E-01
Ethylbenzene	2.66E-01	1.99E-01
2-methylnaphthalene	2.45E-01	2.21E-01
2-methylbutane	2.03E-01	2.80E-02
1,3,5-trimethylbenzene	1.93E-01	4.08E-01
n-propylbenzene	1.88E-01	3.43E-01
1-decene	1.78E-01	5.14E-02
2,3-dimethylnaphthalene	1.76E-01	1.64E-01
Fluorene	1.69E-01	1.87E-02
Benzonitrile	1.63E-01	1.56E-03
2,4,6-trimethyl-1-nonene (meso form)(D)	1.49E-01	1.56E-03
n-heptane	1.44E-01	1.53E-01
p-alpha-dimethylstyrene	1.28E-01	4.05E-02
n-undecane	1.28E-01	3.89E-01
Methyl Isobutyl Ketone	1.21E-01	3.12E-02
p-Cresol	1.19E-01	1.56E-03
1-Tridecene	1.13E-01	4.67E-03
n-tetradecane	1.13E-01	1.70E-01
n-tridecane	1.13E-01	1.82E-01
Hexanal	1.11E-01	1.56E-03
1-Tetradecene	1.05E-01	1.56E-03
n-pentadecane	1.05E-01	1.32E-01
n-octane	1.05E-01	3.96E-01
Acenaphthene	1.03E-01	3.12E-03
Indane	1.03E-01	2.26E-01
n-dodecane	1.03E-01	3.88E-01
Limonene	9.20E-02	6.85E-02

Compound	SUB	IL
Biphenyl	8.62E-02	9.35E-03
n-hexane	8.62E-02	6.39E-02
n-decane	8.62E-02	4.72E-01
p-xylene	8.24E-02	1.40E-02
n-nonane	8.24E-02	3.13E-01
Tetrahydrofuran	8.05E-02	3.12E-03
1,3,5-trimethyl cyclohexane	8.05E-02	2.38E-01
1,2,4,5-tetramethylbenzene	7.85E-02	3.15E-01
1,2,3-trimethylbenzene	7.66E-02	3.29E-01
1,5-dimethylnaphthalene	7.47E-02	1.39E-01
1,2,3,5-tetramethylbenzene	7.09E-02	3.82E-01
4-ethyl-1,2-dimethylbenzene	6.90E-02	3.24E-01
Pentanal	6.70E-02	4.67E-03
1,2,4-trimethyl-(1.alpha, 2.beta, 4.beta)-Cyclohexane	6.70E-02	3.21E-01
Methanol	6.51E-02	3.27E-02
2-butanone	6.32E-02	3.27E-02
1-Dodecanol	5.94E-02	1.56E-03
n-hexadecane	5.56E-02	6.70E-02
Octanal	5.17E-02	1.09E-02
benzyl_alcohol	4.98E-02	6.23E-03
Benzyl chloride	4.79E-02	1.56E-03
1-Pentanol	4.79E-02	3.12E-03
4-Nonene	4.41E-02	1.56E-02
alpha terpineol	4.41E-02	2.65E-02
Phthalic acid anhydride	4.21E-02	1.71E-02
3-methylheptane	4.21E-02	2.68E-01
1-dodecene	4.02E-02	1.56E-03
Isopropanol	3.83E-02	1.40E-02
Isopropyl benzene	3.83E-02	1.40E-01
o-ethyltoluene	3.83E-02	1.48E-01
alpha-pinene	3.45E-02	4.52E-02
2-ethyl-1-hexanol	3.26E-02	3.12E-03
Anthracene	2.68E-02	1.56E-03
beta-pinene	2.11E-02	3.12E-03
2-heptanone	2.11E-02	4.67E-03
n-heptadecane	2.11E-02	3.12E-02
2-butoxy ethanol	2.11E-02	6.07E-02
Pristane	2.11E-02	9.03E-02
Methylcyclohexane	2.11E-02	2.87E-01

Compound	SUB	IL
1-undecene	1.92E-02	1.56E-03
Butyl acetate	1.92E-02	4.67E-03
n-Pentane	1.92E-02	6.23E-03
1-Methoxy-2-propyl_acetate	1.92E-02	2.34E-02
Ethanol	1.92E-02	2.49E-02
1_2_4-trimethylbenzene	1.92E-02	3.10E-01
Diethylene glycol monoethyl ether	1.53E-02	4.67E-03
1-methoxy-2-propanol	1.15E-02	4.67E-03
2,6-dimethylundecane	1.15E-02	3.36E-01
Ethylcyclohexane	1.15E-02	3.77E-01
3-methyldecane	1.15E-02	4.74E-01
Di(propylene glycol)methyl ether mixture of isomers-B	9.58E-03	1.25E-02
Di(propylene glycol)methyl ether mixture of isomers-A	9.58E-03	1.40E-02
Ethylcyclopentane	9.58E-03	4.98E-02
4-methylnonane	9.58E-03	4.66E-01
4-methyldecane	9.58E-03	4.72E-01
2,6-dimethyloctane	9.58E-03	5.20E-01
Longifolene	7.66E-03	3.12E-03
Cyclohexane	7.66E-03	5.30E-02
1-decanol	5.75E-03	1.56E-03
Di(propylene_glycol)methyl ether mixture of isomers-C	5.75E-03	3.12E-03
DL-Camphor	5.75E-03	3.12E-03
3-carene	5.75E-03	6.23E-03
Phytane	5.75E-03	2.65E-02
2-methyl hexane	5.75E-03	6.39E-02
3-methyl-5-propylnonane	5.75E-03	7.32E-02
1,1,3-trimethylcyclohexane	5.75E-03	3.80E-01
Dodecanal	3.83E-03	1.56E-03
Benzyl acetate	3.83E-03	1.56E-02
Ethyl acetate	3.83E-03	2.65E-02
Methylcyclopentane	3.83E-03	4.98E-02
2,4-dimethyl pentane	3.83E-03	8.26E-02
2-methylheptane	3.83E-03	2.87E-01
decahydro-2-methylnaphthalene	3.83E-03	3.44E-01
Propylcyclohexane	3.83E-03	4.56E-01
2-methylnonane	3.83E-03	4.69E-01
2,2-oxybis-ethanol	1.92E-03	1.56E-03
Cedrol	1.92E-03	1.56E-03
n-octadecane	1.92E-03	1.56E-03

Compound	SUB	IL
p-anisaldehyde	1.92E-03	3.12E-03
Caryophyllene	1.92E-03	3.12E-03
o-chlorotoluene	1.92E-03	4.67E-03
Isobornyl acetate	1.92E-03	4.67E-03
2,4,4-trimethyl-2-pentene	1.92E-03	7.79E-03
Phenylethyl alcohol	1.92E-03	7.79E-03
Dimethyl glutarate	1.92E-03	9.35E-03
Di(propylene glycol)methyl ether acetate mixture of isomers-B	1.92E-03	9.35E-03
Methyl ester hexadecanoic acid	1.92E-03	1.09E-02
n-Hexyl acetate	1.92E-03	1.25E-02
alpha-terpinolene	1.92E-03	1.87E-02
n-eicosane	1.92E-03	2.18E-02
(cis)-Decahydronaphthalene	1.92E-03	2.34E-02
2,2,6-trimethyl octane	1.92E-03	6.07E-02
3-methylhexane	1.92E-03	7.32E-02
2,3,4-trimethyl-pentane	1.92E-03	1.36E-01
2,3,5-trimethyl-hexane	1.92E-03	1.51E-01
(cis)-Cyclohexane-1,3-dimethyl	1.92E-03	2.90E-01
(trans)-Decahydronaphthalene	1.92E-03	3.33E-01
(cis)-1-ethyl-4-methylcyclohexane	1.92E-03	3.82E-01
2-methyl octane	1.92E-03	3.97E-01
2-methyl undecane	1.92E-03	3.97E-01
(trans)-1-ethyl-4-methylcyclohexane	1.92E-03	4.19E-01

In Table 6, the frequencies of occurrence are organized according to the decreasing frequency in substrates. Relative to the five major peaks analysis, logistic regression analysis resulted in an increase in compounds identified in both SUB and IL and compounds identified only in SUB. The main reasons for this were the identification of both major and minor peaks, and the analysis of a higher number of different substrate types

According to the ASTM E1618-14 classification Tables 3,4, and 5, there are 15, 13 and 26 target compounds in gasoline, medium petroleum distillate (MPD) and heavy petroleum distillate (HPD) respectively[5]. However, from the total of 38 target compounds in these classes, 25 compounds were also identified as pyrolysis or combustion products in substrates. These compounds were identified as major or minor peaks in substrates. These substrate classes in which these compounds can be seen are tabulated in Table 7. Given below. In this table, PTT, PVC and HDPE are polytrimethylene terephthalate, polyvinyl chloride and high-density polyethylene respectively. The frequencies of occurrence of compounds in each ignitable liquid ASTM E1618-14[5] class are tabulated in Table 8.

Table 7: Substrate types that can be seen in target compounds of GAS, MPD and HPD

Compound	Substrate Categories			
1,3,5-trimethylbenzene	carpet magazines roofing engineered garage exercise adhesives dashboard olefin cork vinyl	Hardwood Rope worn footwear Auto car mat window treatments particle board Polyester Polyethylene polystyrene-co-butadiene	recycled material manila asphalt fiberglass isoprene canvas olefin nylon pvc olefin polyester nylon polypropylene	plastic polyurathane wood paper rubber PTT wool olefin vinyl Automobile carpet
1,2,4-trimethylbenzene	roofing isoprene canvas	worn footwear Rubber	garage exercise flooring pvc	polyethylene
1,2,3-trimethylbenzene	hardwood recycled material manila paper	Cork Wood Vinyl Pvc	roofing polyethylene plastic	worn footwear asphalt fiberglass rubber
indane	bedding vinyl linoleum carpet composite decking olefin vinyl	Hardwood Magazines Roofing rug gripper Plastic Manila	cork worn footwear paper polyester recycled material garage exercise flooring	cotton polyurathane asphalt fiberglass wood polyethylene window treatments
1,2,4,5-tetramethylbenzene	magazines worn footwear rope rubber	Roofing Wood Automobile car mat Pvc	adhesives polyethylene asphalt fiberglass polystyrene-co-butadiene	footwear vinyl isoprene canvas fauxleather
1,2,3,5-tetramethylbenzene	roofing wood rope	worn footwear chair couch Automobile car mat	adhesives hardwood isoprene canvas	polyethylene asphalt fiberglass rubber

Compound	Substrate Categories			
1,2,3,5-tetramethylbenzene	garage exercise polystyrene-co-butadiene	window treatments fauxleather	vinyl	pvc
n-dodecane	cork worn footwear composite decking vinyl linoleum pvc cork vinyl	insulation roofing carpet padding plastic polyethylene Automobile carpet	magazines paper chair couch garage exercise flooring leather HDPE	Automobile car mat asphalt fiberglass vinyl fauxleather polyurathane
2-methylnaphthalene	worn clothing wood cardboard insulation worn footwear laminate polyester plastic jute	bedding engineered hardwood vinyl linoleum rope paper cotton recycled material olefin nylon	carpet upholstery cork magazines roofing composite decking olefin polyethylene	pvc vinyl polyurathane isoprene canvas manila leather rubber asphalt fiberglass
1-methynaphthalene	worn clothing carpet engineered cardboard cork magazines roofing laminate rug gripper polyester	bedding wood upholstery hardwood vinyl linoleum rope worn footwear paper window treatments cotton	olefin polyethylene plastic jute leather isoprene canvas olefin nylon PTT wool olefin cork vinyl	recycled material vinyl manila polyurathane asphalt fiberglass rubber pvc olefin polyester nylon polystyrene-co-butadiene
2,3-dimethyl-naphthalene	bedding wood	engineered hardwood	window treatments footwear	adhesives polyester

Compound	Substrate Categories			
2,3-dimethyl-naphthalene	cork carpet worn footwear rope isoprene canvas polystyrene-co-butadiene	vinyl linoleum roofing paper garage exercise rubber fauxleather	recycled material vinyl manila asphalt fiberglass pvc	polyethylene plastic leather cotton cork vinyl
n-nonane	cork worn footwear chair couch plastic leather	insulation roofing packaging materials polyethylene asphalt fiberglass	magazines composite decking vinyl linoleum recycled material polyurathane	vinyl cotton isoprene canvas rubber
propylcyclohexane	roofing	asphalt fiberglass		
n-decane	cork worn footwear chair couch garage exercise	insulation roofing packaging materials polyethylene	magazines composite decking Plastic Leather	asphalt fiberglass fauxleather pvc HDPE
trans-decahydronaphthalene	roofing			
n-undecane	cork worn footwear composite decking garage exercise adhesives Automobile carpet	insulation roofing packaging materials Auto carmat polyethylene	magazines Paper Plastic engineered Leather	asphalt fiberglass HDPE polyurathane fauxleather pvc
n-tridecane	chair couch vinyl linoleum garage exercise window treatments composite decking	insulation Auto upholstry Auto carmat laminate polyethylene	packaging materials Plastic engineered adhesives asphalt fiberglass	Automobile carpet polyurathane HDPE fauxleather
n-tetradecane	cork	insulation	magazines	polyurathane

Compound	Substrate Categories			
n-tetradecane	worn footwear chair couch rope Auto carmat leather	roofing packaging materials plastic adhesives asphalt fiberglass	composite decking vinyl linoleum garage exercise polyethylene Wood	HDPE vinyl fauxleather cork vinyl Automobile carpet
n-pentadecane	insulation worn footwear composite decking vinyl linoleum plastic Auto carmat adhesives	magazines roofing packaging materials rope garage exercise flooring window treatments polyethylene	Leather Cotton polyurathane cork vinyl HDPE Automobile carpet	asphalt fiberglass isoprene canvas vinyl polystyrene-co-butadiene fauxleather
n-hexadecane	cork plastic polyethylene polyurathane	insulation garage exercise asphalt fiberglass cork vinyl	composite decking window treatments Cotton	fauxleather vinyl isoprene canvas
n-heptadecane	window treatments isoprene canvas	plastic wood	polyethylene	asphalt fiberglass
n-heptadecane	window treatments wood	plastic isoprene canvas	polyethylene Plastic	asphalt fiberglass wood
pristane	isoprene canvas	rope	Manila rope	
n-octadecane	polyethylene			
phytane	isoprene canvas			
n-eicosane	wood			

5.3. Compounds Identified in ASTM E1618-14 Ignitable Liquid Classes

The compounds only present in GAS, AR, ISO, OXY and MPD are listed below. According to this analysis, there are 30 compounds seen only in oxygenated, 11 compounds in miscellaneous, 2 compounds in aromatic and MPD and finally one compound in gasoline and isoparaffinic. These compounds are listed below. The compounds that seen only in ignitable liquids are underlined.

Compounds in oxygenated (OXY) IL class: (check nomenclature)

1-decanol, 2,2-oxybis-ethanol, 2-butoxyethyl acetate, 2-heptanone, phenylethyl alcohol, 2-methoxyphenol, 2-methyl-1-propanol, 9,12-octadecadienoic acid-(Z,Z)-methyl ester, 9-octadecenoicacid-(Z)-methyl ester, alpha methylstyrene, benzylchloride, benzyl benzoate alpha methyl (trans)- cinnamaldehyde, benzaldehyde, bis-(2-ethylhexyl)adipate, butyl acetate, caryophyllene, cedrol, di(propylene glycol) methyl ether mixture of isomers – A, di(propylene glycol) methyl ether mixture of isomers – B, di(propylene glycol) methyl ether mixture of isomers-C, diethyl ether, dimethyl glutarate, 3-(4-(tert-Butyl)phenyl)-2-methylpropanal (lilial) methoxy-3-methylbutanol, methyl ester octadecanoic acid, methylene chloride, nitromethane, p-anisaldehyde, and tetrahydrofuran

Compounds in miscellaneous (MISC) IL class:

o-chlorotoluene, 1,3-bis(1-methylethyl)benzene, hexanal, p-cresol, 1-undecene, 1-dodecene, tri(propylene glycol) methyl ether mixture of isomers-B, 1-tetradecene, dodecanal, 1-dodecanol and anthracene.

Compounds in medium petroleum distillate (MPD):

2,4,6-trimethyl-1-nonene (meso form) – D, and decamethylcyclopentasiloxane

Compounds in aromatic (AR): benzonitrile and malathion.

Gasoline contained 4-nonene and isoparaffinic liquids contained 2,2,8-trimethyl decane.

The frequencies of occurrence in each ignitable liquid according to ASTM E1618-14[5] classes are tabulated in Table 8. The frequencies are ordered from the largest to smallest according to the frequencies of gasoline. Calculation of the log-likelihood ratios of fire debris samples and validation of the calculation methods are discussed in Chapter 6.

Table 8 Frequency of occurrences of compounds in ASTM E1618-14[5] IL classes

COMPOUND	GAS	NORMA	AR	ISO	OXY
m-ethyltoluene	0.947	-	0.781	-	0.187
naphthalene	0.921	-	0.500	-	0.122
1,2,4,5-tetramethylbenzene	0.921	-	0.500	-	0.130
4-ethyl-1,2-dimethylbenzene	0.868	-	0.344	-	0.138
1,2,3,5-tetramethylbenzene	0.868	-	0.594	-	0.138
m-xylene	0.842	-	0.438	-	0.203
1,2,4-trimethylbenzene	0.816	-	0.531	-	0.114
1,2,3-trimethylbenzene	0.816	-	0.625	-	0.138
1,3,5-trimethylbenzene	0.816	-	0.594	-	0.163
p-ethyltoluene	0.816	-	0.719	-	0.179
methylcyclohexane	0.789	-	0.031	-	0.171
o-xylene	0.789	-	0.688	0.024	0.179
n-octane	0.789	-	-	-	0.171
2,5-dimethyl,hexane	0.763	-	-	0.146	0.065
toluene	0.763	0.003	0.344	0.024	0.285
Benzene	0.737	-	0.063	0.024	-
2-methylnaphthalene	0.737	-	0.219	-	0.089
2-methylheptane	0.737	-	-	0.098	0.138
2-methyloctane	0.711	-	0.031	0.098	0.098
1-methylnaphthalene	0.684	-	0.250	-	0.033
2,3,5-trimethyl-hexane	0.684	-	-	0.171	0.033
n-propylbenzene	0.684	-	0.688	-	0.146
2,6-dimethyloctane	0.684	-	0.219	0.561	0.138
2,4-dimethylhexane	0.658	-	-	0.171	0.065
4-methylnonane	0.605	-	0.281	0.024	0.130
2,3,4-trimethylpentane	0.553	-	-	0.195	0.057

COMPOUND	GAS	NORMA	AR	ISO	OXY
2,3-dimethyl-naphthalene	0.526	-	0.094	-	0.073
2-methyl,nonane	0.500	-	0.313	0.049	0.146
n-decane	0.474	0.013	0.250	0.024	0.163
n-hexane	0.447	-	-	-	0.033
2,4-dimethylpentane	0.447	-	-	0.073	0.033
(cis)-1,3-dimethylcyclohexane	0.447	-	-	0.024	0.122
2,2,5-trimethylhexane	0.421	-	-	0.220	0.024
3-methyldecane	0.421	0.011	0.063	-	0.146
3-methylpentane	0.395	-	-	-	0.033
2,3,3-trimethyl-pentane	0.395	-	-	0.171	0.016
2-methylbutane	0.368	-	-	0.024	-
indane	0.368	-	0.594	-	0.098
ethylcyclohexane	0.368	-	0.031	-	0.122
2,3-dimethyl-pentane	0.342	-	0.031	0.049	0.033
1,5-dimethylnaphthalene	0.342	-	0.094	-	0.049
propylcyclohexane	0.342	-	-	-	0.138
methylcyclopentane	0.316	-	-	-	0.024
(trans)-1-ethyl-4-methylcyclohexane	0.316	-	-	-	0.122
Ethanol	0.289	-	-	-	0.041
p-alphadimethylstyrene	0.289	-	0.031	-	-
ethylbenzene	0.289	-	0.313	-	0.089
3-methylheptane	0.289	-	-	0.073	0.122
4-Nonene	0.263	-	-	-	-
n-undecane	0.263	0.024	0.031	0.024	0.130
cyclohexane	0.237	-	-	-	0.024
n-heptane	0.237	-	0.031	0.024	0.065
n-nonane	0.211	-	0.125	-	0.081
1,2,4-trimethyl cyclohexane	0.211	-	-	0.146	0.073
4-methyldecane	0.211	0.011	0.031	0.463	0.130

COMPOUND	GAS	NORMA	AR	ISO	OXY
2-methylundecane	0.158	0.008	0.063	0.293	0.114
2,2,4-trimethylpentane	0.132	-	-	0.098	-
2-methylhexane	0.132	-	0.031	0.049	0.049
3-methylhexane	0.132	-	0.031	0.049	0.033
3-methyl-5-propylnonane	0.132	0.003	0.031	0.707	-
Isopropylbenzene	0.132	-	0.375	-	0.073
o-ethyltoluene	0.132	-	0.438	-	0.073
2,4,4-trimethyl-2-pentene	0.105	-	-	-	-
ethylcyclopentane	0.105	-	-	-	0.016
2,2,6-trimethyloctane	0.105	-	-	0.537	-
(cis)-1-ethyl-4-methylcyclohexane	0.105	-	-	-	0.106
n-tridecane	0.079	0.026	0.031	0.171	0.041
2-methylpentane	0.053	-	-	0.024	0.033
2,2-dimethylbutane	0.053	-	-	0.024	0.024
1,1,3-trimethylcyclohexane	0.026	-	-	-	0.122
decahydro-2-methylnaphthalene	0.026	-	0.031	-	0.073
Hexanal	-	-	-	-	-
p-Cresol	-	-	-	-	-
1-undecene	-	-	-	-	-
1-dodecene	-	-	-	-	-
Tri(propylene glycol)methyl ether mixture of isomers-B	-	-	-	-	-
1-Tetradecene	-	-	-	-	-
Dodecanal	-	-	-	-	-
1-Dodecanol	-	-	-	-	-
Anthracene	-	-	-	-	-
Diethylether	-	-	-	-	0.008
2,2-oxybisethanol	-	-	-	-	0.008
alphamethylstyrene	-	-	-	-	0.008
Benzylchloride	-	-	-	-	0.008

COMPOUND	GAS	NORMA	AR	ISO	OXY
2-methoxyphenol	-	-	-	-	0.008
1-decanol	-	-	-	-	0.008
Cedrol	-	-	-	-	0.008
Benzylbenzoate	-	-	-	-	0.008
9-Octadecenoicacid(Z)-methylester	-	-	-	-	0.008
2,4,6-trimethyl-1-nonene (meso form)(D)	-	-	-	-	-
Decamethylcyclopentasiloxane	-	-	-	-	-
1,3-bis(1-methylethyl)benzene	-	-	-	-	-
betapinene	-	-	-	-	0.008
DL-Camphor	-	-	-	-	0.008
Longifolene	-	-	-	-	0.008
Diethylphthalate	-	-	-	-	0.008
n-octadecane	-	-	-	-	-
Tetrahydrofuran	-	-	-	-	0.016
Di(propylene glycol methyl ether mixture of isomers-C	-	-	-	-	0.016
2-butoxyethyl,acetate	-	-	-	-	0.016
p-anisaldehyde	-	-	-	-	0.016
Caryophyllene	-	-	-	-	0.016
bis(2-ethylhexyl)adipate	-	-	-	-	0.016
2-ethyl-1-hexanol	-	-	-	-	-
o-chlorotoluene	-	-	-	-	-
Isobutylisobutyrate	-	-	-	-	0.008
Isobornylacetate	-	-	-	-	0.008
1-methoxy-2-propanol	-	-	-	-	0.016
Pentanal	-	-	-	-	0.016
Diethylene glycol monoethyl ether	-	-	-	-	0.016
1-Pentanol	-	-	-	-	0.008
Butyl,acetate	-	-	-	-	0.024
2-heptanone	-	-	-	-	0.024

COMPOUND	GAS	NORMA	AR	ISO	OXY
2,2,8-trimethyl-decane	-	-	-	0.024	-
alpha-methyl-trans-cinnamaldehyde	-	-	-	-	0.024
Benzyl alcohol	-	-	-	-	0.016
3-carene	-	-	-	-	0.016
n-Pentane	-	-	-	-	0.016
Benzonitrile	-	-	0.031	-	-
Malathion	-	-	0.031	-	-
Methylene,chloride	-	-	-	-	0.033
2-methyl-1-propanol	-	-	-	-	0.033
Methoxy-3-methylbutanol	-	-	-	-	0.033
Benzaldehyde	-	-	-	-	0.033
Methyl ester octadecanoic acid	-	-	-	-	0.033
Acenaphthene	-	-	0.031	-	-
Ethyl vanillin	-	-	-	-	0.033
Nitromethane	-	-	-	-	0.041
Phenylethyl alcohol	-	-	-	-	0.041
Lilial	-	-	-	-	0.041
9,12-Octadecadienoic acid-(Z,Z)-methyl ester	-	-	-	-	0.041
Tri(propylene glycol) methyl ether mixture of isomers-D	-	-	-	-	0.016
Tri(propylene glycol) methyl ether mixture of isomers -E	-	-	-	-	0.016
Dimethyl glutarate	-	-	-	-	0.041
Octanal	-	-	-	-	0.016
Di(propylene glycol) methyl ether acetate mixture of isomers-A	-	-	-	-	0.049
Di(propylene glycol) methyl ether acetate mixture of isomers -B	-	-	-	-	0.049
Tri(propylene glycol) methyl ether mixture of isomers-A	-	-	-	-	0.016
Methyl ester hexadecanoic acid	-	-	-	-	0.057
Alpha terpinene	-	-	-	-	-
Propylene glycol butyl ether	-	-	0.031	-	0.033
1-butoxy-2-propanol	-	-	0.031	-	0.033

COMPOUND	GAS	NORMA	AR	ISO	OXY
Tri(propylene glycol) methyl ether mixture of isomers-C	-	-	-	-	0.016
Isopropanol	-	-	-	-	0.065
1-Tridecene	-	-	-	-	-
Benzyl acetate	-	-	-	-	0.049
Phthalic acid anhydride	-	-	-	-	-
Alpha terpinolene	-	-	-	-	0.008
Biphenyl	-	-	-	-	0.008
Di(propylene glycol) methyl ether mixture of isomers-B	-	-	0.031	0.024	0.041
n-heneicosane	-	-	-	-	-
n-Hexyl acetate	-	-	0.063	-	0.041
Di(propylene glycol) methyl ether mixture of isomers-A	-	-	0.031	0.024	0.049
Alpha terpineol	-	-	-	-	0.041
fluorene	-	-	0.031	-	-
n-eicosane	-	-	-	-	0.024
1-Methoxy-2-propyl,acetate	-	-	0.063	-	0.073
2-butanone	-	-	-	-	0.130
p-xylene	-	-	0.125	-	0.008
Ethyl acetate	-	-	0.063	-	0.049
Acetone	-	-	-	-	0.122
Methanol	-	-	0.031	-	0.130
Methyl Isobutyl Ketone	-	-	0.063	-	0.073
Alpha pinene	-	-	-	0.024	0.041
phytane	-	-	-	-	0.008
n-heptadecane	-	0.008	-	-	-
1-decene	-	-	-	-	0.016
2-butoxy ethanol	-	-	-	-	0.171
Limonene	-	0.003	-	0.024	0.114
(cis)-Decahydronaphthalene	-	-	-	-	0.008
n-hexadecane	-	0.013	-	-	0.024

COMPOUND	GAS	NORMA	AR	ISO	OXY
pristane	-	-	-	-	0.024
n-pentadecane	-	0.018	0.031	-	0.033
n-tetradecane	-	0.021	0.031	-	0.049
1,3,5-trimethyl cyclohexane	-	-	-	0.049	0.073
(trans)-Decahydronaphthalene	-	-	0.031	-	0.081
n-dodecane	-	0.032	0.031	-	0.114
2,6-dimethyl undecane	-	0.008	0.031	-	0.081
COMPOUND	**LPD**	**MPD**	**HPD**	**NAP**	**MISC**
m-ethyltoluene	0.074	0.450	0.794	0.118	0.420
naphthalene	-	0.092	0.691	-	0.210
1,2,4,5-tetramethylbenzene	-	0.308	0.765	0.118	0.280
4-ethyl-1,2-dimethylbenzene	-	0.392	0.691	0.176	0.318
1,2,3,5-tetramethylbenzene	-	0.492	0.809	0.235	0.369
m-xylene	0.148	0.300	0.750	0.059	0.465
1,2,4-trimethylbenzene	0.037	0.342	0.632	0.118	0.318
1,2,3-trimethylbenzene	0.074	0.375	0.691	0.118	0.299
1,3,5-trimethylbenzene	0.074	0.583	0.794	0.118	0.408
p-ethyltoluene	0.074	0.458	0.779	0.118	0.408
methylcyclohexane	0.963	0.175	0.500	-	0.325
o-xylene	0.222	0.417	0.750	-	0.490
n-octane	0.963	0.583	0.691	-	0.382
2,5-dimethyl,hexane	0.889	0.008	-	-	0.217
toluene	0.370	0.225	0.603	-	0.490
Benzene	0.222	0.083	0.103	-	0.096
2-methylnaphthalene	-	0.108	0.765	-	0.197
2-methylheptane	0.963	0.250	0.412	-	0.325
2-methyloctane	0.778	0.800	0.544	-	0.363
1-methylnaphthalene	-	0.025	0.500	0.059	0.096

COMPOUND	LPD	MPD	HPD	NAP	MISC
2,3,5-trimethyl-hexane	0.667	0.092	0.015	-	0.191
n-propylbenzene	-	0.392	0.706	0.118	0.363
2,6-dimethyloctane	0.630	0.858	0.779	0.588	0.497
2,4-dimethylhexane	0.889	0.017	-	-	0.255
4-methylnonane	0.519	0.858	0.809	0.647	0.427
2,3,4-trimethylpentane	0.815	0.017	-	-	0.172
2,3-dimethyl-naphthalene	-	-	0.779	0.059	0.121
2-methyl,nonane	0.519	0.875	0.809	0.706	0.420
n-decane	0.370	0.883	0.853	0.235	0.465
n-hexane	0.222	0.008	0.015	-	0.076
2,4-dimethylpentane	0.370	-	-	-	0.121
(cis)-1,3-dimethylcyclohexane	0.926	0.400	0.471	-	0.306
2,2,5-trimethylhexane	0.296	-	-	-	0.121
3-methyldecane	0.074	0.975	0.868	0.941	0.446
3-methylpentane	0.148	0.008	0.015	-	0.096
2,3,3-trimethyl-pentane	0.519	0.008	-	-	0.083
2-methylbutane	0.037	-	-	-	0.013
indane	-	0.242	0.441	-	0.261
ethylcyclohexane	0.926	0.642	0.706	-	0.395
2,3-dimethyl-pentane	0.259	-	0.029	-	0.166
1,5-dimethylnaphthalene	-	-	0.750	0.059	0.096
propylcyclohexane	0.667	0.867	0.794	0.529	0.497
methylcyclopentane	0.259	-	0.015	-	0.057
(trans)-1-ethyl-4-methylcyclohexane	0.741	0.858	0.632	0.529	0.427
Ethanol	-	-	-	-	-
p-alphadimethylstyrene	-	0.033	0.044	-	0.045
ethylbenzene	0.037	0.175	0.426	-	0.287
3-methylheptane	0.815	0.317	0.485	-	0.318
4-Nonene	-	-	-	-	-

COMPOUND	LPD	MPD	HPD	NAP	MISC
n-undecane	0.111	0.683	0.824	0.412	0.414
cyclohexane	0.185	-	0.074	-	0.076
n-heptane	0.556	0.108	0.265	-	0.210
n-nonane	0.556	0.667	0.485	0.059	0.318
1,2,4-trimethyl cyclone	0.630	0.692	0.426	0.235	0.318
4-methyldecane	0.148	0.917	0.838	0.824	0.446
2-methylundecane	-	0.675	0.824	0.941	0.414
2,2,4-trimethylpentane	-	-	-	-	0.045
2-methylhexane	0.222	-	0.044	-	0.115
3-methylhexane	0.333	0.008	0.088	-	0.121
3-methyl-5-propylnonane	-	-	0.015	-	0.064
Isopropylbenzene	-	0.217	0.191	-	0.159
o-ethyltoluene	-	0.217	0.265	-	0.146
2,4,4-trimethyl-2-pentene	-	-	-	-	0.006
ethylcyclopentane	0.148	0.033	0.074	-	0.083
2,2,6-trimethyloctane	-	0.008	0.015	-	0.070
(cis)-1-ethyl-4-methylcyclohexane	0.778	0.850	0.559	0.353	0.389
n-tridecane	-	0.217	0.529	0.235	0.159
2-methylpentane	-	-	-	-	-
2,2-dimethylbutane	-	-	-	-	0.013
1,1,3-trimethylcyclohexane	0.778	0.817	0.721	-	0.382
decahydro-2-methylnaphthalene	-	0.683	0.838	0.941	0.350
Hexanal	-	-	-	-	0.006
p-Cresol	-	-	-	-	0.006
1-undecene	-	-	-	-	0.006
1-dodecene	-	-	-	-	0.006
Tri(propylene glycol),methyl ether mixture of isomers-B	-	-	-	-	0.006
1-Tetradecene	-	-	-	-	0.006
Dodecanal	-	-	-	-	0.006

COMPOUND	LPD	MPD	HPD	NAP	MISC
1-Dodecanol	-	-	-	-	0.006
Anthracene	-	-	-	-	0.006
Diethylether	-	-	-	-	-
2,2-oxybisethanol	-	-	-	-	-
alphamethylstyrene	-	-	-	-	-
Benzylchloride	-	-	-	-	-
2-methoxyphenol	-	-	-	-	-
1-decanol	-	-	-	-	-
Cedrol	-	-	-	-	-
Benzylbenzoate	-	-	-	-	-
9-Octadecenoicacid(Z)-methyl ester	-	-	-	-	-
2,4,6-trimethyl-1-nonene (meso form)(D)	-	0.008	-	-	-
Decamethylcyclopentasiloxane	-	0.008	-	-	-
1,3-bis(1-methylethyl)benzene	-	-	-	-	0.013
betapinene	-	-	-	-	0.006
DL-Camphor	-	-	-	-	0.006
Longifolene	-	-	-	-	0.006
Diethylphthalate	-	-	-	-	0.006
n-octadecane	-	-	0.015	-	-
Tetrahydrofuran	-	-	-	-	-
Di(propylene,glycol) methyl ether mixture of isomers-C	-	-	-	-	-
2-butoxyethyl,acetate	-	-	-	-	-
p-anisaldehyde	-	-	-	-	-
Caryophyllene	-	-	-	-	-
bis(2-ethylhexyl)adipate	-	-	-	-	-
2-ethyl-1-hexanol	-	0.017	-	-	-
o-chlorotoluene	-	-	-	-	0.019
Isobutylisobutyrate	-	-	-	-	0.013
Isobornylacetate	-	-	-	-	0.013

COMPOUND	LPD	MPD	HPD	NAP	MISC
1-methoxy-2-propanol	-	-	-	-	0.006
Pentanal	-	-	-	-	0.006
Diethylene glycol monoethyl ether	-	-	-	-	0.006
1-Pentanol	-	-	0.015	-	-
Butyl,acetate	-	-	-	-	-
2-heptanone	-	-	-	-	-
2,2,8-trimethyl-decane	-	-	-	-	-
alpha-methyl-trans-cinnamaldehyde	-	-	-	-	-
Benzyl alcohol	-	-	-	-	0.013
3-carene	-	-	-	-	0.013
n-Pentane	-	0.008	-	-	0.006
Benzonitrile	-	-	-	-	-
Malathion	-	-	-	-	-
Methylene,chloride	-	-	-	-	-
2-methyl-1-propanol	-	-	-	-	-
Methoxy-3-methylbutanol	-	-	-	-	-
Benzaldehyde	-	-	-	-	-
Methyl ester octadecanoic acid	-	-	-	-	-
Acenaphthene	-	-	-	-	0.006
Ethyl vanillin	-	-	-	-	0.006
Nitromethane	-	-	-	-	-
Phenylethyl alcohol	-	-	-	-	-
Lilial	-	-	-	-	-
9,12-Octadecadienoic acid-(Z,Z)-methyl ester	-	-	-	-	-
Tri(propylene glycol) methyl ether mixture of isomers-D	-	-	-	-	0.025
Tri(propylene glycol) methyl ether mixture of isomers -E	-	-	-	-	0.025
Dimethyl glutarate	-	-	-	-	0.006
Octanal	-	-	-	-	0.032
Di(propylene glycol) methyl ether acetate mixture of isomers-A	-	-	-	-	-

COMPOUND	LPD	MPD	HPD	NAP	MISC
Di(propylene glycol) methyl ether acetate mixture of isomers -B	-	-	-	-	-
Tri(propylene glycol) methyl ether mixture of isomers-A	-	0.008	-	-	0.032
Methyl ester hexadecanoic acid	-	-	-	-	-
Alpha terpinene	-	0.008	-	-	0.057
Propylene glycol butyl ether	-	-	-	-	0.006
1-butoxy-2-propanol	-	-	-	-	0.006
Tri(propylene glycol) methyl ether mixture of isomers -C	-	0.017	-	-	0.038
Isopropanol	-	-	-	-	0.006
1-Tridecene	-	0.008	-	0.059	0.006
Benzyl acetate	-	-	-	-	0.025
Phthalic acid anhydride	-	0.042	-	-	0.038
Alpha terpinolene	-	0.008	-	-	0.064
Biphenyl	-	-	0.074	-	-
Di(propylene glycol) methyl ether mixture of isomers -B	-	-	-	-	0.006
n-heneicosane	-	-	0.088	-	0.019
n-Hexyl acetate	-	-	-	-	0.006
Di(propylene glycol) methyl ether mixture of isomers -A	-	-	-	-	0.006
Alpha terpineol	-	-	-	-	0.076
fluorene	-	-	0.103	-	0.025
n-eicosane	-	-	0.118	-	0.019
1-Methoxy-2-propyl,acetate	-	-	-	-	0.025
2-butanone	-	-	-	-	0.032
p-xylene	-	-	0.015	-	0.019
Ethyl acetate	-	0.008	-	-	0.051
Acetone	0.037	-	-	-	0.019
Methanol	-	-	-	-	0.025
Methyl Isobutyl Ketone	-	-	-	-	0.057
Alpha pinene	-	0.008	-	-	0.140
phytane	0.037	-	0.162	-	0.025

COMPOUND	LPD	MPD	HPD	NAP	MISC
n-heptadecane	-	-	0.206	-	0.019
1-decene	-	0.183	-	-	0.057
2-butoxy ethanol	-	0.008	-	-	0.108
Limonene	-	0.008	0.044	-	0.153
(cis)-Decahydronaphthalene	-	0.033	-	0.412	0.019
n-hexadecane	-	-	0.382	-	0.057
pristane	0.037	-	0.515	0.118	0.108
n-pentadecane	-	0.008	0.721	0.118	0.134
n-tetradecane	-	0.033	0.794	0.412	0.185
1,3,5-trimethyl cyclohexane	0.778	0.625	0.015	0.353	0.248
(trans)-Decahydronaphthalene	-	0.850	0.529	0.941	0.312
n-dodecane	-	0.717	0.912	0.706	0.395
2,6-dimethyl undecane	-	0.592	0.897	-	0.338

5.3. Classification of the Compounds Found in Substrates and Ignitable Liquids Based on the Compound Type

The compounds present in the standard library (294) were categorized based on their organic compound type. According to this classification, there are 44 types of different compound types. These compound types are given in Table 8. The compounds analyzed using logistic regression in substrates and ignitable liquids were categorized using these compound types.

Table 9: Compound types in the standard mass spectral library

Alcohol	Benzimidazoles	Nitro alkane
Aldehyde	Bicyclic sesquiterpene	PAH
Alkane	Branched alkane	Phenol
Alkene	Branched alkene	Phenyl propanoids
Alkyl halide	Carboxylic acid	Phenyl propene
Amide	Cycloalkane	Pyrrole
Anhydride	Cycloalkene	Sesquiterpene alcohol
Aromatic alcohol	Ester	Siloxane
Aromatic aldehyde	Ether	Terpene
Aromatic amide	Furan	Thiozole
Aromatic carboxylic acid	Halogenated alcohol	
Aromatic ester	Halogenated aromatic hydrocarbon	
Aromatic hydrocarbon (including styrenes)	Halogenated branched alkane	
Aromatic ketone	Indane	
Aromatic nitrile	Indene	
Aryl halide	Ketone	
Aryl halide ester	Nitrile	

The distributions of these compound types in substrates and ignitable liquids are given in Figure 1.

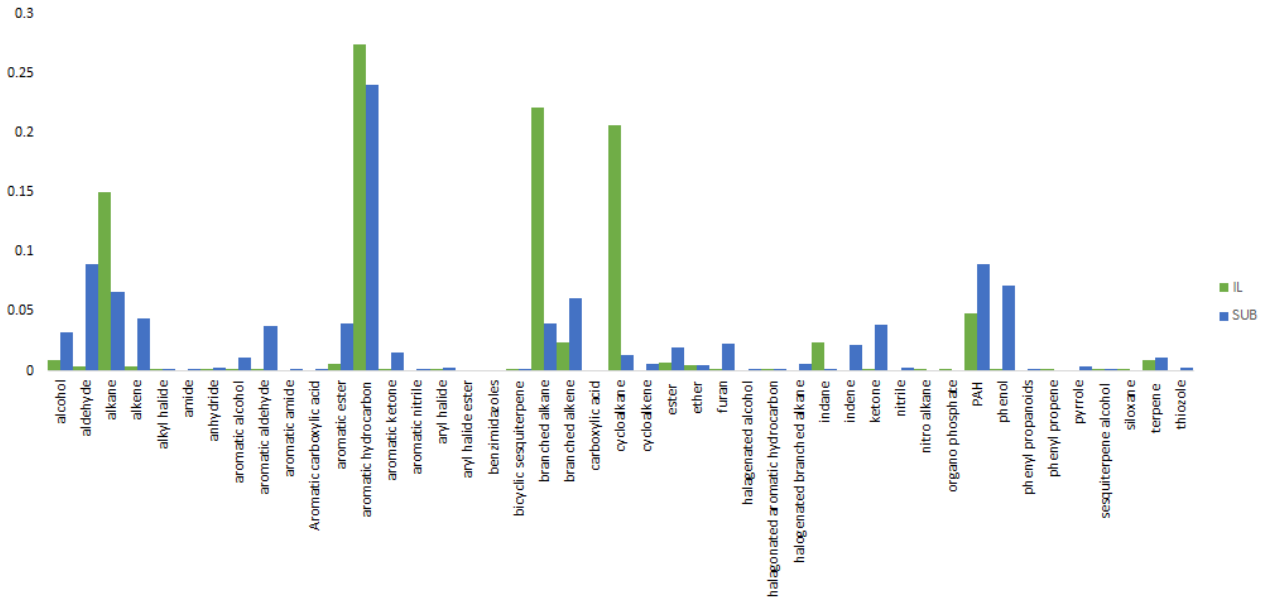


Figure 58: Distribution of compound types in ignitable liquids and substrates

According to the distribution depicted in Figure 58, alkanes, aromatic hydrocarbons, cycloalkanes and PAH (poly nuclear aromatic hydrocarbons) have high frequencies in ignitable liquids whereas in substrates, aldehydes, alkanes, aromatic hydrocarbons, branched alkenes, ketones, PAH and phenols have high frequencies. The classification of these compounds is provided in Appendix A. Distribution of compound types in each ignitable liquid classes is provided in Figure 59, 60, 61 and 62.

In ignitable liquids, there were 12473 peaks identified in total, whereas in substrates 9460 total peaks were identified. The frequency of the compound types was calculated by the equation 5.1 below. In this equation alcohol was provided as an example. In IL and SUB the frequencies of alcohols are 0.009 and 0.033 respectively. The data tables used for this chart is also provided in Appendix A.

(5.1)

$$Frequency\ of\ an\ identified\ compound\ type\ in\ IL\ (or\ SUB)\ (eg{:}alcohol) = \frac{Total\ number\ of\ alcohol\ peaks\ identified\ in\ IL\ (or\ SUB)}{Total\ number\ of\ peaks\ identified\ in\ IL\ (or\ SUB)}$$

Total number of peaks identified in each ignitable liquid class are tabulated in Table 10.

Table 10: Total Number of peaks identified in each IL class

IL CLASS	TOTAL NUMBER OF PEAKS IDENTIFIED IN EACH CLASS
AROMATIC	393
GASOLINE	1202
HEAVY PETROLEUM DISTILLATE	2406
ISOPARAFFINIC	206
LIGHT PETROLEUM DISTILLATE	591
MEDIUM PETROLEUM DISTILLATE	3245
MISCELLANIOUS	2999
NORMAL ALKANE	76
NAPHTHENIC PARAFFINIC	234
OXYGENATED	1121

The frequency of compound types identified in each IL class were calculated by Equation 5.2.

(5.2)

$$Frequency\ of\ an\ identified\ compound\ type\ in\ AR\ (or\ GAS, HPD\ etc.) = \frac{Total\ number\ of\ alcohol\ peaks\ identified\ in\ AR\ (or\ GAS, HPD\ etc.)}{Total\ number\ of\ peaks\ identified\ in\ AR\ (or\ GAS, HPD\ etc.)}$$

Frequency of alcohols in aromatic IL class (AR) is 2/393 = 0.0051. The data tables used for the graphs below are also provided in Appendix 2.

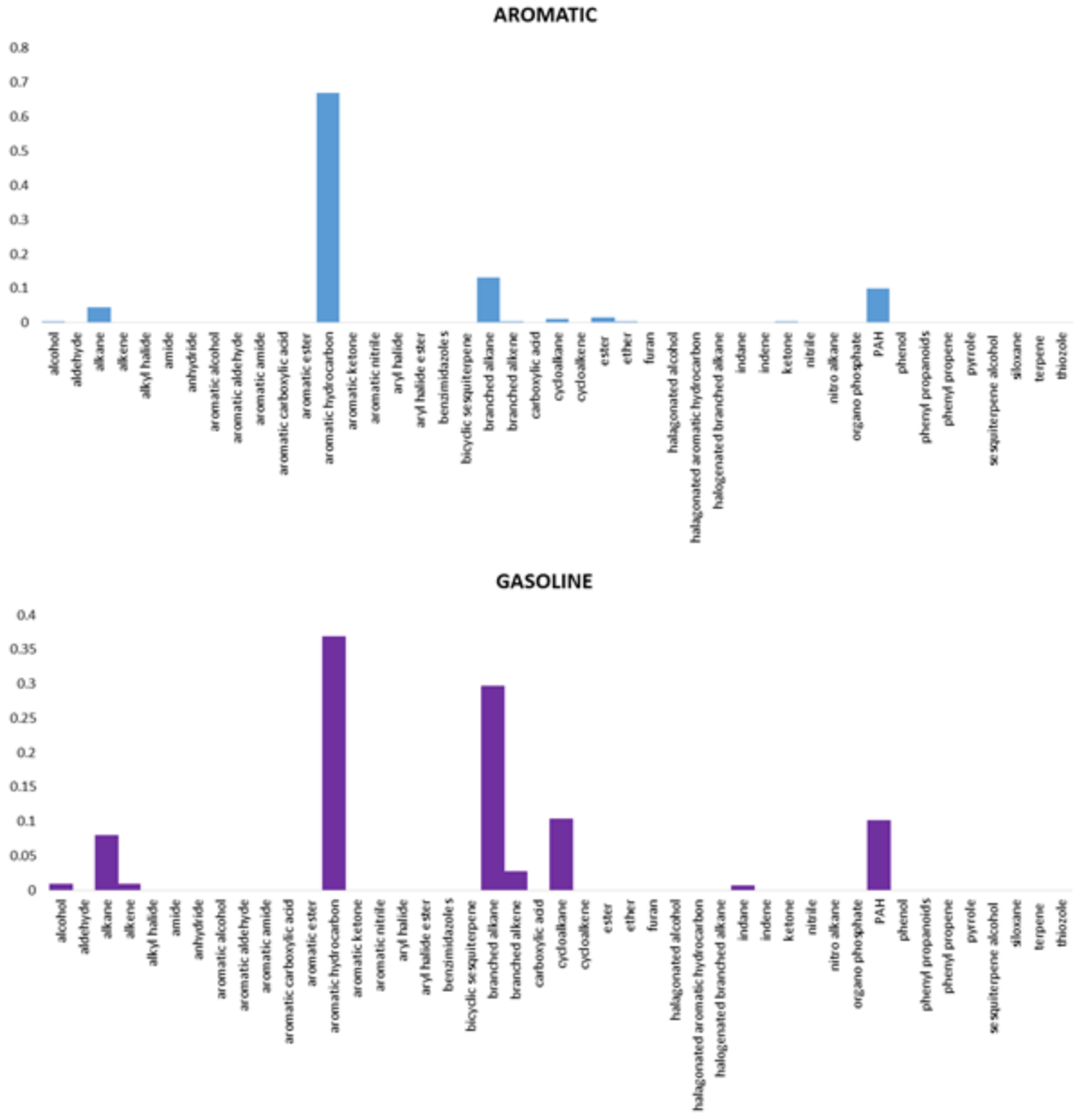


Figure 59: Compound types seen in aromatic and gasoline classes

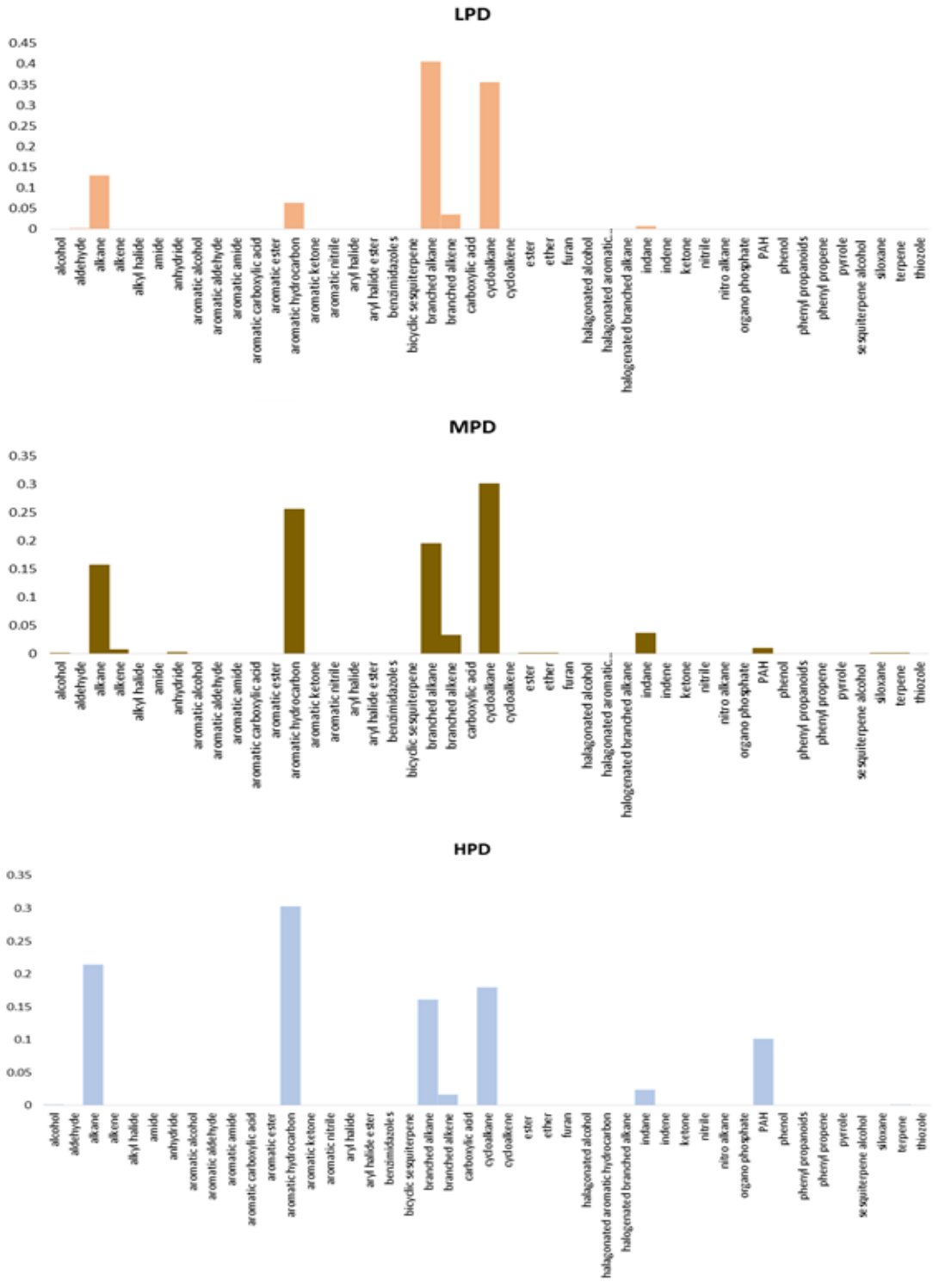


Figure 60: Compound types identified in the petroleum distillate classes

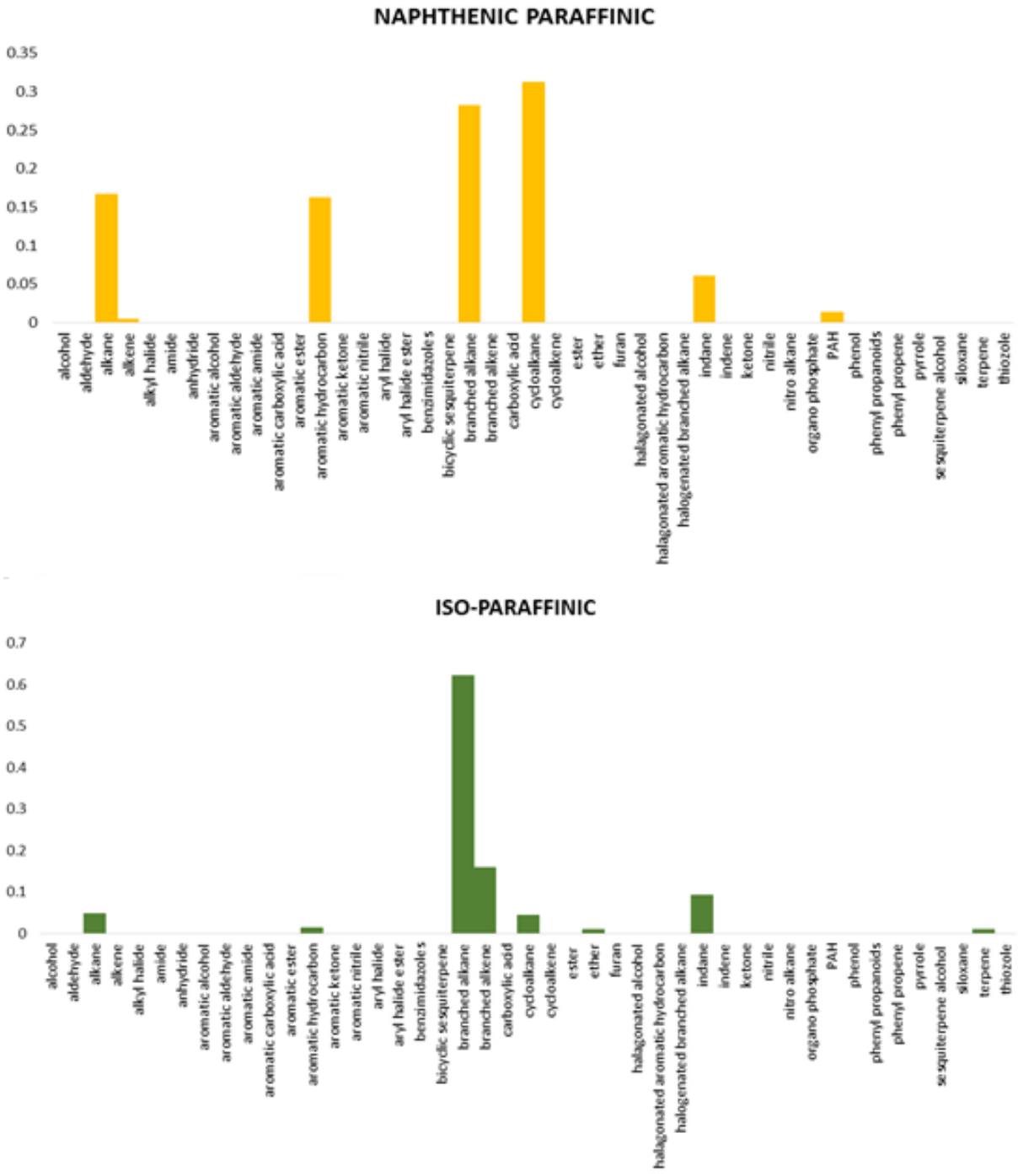


Figure 61: Compound types identified in naphthenic paraffinic and iso-paraffinic ignitable liquid classes.

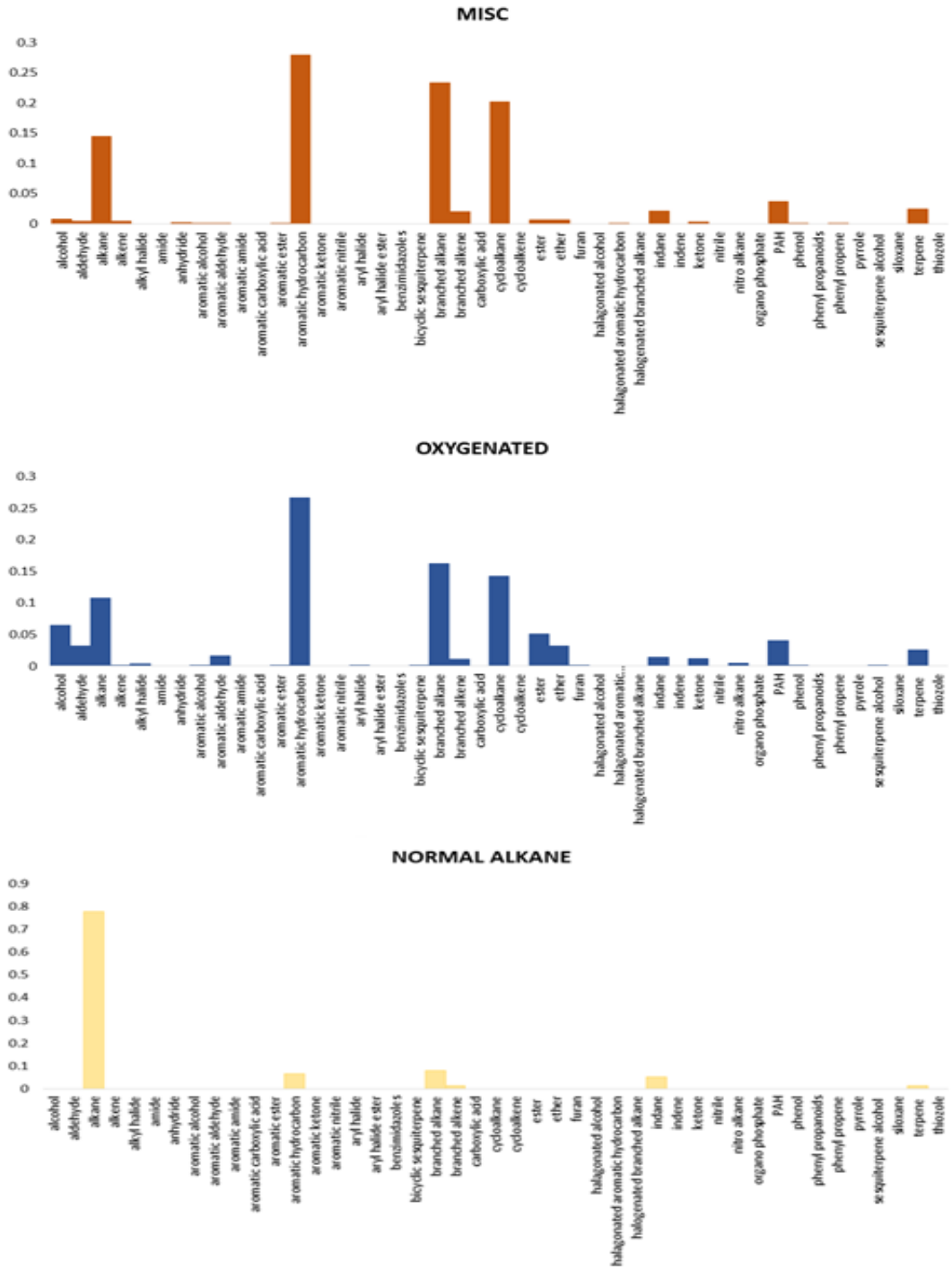


Figure 62: Compound types distribution in miscellaneous, oxygenated and normal alkane ignitable liquid classes

The major compound types in aromatic and gasolines are aromatic hydrocarbons (ex: toluene, xylenes, tri-methyl and tetra-methyl benzenes). In gasoline the other major compound types are iso-parrafins, PAH (polynuclear aromatic hydrocarbons) and cycloalkanes (Figure 59). In petroleum distillates (LPD, MPD and HPD), the major compound types are aromatic hydrocarbons, cycloalkanes, iso-paraffins and alkanes (Figure 60).

In naphthenic-paraffinic IL class, the major compound types identified were branched alkanes, cycloalkanes, aromatic hydrocarbons, alkanes, PAH and indanes and in iso-paraffinic class the major compound types were branched alkanes and branched alkenes (Figure 61). Miscellaneous and oxygenated ignitable classes have a variety of distributions of compound types. However, they have higher frequencies in the compound types of aromatic hydrocarbons, iso-paraffins, cycloalkanes and alkanes whereas in normal alkanes have a high abundance of alkanes (Figure 62).

CHAPTER 6: VALIDATION OF THE METHODS

6.1. Validation of the Naïve Bayes LLR Calculation Method

The log-likelihood ratio (LLR) calculation method using Naïve Bayes was validated using 16 laboratory-prepared fire debris samples. These samples were a mixture of substrates and substrates with ignitable liquids. Details of these fire debris samples are presented in Table 11.

Table 11: Details of the laboratory-prepared fire debris samples

Sample	Substrate	IL (SRN) 75% Evaporated	IL to Sub
A (1)	olefin carpet and padding	None	0
B (2)	leather jacket	120 (ISO)	3.5
C (3)	vinyl flooring	259 (GAS)	1
D (4)	milk jug and duct tape	None	0
E (5)	roofing shingle	46 (MPD)	1.76
F (6)	vinyl flooring	None	0
G (7)	polyester carpet	None	0
H (8)	polyester carpet	120 (ISO)	0.25
I (9)	olefin carpet and padding	73 (AR)	0.25
J (10)	laminate flooring and newspaper	None	0
K (11)	polyester carpet and padding	73 (AR)	1
L (12)	polyester carpet and padding	None	0
M (13)	leather jacket	None	0
N (14)	milk jug and duct tape	259 (GAS)	0.25
O (15)	laminate flooring and newspaper	46 (MPD)	1
P (16)	roofing shingle	None	0

*ISO-Isoparaffinic GAS-Gasoline MPD-Medium petroleum distillate AR-Aromatic

a) Validation of Method A (All Compounds in Substrates and Ignitable Liquids)

All the samples without ignitable liquid residue (ILR), and 2 samples containing ILR (C and E) were classified correctly from this method. The main reason for misclassification of the samples was that the fire debris samples (except I and K) contained some compounds that were only seen in substrates, and not in ignitable liquids. These compounds in which are not seen in the sample of ignitable liquids have been assigned the frequencies calculated by Good-Turing estimation, which were $\approx 10^{-5}$. Presence of these compounds drastically minimized the numerator in Equation 4.2 (Chapter 4), in turn, provided a smaller value (<1) for the likelihood ratio. This provided a negative LLR and affected the identification of fire debris samples which contained ILR. The fire debris sample numbers were projected to the initially generated calibrated ROC curve based on their calibrated LLR value. This is illustrated in Figure 63 and the AUC of this ROC plot was 0.99.

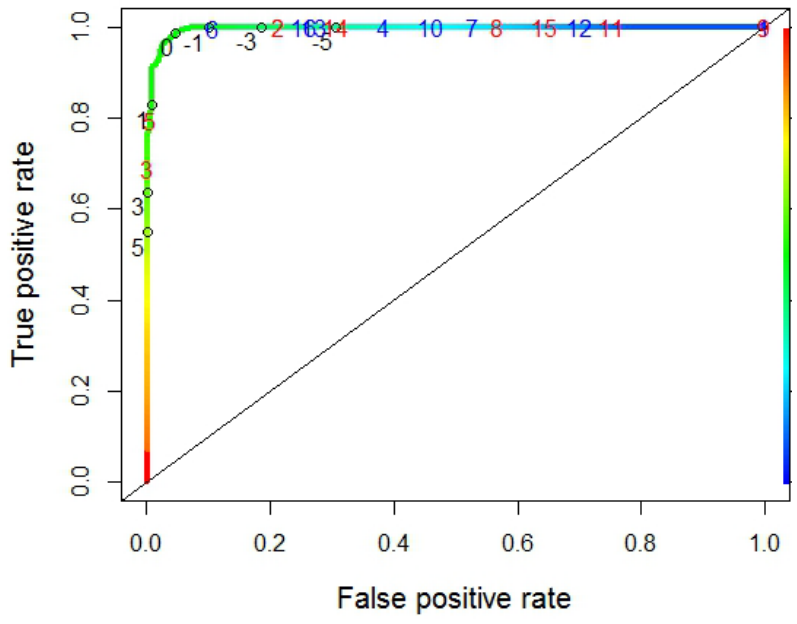


Figure 63: Projection of calculated LLR (method A) for fire debris samples to the ROC curve (red: samples with ILR, blue: samples without ILR)

a) Validation of Method B (Compounds in both Substrates and Ignitable Liquids Only)

This method correctly identified all the substrates without ILR and misclassified samples B, H, I and K. The reason for the misclassification in this method was the higher frequency of occurrence of some of the common compounds (seen in both SUB and IL) in substrates which reduced the numerator in Equation 4.2, which in turn provided a LLR value which is <1. Projection of the fire debris samples on the ROC curve based on their calibrated LLR is presented on Figure 64. The area under the curve obtained for this plot was 0.98.

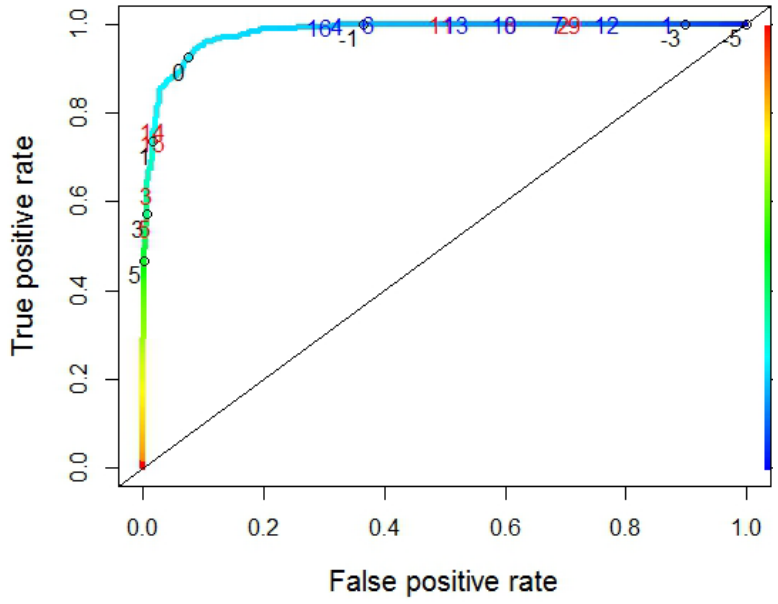


Figure 64: Projection of calculated LLR (method B) for fire debris samples to the ROC curve (red: samples with ILR, blue: samples without ILR)

b) Validation of Method C (All Compounds in Ignitable Liquids)

Out of all the three methods, this worked best in the classification of fire debris samples. This method only misclassified samples I and K which contained aromatic ILR. The results of these studies showed that some of the aromatic compounds have a higher frequency of occurrence in SUB than in IL. When the compounds present only in IL were included for the method, it drastically reduced the denominator since those compounds in substrates have estimated frequencies by Good-Turing method. This increased the LR ratio and provided a positive LLR value when ignitable liquid residue was present in the sample. Projection of the fire debris sample numbers on the ROC curve based on the calibrated LLR is presented in Figure 65, and the area under the curve of this ROC plot was 0.98.

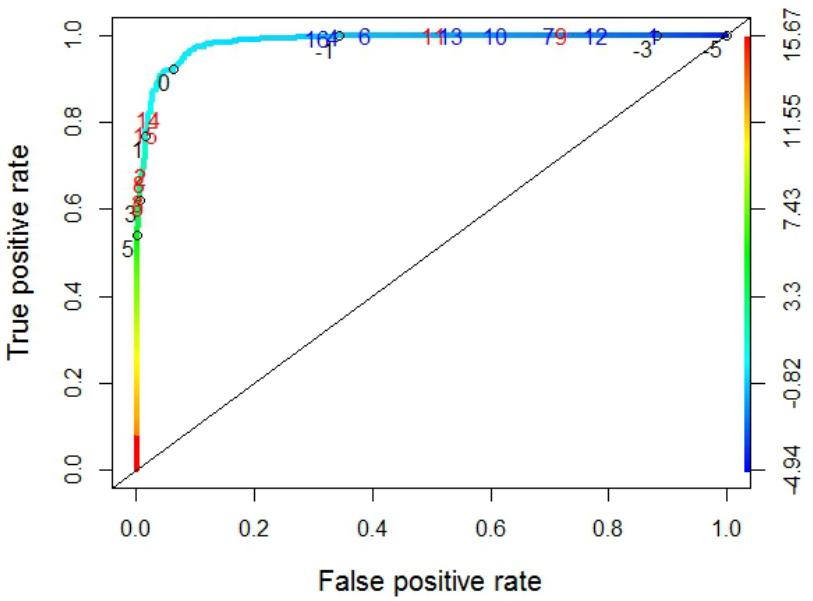


Figure 65: Projection of calculated LLR (method C) for fire debris samples to the ROC curve (red: samples with ILR, blue: samples without ILR)

Uncalibrated and calibrated log likelihood ratios calculated for these samples are presented in Table 12.

Table 12: Uncalibrated and calibrated log-likelihood ratios (LLR) obtained by the 3 methods described above

Sample	Ground Truth	Method (A)	Calibrated LLR (A)	Method (B)	Calibrated LLR (B)	Method (C)	Calibrated LLR (C)
A	SUB	-61.14	-15.65	-9.66	-2.79	-9.66	-2.97
B	SUB & IL	-16.03	-3.43	-7.13	-2.01	6.32	2.04
C	SUB & IL	6.67	2.35	6.93	2.29	10.3	3.28
D	SUB	-26.94	-6.21	-3.29	-0.84	-3.29	-0.97
E	SUB & IL	2.61	1.32	10.42	3.35	10.42	3.32
F	SUB	-7.54	-1.27	-3.91	-1.03	-3.91	-1.16
G	SUB	-35.49	-8.39	-7.02	-1.98	-7.02	-2.14
H	SUB & IL	-37.49	-8.90	-6.28	-1.75	7.17	2.31
I	SUB & IL	-62.04	-15.65	-7.32	-2.07	-7.32	-2.23
J	SUB	-31.42	-7.36	-6.14	-1.71	-6.14	-1.86
K	SUB & IL	-51.58	-12.50	-4.94	-1.34	-4.94	-1.49
L	SUB	-46.59	-11.22	-8.18	-2.33	-8.18	-2.50
M	SUB	-20.18	-4.49	-5.34	-1.47	-5.34	-1.61
N	SUB & IL	-22.18	-5.00	1.82	0.72	1.82	0.63
O	SUB & IL	-43.27	-10.38	2.77	1.02	2.77	0.93
P	SUB	-18.4	-4.04	-2.95	-0.73	-2.95	-0.86

As mentioned above in method A, calculation based on the compounds present only in substrates reduced the value of log likelihood ratios of the samples containing ignitable liquid residue. The samples A, D, F, G, I, J, K, L, M and P have the same negative uncalibrated LLR values in methods B and C, and samples E, N and O have same positive uncalibrated LLR. The reason for this similarity between LLRs was that these samples only contained some compounds which can be seen in both SUB and IL, therefore the calculated LLRs were the same in both

methods. The samples B and H obtained negative LLRs in method B but a positive value in method C. The reason for this was that those samples contained some compounds that were found only in ignitable liquids which affected the magnitude of the likelihood ratio.

6.2. Classification of Fire Debris Data without Frequency Adjustments

The classification model was validated on a large number of laboratory generated fire debris samples. The log likelihood ratios of these fire debris samples were calculated using the three sets of compounds described above. These fire debris samples contained 112 substrates and 293 of the substrate and ignitable liquid mixture burns. The substrate and ignitable liquid mixture burns contained ILR from all 10 IL classes. The summary of the classification of fire debris samples is presented in Table 13. The cutoff LLR is 0, hence if the calculated LLR ≥ 0, it was identified as positive for ILR whereas LLR < 0, was identified as negative for ILR.

Table 13: Summary of the classification of fire debris samples

	Using all compounds	Using compounds in SUB and IL	Using compounds in IL
SUB only	112/112	103/112	103/112
SUB and IL mixtures	43/293	107/293	130/293

It was apparent that all these sets of compounds performed well in the identification of pure substrate samples. In the identification of pure substrate samples, accuracy was the highest when all compounds were used for the calculation of log likelihood ratios. However, in this method the accuracy was the lowest (8.78%) in the identification of samples containing ILR. The accuracy was the highest (45.95%) in the identification of ILR in samples when compounds found in IL were used. Using the LLRs calculated by compounds in IL, the fire debris samples were

classified by the respective IL class. These results are tabulated in Table 14. According to these results, aromatic (AR) had the lowest percentage of correct classification (2.56%) whereas light petroleum distillate (LPD) and miscellaneous (MISC) had the highest percentage of 100% and 90% respectively.

Table 14: Classification of fire debris based on IL class by the calculated LLRs using compounds present in IL

IL Class	Correct	Incorrect	Total	Total %	Correct %	Incorrect %
AR	1	38	39	13.31	2.56	97.44
GAS	13	7	20	6.83	65.00	35.00
HPD	0	11	11	3.75	0.00	100.00
ISO	2	45	47	16.04	4.26	95.74
LPD	24	0	24	8.19	100.00	0.00
MISC	36	4	40	13.65	90.00	10.00
MPD	19	5	24	8.19	79.17	20.83
NA	5	34	39	13.31	12.82	87.18
NP	22	3	25	8.53	88.00	12.00
OXY	5	19	24	8.19	20.83	79.17

6.3. Validation of Fire Debris Samples Using Frequency Adjustments

Frequencies of compounds obtained for ignitable liquid classes (AR, GAS, HPD, ISO, LPD, MISC, MPD, NA, NP and OXY) and substrates were adjusted using three different distributions. They were,

1. IL and SUB distribution of known ground-truth real world fire debris data obtained from the Florida Fire Marshall
2. IL and SUB distribution of the substrate and ignitable liquid databases of the National Center for Forensic Science[8,6]
3. Equal distribution of substrate and ignitable liquids

The adjusted frequencies of the compounds present in ignitable liquids calculated by matrix multiplication of frequency of compounds in each IL class and the ratios of IL classes in each distribution mentioned above.

These IL classes and SUB ratio distributions are presented in Table 15. The adjusted frequencies of occurrence of compounds in substrates were calculated by multiplying the frequency of compounds in substrates with the ratio of substrates in each distribution. For all these calculations, heavy, medium and light petroleum distillates were combined as petroleum distillates.

Table 15: The ratios between IL classes and SUB in the 3 distributions mentioned above

	FL Fire Marshall	Database	Equal
Gasoline	0.33	0.033	0.0625
Normal alkane	0.003	0.017	0.0625
Aromatic	0.005	0.027	0.0625
Iso-paraffinic	0.003	0.036	0.0625
Oxygenated	0.012	0.109	0.0625
Petroleum distillates	0.062	0.190	0.0625
Naphthenic paraffinic	0.002	0.014	0.0625
Miscellaneous	0.058	0.125	0.0625
Substrates	0.525	0.448	0.5

6.3.1. Laplace Estimation

Laplace estimation is another technique which can be used to estimate the frequencies in unseen compounds (not observed in the sample) in ignitable liquids and substrates. The probability of the unseen species (P_c) is calculated by Equation 6.1. In this equation, n_c is the number of samples containing a specific compound, k is the smoothing parameter and N is the total number of samples of substrates or ignitable liquids. For this specific estimation of unseen compounds in IL or SUB, $k = 1$ and $n_c = 0$.

$$P_c = \frac{n_c + k}{N + n_c * k} \quad \textbf{(6.1)}$$

6.3.2. Calculation of Naïve Bayes Log-Likelihood Ratios

Three different compound sets were selected (as mentioned in Section 6.1) when calculating the log-likelihood ratios (LLR) using this approach. These compounds sets were,

1. All compounds in ignitable liquids and substrates
2. Compounds in ignitable liquids
3. Compounds in both substrates and ignitable liquids only

The same data set mentioned in Section 5.2 was used for the validation of these methods. In this data set, 112 samples contained only burned substrates whereas 293 samples were mixtures of substrates and ignitable liquids. As mentioned in Chapter 5, the total compounds present in substrates and ignitable liquids were 252 and 177 compounds were present only in ignitable liquids. The analysis of the calculated LLR by the four frequency sets (without frequency adjustments and frequency adjustment based on the 3 distributions given above) are presented in Table 16, 17 and 18. The results were separated based on the compound sets given above. These LLRs were not calibrated. If the calculated LLR is ≥ 0, it was considered a positive sample for ILR and if it was < 0, then it was considered a negative sample for ILR.

6.3.2.1. All Compounds in Ignitable Liquids and Substrates

Table 16: Summary of the analysis of LLRs calculated by all compounds in SUB and IL

	Without frequency adjustments	Equal dist. frequency adjustments	FL fire Marshall dist. frequency adjustments	Database dist. frequency adjustments
Samples containing IL (TP)	43 (14.7%)	113 (38.6%)	135 (46.1%)	116 (39.6%)
Samples without IL (TN)	112 (99.1%)	110 (98.2%)	109 (97.3%)	110 (98.2%)
Misclassified as a sample with IL (FP)	0 (100%)	2 (1.79%)	3 (1.02%)	2 (1.79%)
Misclassified as a sample without IL (FN)	250 (85.3%)	180 (61.4%)	158 (53.9%)	177 (60.4%)

When all the compounds identified in IL and SUB were used in the LLR calculation, the highest true positive rate (46.1 %) was observed in the frequency adjustments made by the data obtained from the FL Fire Marshall (Table 16).

6.3.2.2. Compounds Present in Ignitable Liquids

Table 17: Summary of the analysis of LLRs calculated by all compounds in IL

	Without frequency adjustments	Equal dist. frequency adjustments	FL fire Marshall dist. frequency adjustments	Database dist. frequency adjustments
Samples containing IL (TP)	137 (46.8%)	165 (56.3%)	234 ((79.9%)	196 (66.9%)
Samples without IL (TN)	103 (91.9%)	101 (90.2%)	89 (79.5%)	98 (87.5%)
Misclassified as a sample with IL (FP)	9 (8.03%)	11 (9.82%)	23 (20.5%)	14 (12.5%)
Misclassified as a sample without IL (FN)	156 (53.2%)	128 (43.7%)	59 (20.1%)	97 (33.1%)

As in the above section, the highest true positive rate (79.9 %) and the lowest true negative rate (79.5 %) were observed in the frequency adjustments made by the data obtained from the FL Fire Marshall. The highest true negative rate (91.9%) was observed in the LLRs calculated using the unadjusted frequencies (Table 17).

6.3.2.3. Compounds Present in both Ignitable Liquids and Substrates Only

Table 18: Summary of the analysis of LLRs calculated by compounds in both SUB and IL only

	Without frequency adjustments	Equal dist. frequency adjustments	FL fire Marshall dist. frequency adjustments	Database dist. frequency adjustments
Samples containing IL (TP)	110 (37.5%)	159 (54.3%)	227 (77.5%)	185 (63.1%)
Samples without IL (TN)	103 (91.9%)	101 (90.2%)	88 (78.6%)	99 (88.4%)
Misclassified as a sample with IL (FP)	9 (8.03%)	7 (6.25%)	24 (21.4%)	13 (11.6%)
Misclassified as a sample without IL (FN)	183 (62.5%)	127 (43.3%)	66 (22.5%)	108 (36.9%)

The highest true positive rate was obtained in the LLRs calculated by the frequency adjustments made by the data obtained from FL Fire Marshall (77.5 %). The lowest true negative rate was also obtained in the calculated LLRs by using this adjusted set of frequency (78.6%) (Table 18).

6.3.3. Fire Debris LLR Projection on Pure Substrate and Ignitable Liquids Data

In this section, the projection of the calculated LLRs for laboratory generated fire debris samples to the ROC curves obtained from the cross-validation of pure substrates and ignitable liquids will be discussed (Chapter 4, Section 4.2.). For this procedure, the likelihood ratios for fire debris samples were calculated by randomly selecting fire debris samples. Three sets selected from these samples based on the population distributions. The number of fire debris samples selected for the calculation of likelihood ratios from each distribution is given in Table 19.

Table 19: The number of samples selected for the calculation of likelihood ratios from each distribution

	Equal	FL Fire Marshall	Database
Gasoline	8	17	3
Normal alkane	8	1	2
Aromatic	8	1	3
Iso-paraffinic	8	1	4
Oxygenated	8	1	11
Petroleum distillates	8	3	19
Naphthenic paraffinic	8	1	1
Miscellaneous	8	3	13
Substrates	64	26	45
Total	128	54	101

6.3.3.1. Frequency Adjustments based on Equal Distribution

The ROC curves were generated using the pure substrates and ignitable liquid samples from the ILRC and Substrate Databases of the National Center for Forensic Science. The LLRs were calculated by the three methods mentioned above. They were the calculations of LLRs using all compounds, compounds in IL and compounds only in SUB and IL. The AUC values obtained for these three methods were 0.99, 0.98, and 0.98 respectively. The likelihood ratios for were calculated for 128 fire debris samples. Likelihood ratios and AUC calculation for selected fire debris samples were repeated 10 times. The AUC for the LLRs of 128 fire debris samples were calculated by the Wilcoxon-Mann-Whitney U-statistic[53] (Equation 6.2).

$$AUC = \frac{U}{N_{(p)} N_{(n)}} \quad \textbf{(6.2)}$$

In this equation, U is the Wilcoxon-Mann-Whitney U-statistic, N_p is the number of correct positive identification and N_n is the number of correct negative identification. The average AUC values obtained for these calculated fire debris samples are:

All compounds – 0.73 ± 0.032

Compounds in IL – 0.78 ± 0.034

Compounds in SUB and IL only – 0.75 ± 0.037

The calculated LLRs of 128 fire debris samples were projected on these ROC plots. Blue indicated the calculated LLRs for SUB and red indicates the calculated LLRs for IL. The LLRs calculated for fire debris samples were calibrated using the logistic regression model created by the training set of the pure substrates and ignitable liquids data (Section 4.2). The classification of the data is tabulated in Table 20. The TP and TN are the average of the reported values.

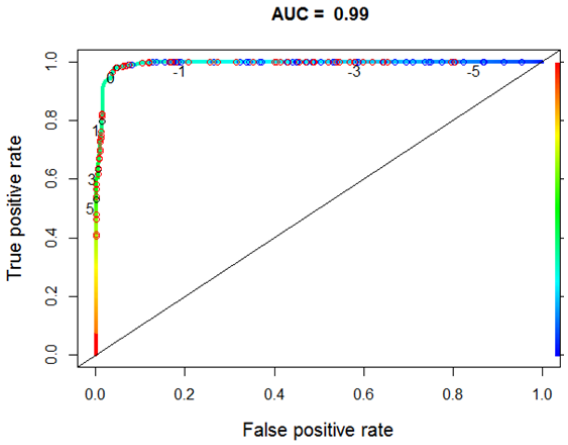
Table 20: Correct classification of SUB and SUB/IL mixture samples (equal distribution)

	TP (64)	**TN (64)**
All compounds	27 ± 2.4	63 ± 1.1
Compounds in IL	37 ± 3.3	62 ± 1.3
Compounds in IL and SUB	34 ± 3.2	61 ± 1.6

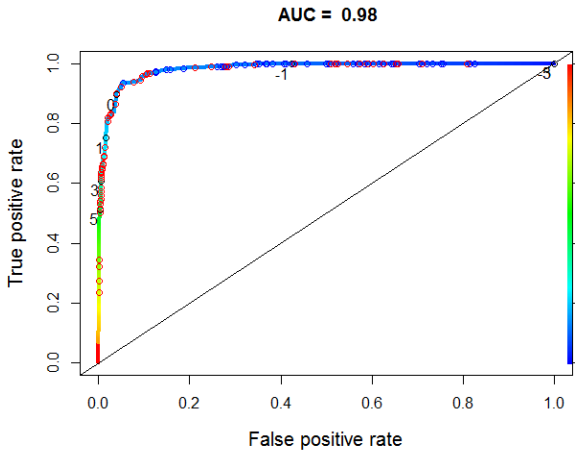
In this, TP is the correctly classified number of samples positive for ILR out of 64 samples whereas TN is the correctly classified number of samples negative for ILR. The ROC curves presented in Figure 66a, 66b and 66c are the projections of the log likelihood ratios of the final iteration (10) of the projection of data.

The calculation of AUC using Mann-Whitney U statistic indicates that the highest AUC for the calculated LLRs were obtained using the compounds in ignitable liquids. The generated ROC plot for these LLRs for 128 samples is depicted in Figure 67. The classification based on IL classes is provided in Table 21.

(a)



(b)



(c)

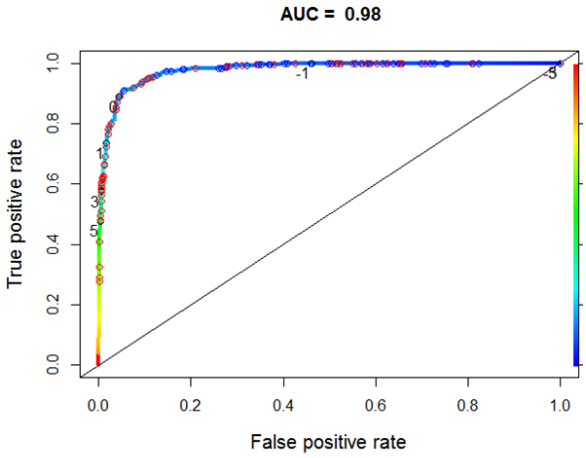


Figure 66: ROC plots obtained for the calculated LLRs for pure SUB and IL by equal distribution frequency adjustments using a) all compounds b) compounds in IL c) compounds only in both SUB and IL

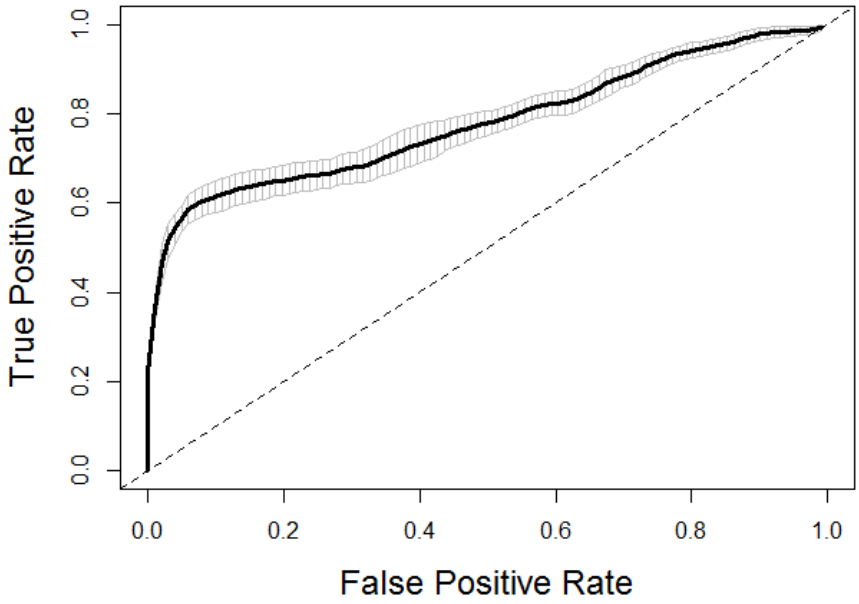


Figure 67: ROC curve generated from the calculated LLRs for 128 samples using compounds in IL (with 95% Confidence interval)

Table 21: Correct and incorrect IL class classification based on the calculated LLRs using compounds in ignitable liquids

IL class	Correct	Incorrect	Total	Correct %	Incorrect %
AR	2.7 ± 1.1	5.3 ± 1.1	8	33.8	66.3
GAS	5.7 ± 1.8	2.3 ± 1.8	8	71.3	28.8
PD	7.7 ± 0.48	0.3 ± 0.48	8	96.3	3.8
ISO	2.1 ± 1.7	5.9 ± 1.7	8	26.3	73.8
MISC	8	0	8	100	0
NP	8	0	8	100	0
NORMA	2.0 ± 1.1	6.0 ± 1.1	8	25	75
OXY	4.2 ± 0.79	3.8 ± 0.79	8	52.5	47.5
SUB	57.9 ± 1.6	6.1 ± 1.6	64	90.5	9.5

According to this data table, the calculated LLRs for GAS, PD, MISC, NP have a correct classification of more than 70% whereas AR, ISO, OXY and NORMA have a correct classification less than 60%. One of the possible reasons for the misclassification of AR IL class is, that the burned substrates contained more aromatic compounds and most of these compounds have higher frequencies of occurrence in substrates than in ignitable liquids. This also raised problem in the IL class normal alkane (NORMA). Majority of the normal alkanes seen in NORMA ignitable liquids can also be seen in substrates, and some of these compounds with a higher frequency of occurrence which in turn provide a low evidentiary value for the presence of ignitable liquids in the samples. Most of the compounds present in iso-paraffinic (ISO) ignitable liquid classes (used in this validation data set) were not identified due to the complexity of their chromatograms and the resolution of the mass spectral data.

6.3.3.2. Frequency Adjustments based on FL Fire Marshall Data Distribution

The AUC values obtained for the calculated LLRs for pure SUB and IL using these adjusted frequencies for the three methods (all compounds, compounds in IL and compounds only in IL and SUB) were 0.99, 0.96, and 0.95 respectively. The calculated LLRs of 54 fire debris samples were projected on these ROC plots and they are presented in Figure 67a, 67b and 67c. As above, AUC calculation for randomly selected fire debris samples were repeated 10 times. When selecting the number of samples for this method, the frequency distribution of NA, AR, ISO and NP were lower than the other IL classes, therefore, 1 from each of these classes were randomly selected for the calculations to represent all the IL classes. Due to this reason the original frequency distribution was changed in the sample selection. The calculated average AUC using Mann-Whitney-U-statistics for the LLRs calculated by the given sets of compounds (three methods) were,

All compounds – 0.88 ± 0.064

Compounds in IL – 0.92 ± 0.045

Compounds in SUB and IL only – 0.92 ± 0.03

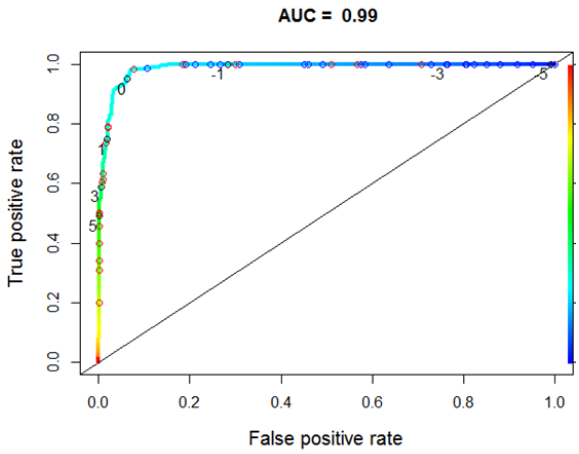
When these calculated AUC for pure SUB and IL were compared to the previous ROC curves obtained by the equal distribution adjusted frequencies, there is a decrease in the AUC. The number of TP and FP are tabulated in Table 21.

Table 22: Correct classification of SUB and SUB/IL mixture samples (Florida Fire Marshall Data distribution)

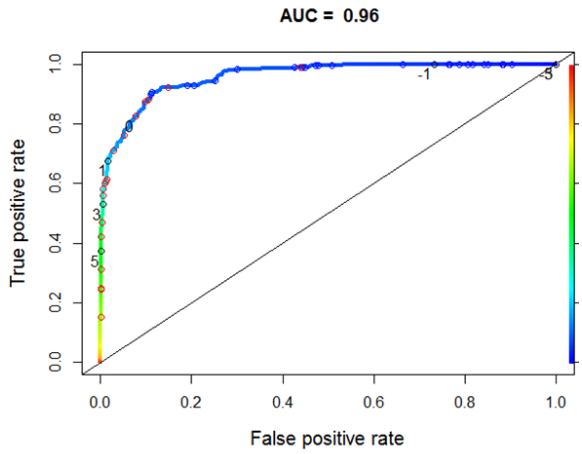
	TP (28)	**TN (26)**
All compounds	17 ± 1.9	25 ± 0.94
Compounds in IL	19 ± 1.8	25 ± 1.2
Compounds in IL and SUB	19 ± 2.1	25 ± 0.99

The TP is the correctly classified number of samples positive for ILR out of 28 fire debris samples with ILR and whereas TN is the correctly classified number of samples without ILR. The ROC curves presented are the projections of the log likelihood ratios of the final iteration (10) of the projection of data.

(a)



(b)



(c)

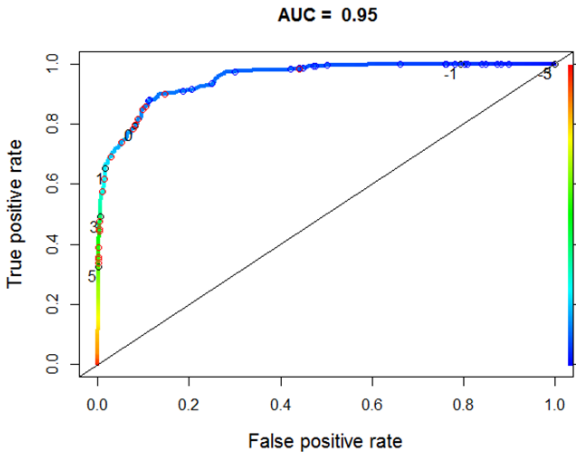


Figure 68: ROC plots obtained for the calculated LLRs for pure SUB and IL by Florida Fire Marshall data distribution frequency adjustments using a) all compounds b) compounds in IL c) compounds in both SUB and IL only.

The calculation of AUC using Mann-Whitney U statistic indicates that the highest AUC for the calculated LLRs were obtained using the compounds in ignitable liquids. The generated ROC plot for these LLRs for 54 samples is depicted in Figure 69.

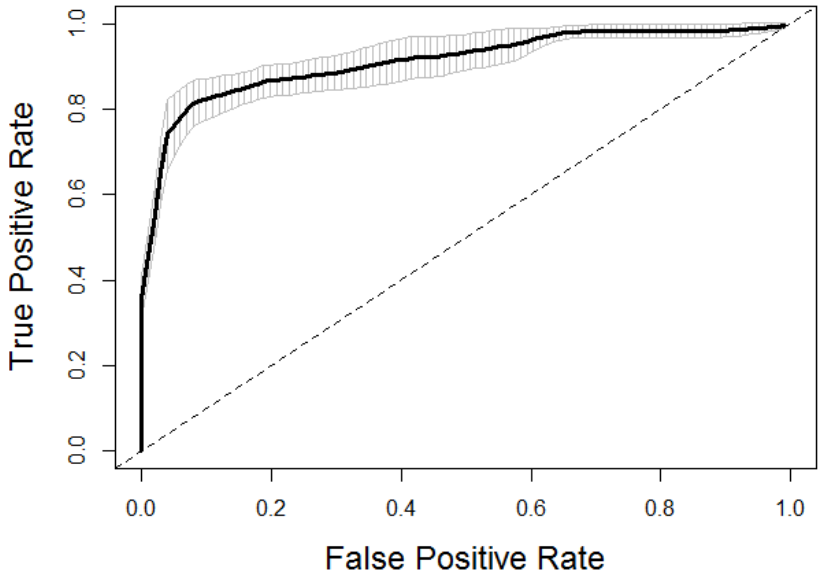


Figure 69: ROC curve generated from the calculated LLRs for 54 samples using compounds in IL (with 95% Confidence interval)

The reason for higher AUC for this distribution was the samples selection. In this, most of the selected samples with IL contained gasoline. As previously discussed, GAS containing samples have a correct identification of more than 60%.

6.3.3.4. Frequency Adjustments based on SUB and IL Databases (NCFS) Data Distribution

The AUC values obtained for the LLRs for the three methods (all compounds, compounds in IL and compounds in IL and SUB only) for pure SUB and IL using this set of frequencies were 0.99, 0.98, and 0.97 respectively. The calculated LLRs of 101 fire debris samples were projected on these ROC plots and the 10^{th} iteration of the ROC plots are presented in Figure 68a, 68b and 68c. The calculated average AUC using Mann-Whitney-U-statistic for the LLRs of the 101 fire debris samples calculated using the above three methods were,

All compounds – 0.80 ± 0.031

Compounds in IL – 0.89 ± 0.024

Compounds in SUB and IL only – 0.85 ± 0.029

When these calculated AUC values for pure SUB and IL were compared to the previous ROC curves obtained by the equal distribution adjusted frequencies, the AUC values were nearly the same. The average number of TP and TN are tabulated in Table 22.

Table 23: Correct classification of SUB and SUB/IL mixture samples (SUB and IL Database distribution)

	TP (56)	**TN (45)**
All compounds	29 ± 2.8	44 ± 0.92
Compounds in IL	37 ± 2.4	43 ± 1.1
Compounds in IL and SUB	33 ± 1.9	44 ± 0.99

Overall, the adjusted frequency sets using FL Fire Marshall and Database distributions performed well in LLR calculations in the fire debris samples.

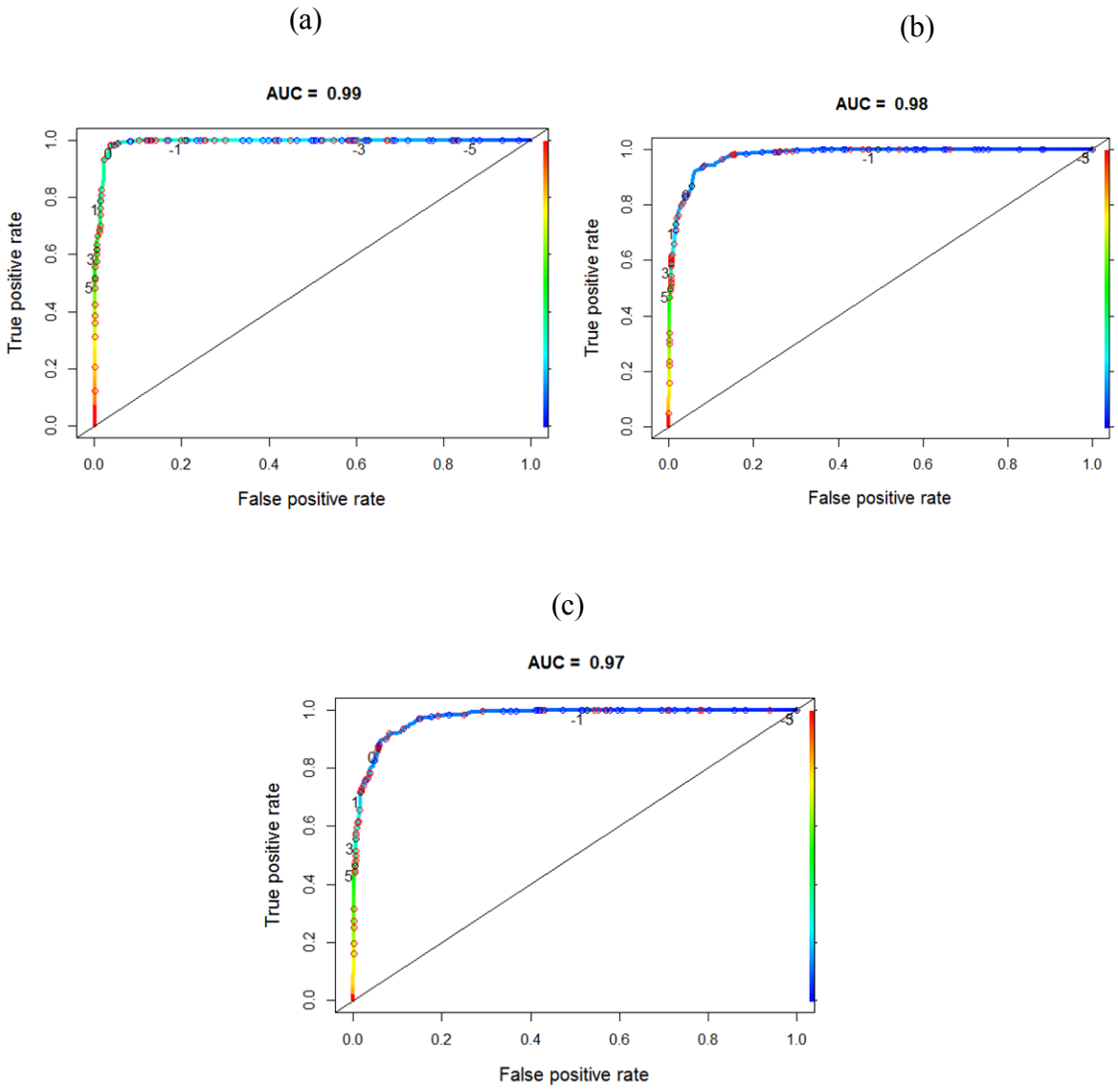


Figure 70: ROC plots obtained for the calculated LLRs for pure SUB and IL by SUB and IL database data (at NCFS) distribution frequency adjustments using a) all compounds b) compounds in IL c) compounds in both SUB and IL only.

As same as the previously discussed distributions, AUC was calculated using Mann-Whitney U Statistic. The calculated AUC for these LLRs was 0.89. The generated ROC lot for the calculated LLRs for 101 samples is provided in Figure 71.

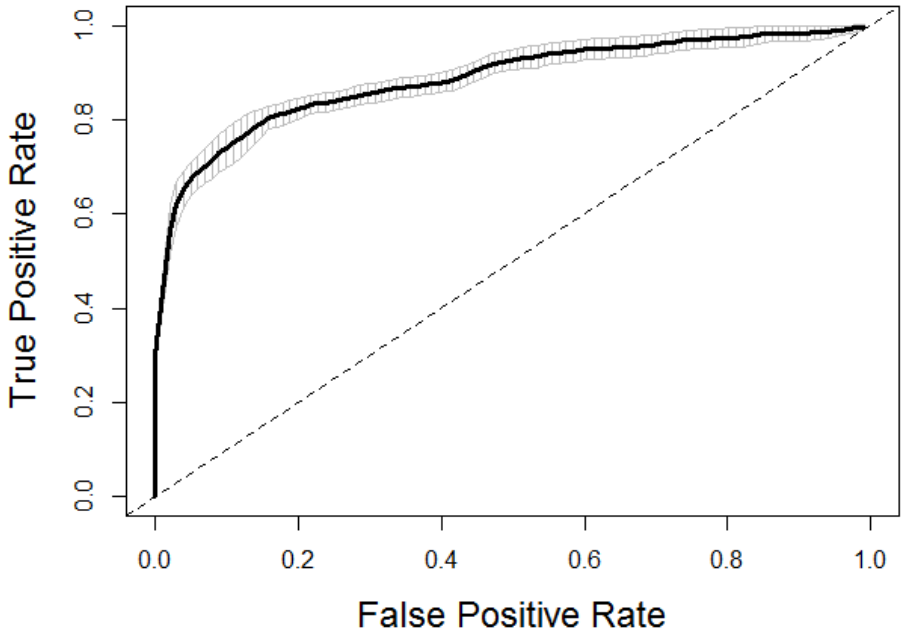


Figure 71: ROC curve generated from the calculated LLRs for 101 samples using compounds in IL (with 95% Confidence interval)

6.4. Likelihood Ratio Calibration using Logistic Regression

The likelihood ratios for the laboratory generated fire debris samples were calculated using the equal data distributions as discussed above. As above, 128 samples were randomly selected from the fire debris data (64 substrates, and 8 mixtures from each class). These likelihood ratios were calibrated using logistic regression as discussed in Section 4.2.2. One of the important observations in the calibration of LLRs was that inclusion of prior odds. But in this case, $\log_{10}$ (prior odds) is equal to zero since prior odds = 1.

6.4.1. Calibration of Likelihood Ratios Calculated using All Compounds

In this method, the AUC obtained for the calculated likelihood ratios was 0.74. The calibrated and uncalibrated ROC plots are presented in Figure 69a and 69b.

(a) (b)

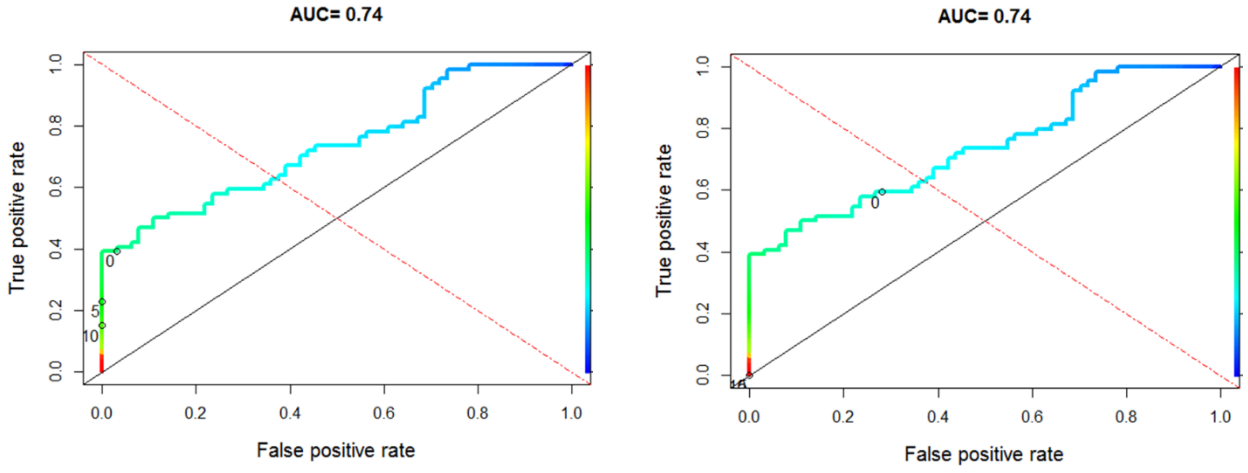


Figure 72: ROC plots obtained for the likelihood ratios calculated for known ground truth fire debris samples a) uncalibrated b) calibrated

The position of LLR=0 in the uncalibrated ROC plot is biased towards the positive LLRs whereas in the calibrated ROC plot the bisecting line on the plot intersects closer to LLR=0. The calibration of the calculated LLRs can be interpreted by the ECE plots. These plots are interpreted in Figure 70a and 70b.

(a) (b)

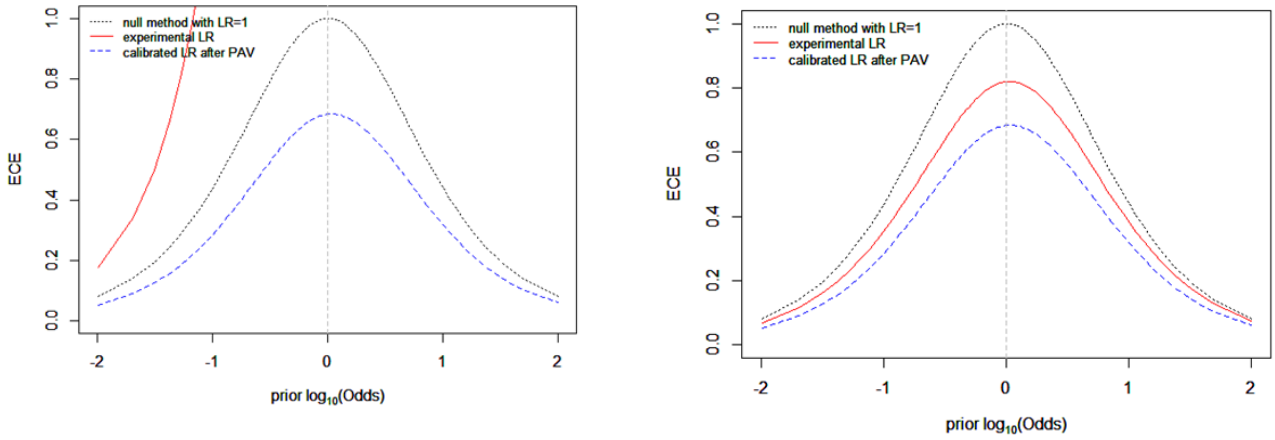


Figure 73: The ECE plots obtained for the calculated log-likelihood ratios a) Uncalibrated b) Calibrated

The calibration and the accuracy of the likelihood ratios were improved after the calibration by logistic regression (Figure 70b). The distributions of the uncalibrated and calibrated likelihood ratios are depicted by histograms in Figure 71a and 71b. According to these histograms, uncalibrated log likelihood ratios for substrates and mixtures were distributed from 0 to -20 and, -12 to 15 respectively. However, the calibrated log likelihood ratios were distributed from -2 to 0.5 in substrates and -0.5 to 3 in fire debris samples with IL residue.

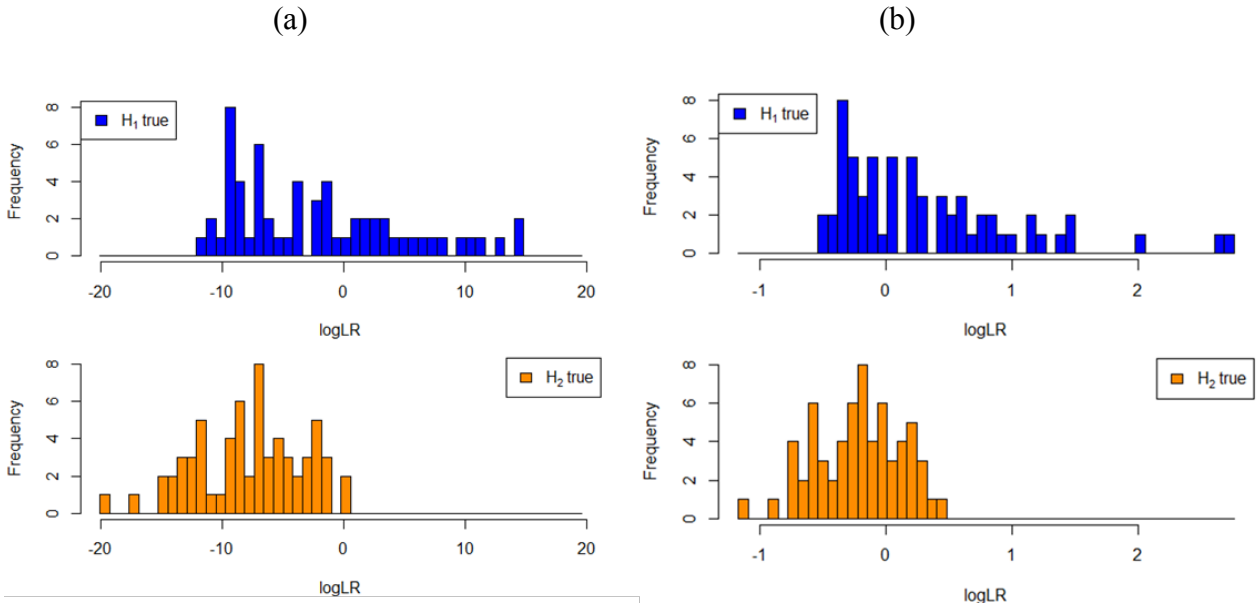


Figure 74: The distribution of log likelihood ratios a) uncalibrated b) calibrated

The discrimination power of the calculated log-likelihood ratios can be explained using tippet plots which are provided in Figure 72a and 72b.

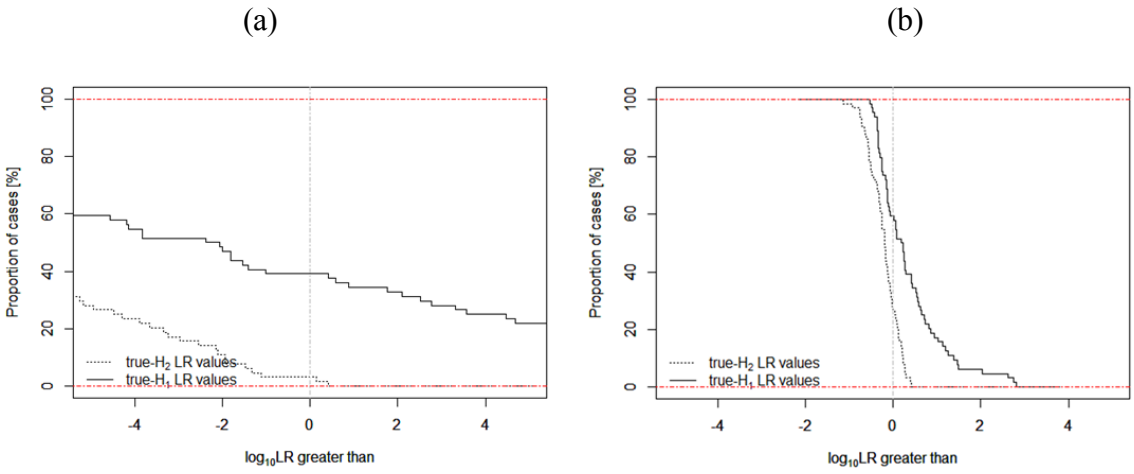


Figure 75: Tippet plots obtained for calculated log-likelihood ratios a) uncalibrated b) calibrated

In Figure 72a, the proportion of misleading evidence where true H_2 (calculated LLRs support the absence of IL) is less than 5 %. The proportion of misleading evidence for true H_1 (calculated LLRs support the presence of IL) is 60%. The proportion of misleading evidence increased for H_2 and decreased for H_1 after these LLRs were calibrated, but the discriminating power of these calculated log-likelihood ratios was not improved.

6.4.2. Calibration of the Likelihood Ratios Calculated using Compounds in IL

The calculated likelihood ratios were calibrated using logistic regression as discussed above. The ROC, ECE, tippet and histograms before calibration are provided in Figure 73a, 73b, 74a and 74b.

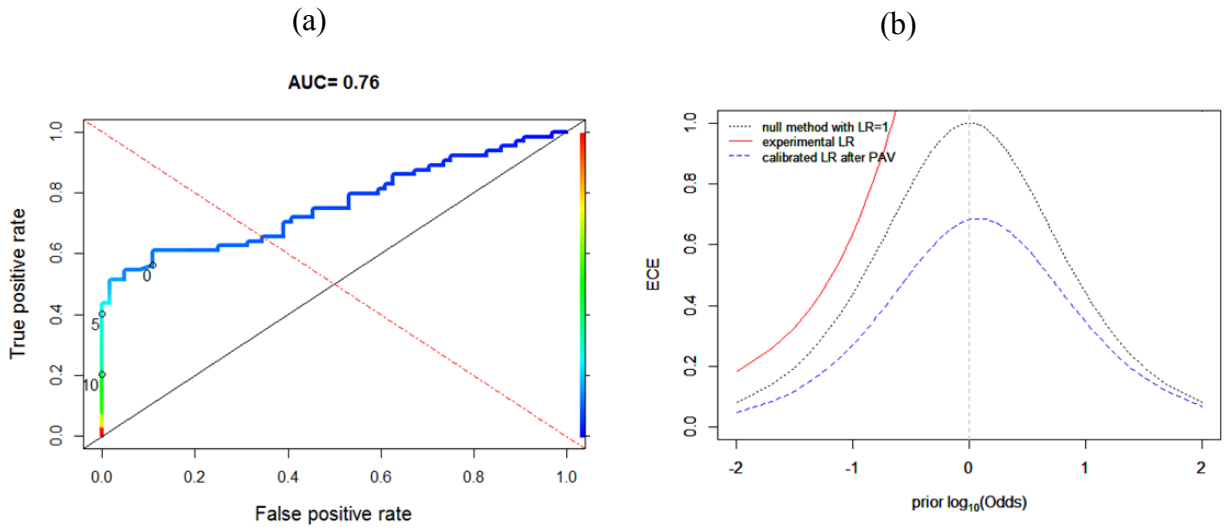


Figure 76: a) ROC plot b) ECE plot obtained for the log-likelihood ratios before calibration

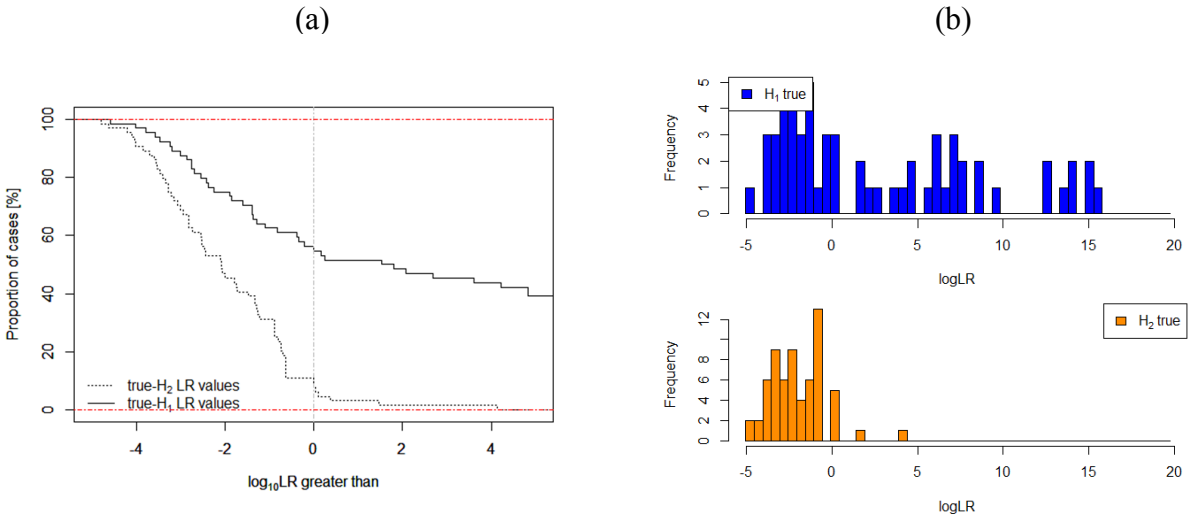


Figure 77: a) Tippet plot b) histogram obtained for the log-likelihood ratios before calibration

The AUC obtained for the ROC curve from the calculated log likelihood ratios was 0.76, and according to the ECE plot, the calibration and the accuracy of the calculated log-likelihood ratios were significantly decreased. The tippet plot also indicates that the proportion of misleading evidence which supports the proposition for the presence of ILR was approximately 50% and the proportion of misleading evidence which supports the proposition for the absence of ILR was approximately less than 10%. This also can be visualized in the log-likelihood ratio distribution in the histograms. The ROC, ECE, tippet and histograms obtained after calibrated are presented in Figure 75a, 75b, 75c and 75d respectively.

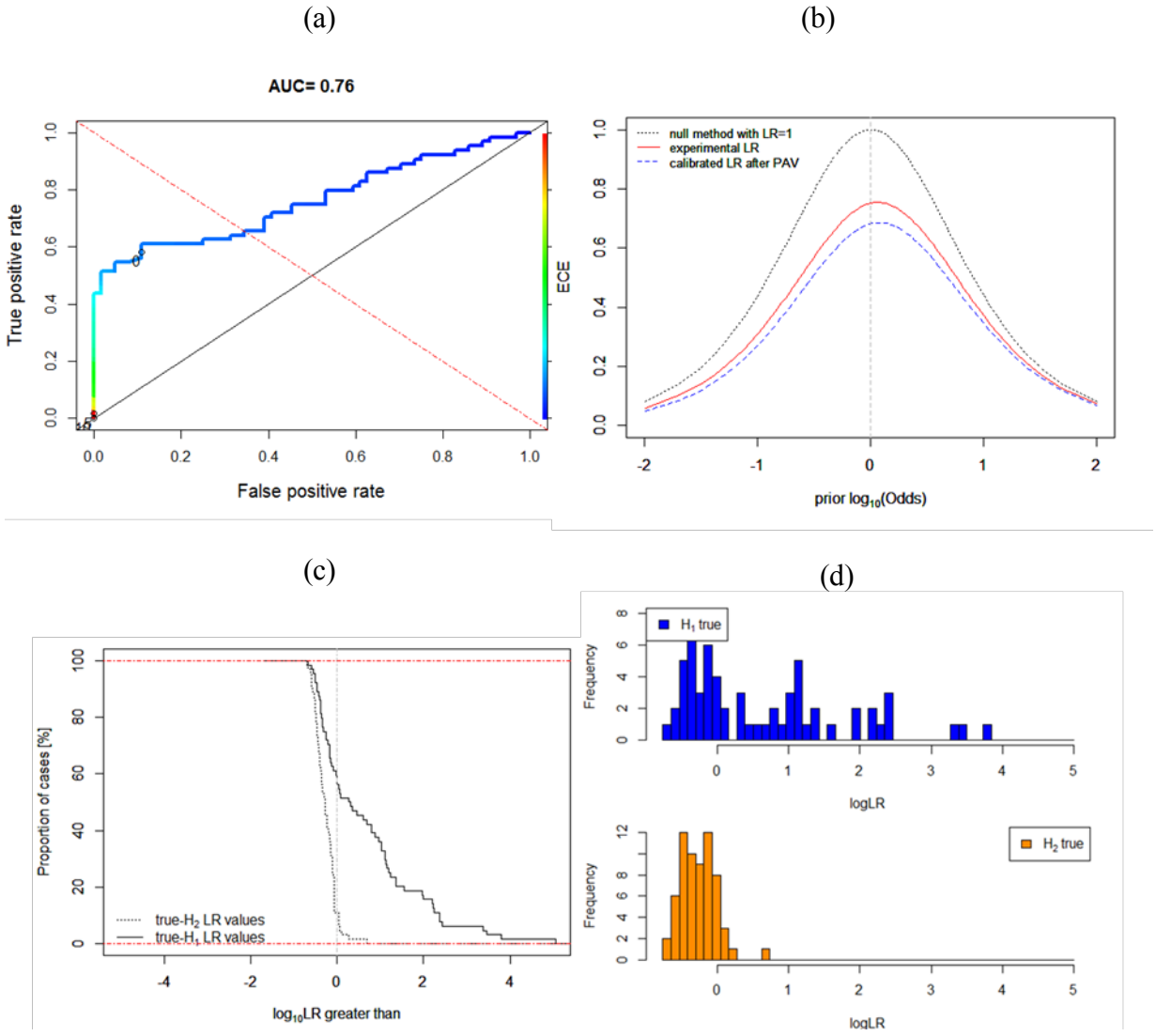


Figure 78: a) ROC b) ECE c) tippet and d) histogram of the calibrated likelihood ratios calculated by compounds in IL

The position of LLR = 0 did not change in the ROC plot after calibration of the log-likelihood ratios. But, this has not affected the calibration and the accuracy according to the ECE plot. The calibrated log likelihood ratios have better accuracy than the uncalibrated LLRs. The proportion

of the misleading evidence in both propositions did not change drastically. Nevertheless, the discriminating power did not improve after calibration.

The validation results of these methods show that the Naïve Bayes likelihood ratios can be used as a technique to classify fire debris data. However as discussed earlier, these methods do not perform well with AR, OXY, ISO and NA ignitable liquids used in these samples since they provide a lower evidentiary value in the log likelihood ratio calculation.

CHAPTER 7: CONCLUSION AND FUTURE WORK

As discussed in the previous chapters, statistical approaches can be used in the field of fire debris analysis to provide a numeric strength of evidence which will help the analysts in multiple ways. This chapter will conclude the results and will discuss the future work that can be performed using the statistical approaches discussed in this chapter.

7.1. Conclusion

One of the sections of this study was focused on the identification of major compounds in burned substrates (SUB) and ignitable liquids (IL) and to determine a probability of the presence of these compounds in IL and SUB using a logistic regression model. The logistic regression analysis of these compounds showed that there were 177 compounds can be seen in both ignitable liquids and substrates, 39 compounds only in ignitable liquids and 77 compounds only in substrates. These compounds were determined by specific probability cutoffs for IL and SUB based on these compounds' retention time similarity. The number of these compounds may subject to change with the change of probability cutoffs. For example, if the probability cutoff was lowered, then more compounds would have been identified but this also would increase the risk of false positive identification. However, the method used to determine the probability cutoffs in this work reduced the possibility of false positives identification which can be a major advantage in the fire debris analysis, even though this method did not identify all the true positives. **Absence of certain compounds in the sample of these ignitable liquids or substrates examined in this work does not mean that they would not be identified in the general population**.

The other important part of this work is the use of Naïve Bayes likelihood ratios to classify the fire debris samples. This classification worked well in the identification of pure substrates and ignitable liquid samples. However, when this method was applied to the substrates and ignitable liquid mixtures it did not work well. According to the Table 4 in section 6.1, the correct percentage of LPD, MISC, MPD, NP and GAS samples were more than 60% and except gasoline all the other classes have more than 75% correct classification.

The main reason for the misclassification of the fire debris samples is that some of the compounds produced in burned substrates have a higher frequency of occurrence in substrates than in ignitable liquids. The 10 ignitable liquid samples used in this technique were 50 % - 75 % weathered, therefore a significant number of peaks were reduced in these ignitable liquids which in turn could affect the magnitude of the likelihood ratios. The results showed that this method worked well with the compounds identified in ignitable liquids.

At present, fire debris analysis is mainly based on pattern identification from the total ion chromatograms (TIC) and extracted ion profiles (EIP). If there is no specific pattern identified in the samples or there are no target compounds identified, it is more likely to be classified as negative for the presence of ignitable liquid residue (ILR). This may or may not be true. The importance of this method is that even when there is no significant chromatographic pattern, if the peaks (major or minor) are identified in the TIC, it will provide a likelihood ratio. This method is independent of the abundance of the identified peaks which can be considered as a major advantage.

7.2. Future Work

Identification of more compounds in ignitable liquids and compounds only identified in each ASTM E1618-14[5] class will significantly improve this method. This project only classifies the fire debris samples as positive or negative for ILR. However, the identification of more compounds unique to each class can be used to determine the specific IL class by the Naïve Bayes likelihood approach. Proposed future work based on this project is the expansion of the ILRC database to identify more compounds unique to ignitable liquids and the usage of these compound sets in the IL class determination.

APPENDIX A: DATA TABLES REQUIRED FOR CHAPTER 5

Table 24: Compounds and compound type

Compound	Compound Type
Methanol	alcohol
Acetaldehyde	aldehyde
Ethanol	alcohol
Acetonitrile	nitrile
2-methyl butane	branched alkane
Acetone	aldehyde
Acrolein	aldehyde
Propanal	aldehyde
Isopropanol	alcohol
Diethyl ether	ether
n-Pentane	Alkane
Methylene chloride	alkyl halide
Nitromethane	nitro alkane
2,2-dimethyl butane	branched alkane
2-methyl pentane	branched alkane
Butanal	aldehyde
2-butanone	ketone
3-methylpentane	branched alkane
2-methylfuran	furan
Ethyl acetate	ester
n-hexane	Alkane
2-methyl-1-propanol	alcohol
Tetrahydrofuran	furan
1,2-dichloroethane	alkyl halide
2,4-dimethyl pentane	branched alkane
methylcyclopentane	cycloalkane
2-methyl-butanal	aldehyde
1-butanol	alcohol
Benzene	aromatic hydrocarbon
1-methoxy-2-propanol	alcohol
cyclohexane	cycloalkane
2-methyl hexane	branched alkane
2,3-dimethyl-pentane	branched alkane
Pentanal	aldehyde
3-methylhexane	branched alkane
2,2,4-trimethyl pentane	branched alkane
methyl methacrylate	ester
n-heptane	alkane

Compound	Compound Type
2,4,4-trimethyl pentane	branched alkane
1-methylpyrrole	pyrrole
Propionic acid	carboxylic acid
methyl isobutyl ketone	ketone
methylcyclohexane	cycloalkane
2,4,4-trimethyl-2-pentene	branched alkane
2,5-dimethylhexane	branched alkane
ethylcyclopentane	cycloalkane
2,4-dimethyl hexane	branched alkane
dimethylformamide	amide
2-methylpentanal	aldehyde
1-Pentanol	alcohol
2,3,4-trimethylpentane	branched alkane
toluene	aromatic hydrocarbon
2,3,3-trimethylpentane	branched alkane
Cyclopentanone	ketone
2-methylheptane	branched alkane
3,5,5-trimethyl-1-hexane	branched alkane
Hexanal	aldehyde
3-methylheptane	branched alkane
(cis)-1,3-dimethyl cyclohexane	cycloalkane
2-ethyl-1-hexene	branched alkene
2,2,5-trimethylhexane	branched alkane
2-cyclopenten-1-one	ketone
(cis)-3-methyl-2-heptene	branched alkene
furfural	aldehyde
Butyl acetate	ester
n-octane	normal alkane
(trans)-3-methyl-2-heptene	branched alkene
2,3,5-trimethylhexane	branched alkane
hexamethylcyclotrisiloxane	siloxane
4-vinyl-1-cyclohexene	cyclohexene
Furfuryl alcohol	alcohol
1,3,5-trimethylcyclohexane	cycloalkane
ethylcyclohexane	cycloalkane
1,1,3-trimethylcyclohexane	cycloalkane
2,4-dimethyl-1-heptene	branched alkene
1-Methoxy-2-propylacetate	ester
ethylbenzene	aromatic hydrocarbon

Compound	Compound Type
1,2,4-trimethylcyclohexane	cycloalkane
m-xylene	aromatic hydrocarbon
Cyclohexanone	ketone
p-xylene	aromatic hydrocarbon
1,3-dichloro-2-propanol	halagenated alcohol
2-methyl octane	branched alkeane
2-heptanone	ketone
Styrene	aromatic hydrocarbon
Heptanal	aldehyde
o-xylene	aromatic hydrocarbon
2-butoxy ethanol	alcohol
(cis)-1-ethyl-4-methylcyclohexane	cycloalkane
1-nonene	alkene
Methoxy-3-methylbutanol	alcohol
4-Nonene	alkene
Isobutyl isobutyrate	ester
n-nonane	normal alkane
(trans)-1-ethyl-4-methylcyclohexane	cycloalkane
Isopropyl benzene	aromatic hydrocarbon
1-butoxy-2-propanol	alcohol
5-Methylfurfural	aldehyde
Benzaldehyde	aromatic aldehyde
2,2-oxybis ethanol	alcohol
propylcyclohexane	cycloalkane
2-(2-chloroethoxy) ethanol	alcohol
alpha pinene	terpene
2,6-dimethyl octane	branched alkane
o-chlorotoluene	halagonated aromatic hydrocarbon
n-propylbenzene	aromatic hydrocarbon
Benzonitrile	aromatic nitrile
bis(2-chloroethyl) ether	ether
m-ethyltoluene	aromatic hydrocarbon
p-ethyltoluene	aromatic hydrocarbon
Phenol	phenol
1,3,5-trimethylbenzene	aromatic hydrocarbon
4-methylnonane	branched alkane
Ethyl-3-ethoxy-propionate	ester
Methacrolein	aldehyde
2-methyl nonane	branched alkane

Compound	Compound Type
2,2,6-trimethyl octane	branched alkane
alpha methylstyrene	aromatic hydrocarbon
o-ethyltoluene	aromatic hydrocarbon
3-methylnonane	branched alkane
beta pinene	terpene
Diethylene glycol monoethyl ether	ether
Di(propylene glycol) methyl ether mixture of isomers-A	ether
Octanal	aldehyde
Di(propylene glycol) methyl ether mixture of isomers-C	ether
2-pentyl furan	furan
1,2,4-trimethylbenzene	aromatic hydrocarbon
Benzyl chloride	aryl halide
1-decene	alkene
Di(propylene glycol) methyl ether mixture of isomers-B	ether
isododecane	branched alkane
n-Hexyl acetate	ester
3-chloromethylheptane	halogenated branched alkane
n-decane	alkane
Octamethylcyclotetrasiloxane	siloxane
1-methyl-2-Pyrrolidinone	aromatic ketone
benzyl alcohol	aromatic alcohol
3-carene	terpene
1,2,3-trimethylbenzene	aromatic hydrocarbon
alpha terpinene	terpene
2-ethyl-1-hexanol	alcohol
indane	indane
4-methyldecane	branched alkane
2-methylphenol	phenol
Limonene	terpene
indene	indene
2,2,8-trimethyl-decane	branched alkane
Acetophenone	aromatic ketone
p-Cresol	phenol
1-chloro octane	alkyl halide
(trans)Decahydronaphthalene	cycloalkane
2-methoxyphenol	phenol
2-butoxyethylacetate	ester
p-alpha dimethylstyrene	aromatic hydrocarbon
3-methyldecane	branched alkane

Compound	Compound Type
4-ethyl-1,2-dimethylbenzene	aromatic hydrocarbon
2,4,6-trimethyl-1-nonene (meso form)(D)	branched alkene
2,4,6-trimethyl-1-nonene (racemic form)(E)	branched alkene
3-methyl-5-propylnonane	branched alkane
Nonanal	aldehyde
alpha terpinolene	terpene
Phenylethyl alcohol	alcohol
1-undecene	alkene
n-Heptyl acetate	ester
n-undecane	alkane
dimethyl glutarate	ester
(cis)-Decahydronaphthalene	cycloalkane
1,2,4,5-tetramethylbenzene	aromatic hydrocarbon
1,2,3,5-tetramethylbenzene	aromatic hydrocarbon
vinyl benzoate	ester
decahydro-2-methylnaphthalene	cycloalkane
DL-Camphor	ketone
Benzyl acetate	ester
Di(propylene glycol)methyl ether acetate mixture of isomers-A	ether
Citronellal	aldehyde
1_3-bis(1-methylethyl)benzene	aromatic hydrocarbon
Di(propylene glycol)methyl ether acetate mixture of isomers-B	ether
Benzoic acid	Aromtatic carboxylic acid
Decamethylcyclopentasiloxane	siloxane
2-methyl undecane	branched alkane
creosol	phenol
methyl salicylate	aromatic ester
naphthalene	PAH
alpha terpineol	terpene
1-methoxy-4-(2-propenyl)benzene	ether
2-phenoxyethanol	aromatic alcohol
Decanal	aldehyde
1-dodecene	alkene
benzothiazole	thiozole
n-Octyl acetate	ester
n-dodecane	alkane
Caprolactam	ketone
Dimethyl ester hexanedioic acid	ester
1-phenoxypropan-2-ol	aromatic alcohol

Compound	Compound Type
2,6-dimethyl undecane	branched alkane
p-anisaldehyde	aromatic aldehyde
allyl benzoate	ester
4-Ethoxyphenol	phenol
1-decanol	alcohol
methyl benzoylformate	aromatic ester
4-Ethyl-2-methoxyphenol	phenol
p-tert butylphenol	phenol
Tri(propylene glycol)methyl ether mixture of isomers-A	ether
Phthalic_acid_anhydride	anhydride
Tri(propylene glycol)methyl ether mixture of isomers-B	ether
Safrole	phenyl propene
3-tert-butylphenol	phenol
Tri(propylene glycol)methyl ether mixture of isomers-C	ether
Isobornyl acetate	ester
Tri(propylene glycol)methyl ether mixture of isomers-D	ether
2-methyl naphthalene	PAH
Tri(propylene glycol)methyl ether mixture of isomers-E	ether
1-Tridecene	alkene
Undecanal	aldehyde
4-phenylbutronitrile	aromatic nitrile
Azulene	PAH
Triacetin	ester
alpha-methyl-trans-cinnamaldehyde	aromatic aldehyde
n-tridecane	alkane
1-methylnaphthalene	PAH
2,4,6,8-tetramethyl-1-undecene (isotactic)(A)	branched alkene
2,4,6,8-tetramethyl-1-undecene (heterotactic)(B)	branched alkene
2_6-dimethoxy_phenol	phenol
2,4,6,8-tetramethyl-1-undecene (syndiotactic)(C)	branched alkene
2-methyl propanal	alcohol
eugenol	aromatic alcohol
Ethanol-2-(2-butoxyethoxy)-acetate	ester
1,3,5-tris-(1-methylethyl) benzene	aromatic hydrocarbon
1,2-bis(2-aminophenoxy)ethane-N,N,N,N-tetraacetic acid	aromatic carboxylic acid
Cyclododecane	cycloalkane
Butyl benzoate	aromatic ester
1-(2,4,5-trimethylphenyl)ethanone	ketone
Biphenyl	aromatic hydrocarbon

Compound	Compound Type
2-Chloroethyl benzoate	aryl halide ester
1-Tetradecene	alkene
Dodecanal	aldehyde
Decyl acetate	ester
1,5-dimethylnaphthalene	PAH
n-tetradecane	normal alkane
3,5-Dimethoxy-4-hydroxytoluene	phenol
2,3-dimethyl-naphthalene	PAH
3-hydroxy-4-methoxy benzaldehyde	aromatic aldehyde
Ethyl vanillin	aromatic aldehyde
Longifolene	cycloalkane
Caryophyllene	bicyclic sesquiterpene
1-Dodecanol	alcohol
Acenaphthene	PAH
1-Pentadecene	alkene
Tridecanal	aldehyde
Butylated_hydroxytoluene	phenol
n-pentadecane	alkane
Lilial	aromatic aldehyde
Bibenzyl	aromatic hydrocarbon
1,2-diphenylpropane	branched alkane
Diethyltoluamide	aromatic amide
Diethyl phthalate	aromatic ester
fluorene	PAH
TXIB	ester
Tetradecanal	aldehyde
1-hexadecene	alkene
n-hexadecane	alkane
Benzophenone	aromatic aldehyde
Cedrol	sesquiterpene alcohol
1,3-diphenylpropane	aromatic hydrocarbon
1-Heptadecene	alkene
Benzoic acid, 2-ethylhexyl ester	aromatic ester
n-heptadecane	alkane
3-butene-1,3-diyldibenzene	aromatic hydrocarbon
pristane	alkane
1H-Indene-2,3-dihydro-1,1,3-trimethyl-3-phenyl	indene
Benzyl benzoate	aromatic ester
3,5-di-tert-Butyl-4-hydroxybenzaldehyde	aromatic aldehyde

Compound	Compound Type
Anthracene	PAH
1-Octadecene	alkene
n-octadecane	alkane
phytane	alkane
n-nonadecane	alkane
Methyl ester hexadecanoic acid	ester
Dibutyl phthalate	aromatic ester
Malathion	organo phosphate
n-eicosane	alkane
1-benzylbenzoimidazole	benzimidazoles
Heptadecanoic acid-16-methyl methyl ester	ester
9,12-Octadecadienoic acid(Z,Z)-methyl ester	ester
9-Octadecenoic acid(Z)methyl ester	ester
9-Octadecenoic acid(E)methylester	ester
n-heneicosane	alkane
methyl ester octadecanoic acid	ester
1-docosene	alkene
n-docosane	alkane
bis(2-ethylhexyl)adipate	ester
n-Tetracosane	alkane
n-Hexacosane	alkane

Table 25: Data obtained for compound type charts in Chapter 5

Compound Type	AR	GAS	HPD	ISO	LPD
alcohol	0.005	0.009	0.000	0.000	0.000
aldehyde	0.000	0.000	0.000	0.000	0.002
alkane	0.046	0.079	0.214	0.049	0.130
alkene	0.000	0.008	0.000	0.000	0.000
alkyl halide	0.000	0.000	0.000	0.000	0.000
amide	0.000	0.000	0.000	0.000	0.000
anhydride	0.000	0.000	0.000	0.000	0.000
aromatic alcohol	0.000	0.000	0.000	0.000	0.000
aromatic aldehyde	0.000	0.000	0.000	0.000	0.000
aromatic amide	0.000	0.000	0.000	0.000	0.000
aromatic carboxylic acid	0.000	0.000	0.000	0.000	0.000
aromatic ester	0.000	0.000	0.000	0.000	0.000
aromatic hydrocarbon	0.669	0.369	0.303	0.015	0.064
aromatic ketone	0.000	0.000	0.000	0.000	0.000
aromatic nitrile	0.003	0.000	0.000	0.000	0.000
aryl halide	0.000	0.000	0.000	0.000	0.000
aryl halide ester	0.000	0.000	0.000	0.000	0.000
benzimidazoles	0.000	0.000	0.000	0.000	0.000
bicyclic sesquiterpene	0.000	0.000	0.000	0.000	0.000
branched alkane	0.132	0.296	0.161	0.621	0.406
branched alkene	0.005	0.027	0.016	0.160	0.036
carboxylic acid	0.000	0.000	0.000	0.000	0.000
cycloalkane	0.010	0.104	0.180	0.044	0.355
cycloalkene	0.000	0.000	0.000	0.000	0.000
ester	0.015	0.000	0.000	0.000	0.000
ether	0.005	0.000	0.000	0.010	0.000
furan	0.000	0.000	0.000	0.000	0.000
halagonated alcohol	0.000	0.000	0.000	0.000	0.000
halagonated aromatic hydrocarbon	0.000	0.000	0.000	0.000	0.000
halogenated branched alkane	0.000	0.000	0.000	0.000	0.000
indane	0.003	0.007	0.024	0.092	0.007
indene	0.000	0.000	0.000	0.000	0.000
ketone	0.005	0.000	0.000	0.000	0.000
nitrile	0.000	0.000	0.000	0.000	0.000
nitro alkane	0.000	0.000	0.000	0.000	0.000
organo phosphate	0.003	0.000	0.000	0.000	0.000
PAH	0.099	0.101	0.101	0.000	0.000
phenol	0.000	0.000	0.000	0.000	0.000

Compound Type	AR	GAS	HPD	ISO	LPD
phenyl propanoids	0.000	0.000	0.000	0.000	0.000
phenyl propene	0.000	0.000	0.000	0.000	0.000
pyrrole	0.000	0.000	0.000	0.000	0.000
sesquiterpene alcohol	0.000	0.000	0.000	0.000	0.000
siloxane	0.000	0.000	0.000	0.000	0.000
terpene	0.000	0.000	0.001	0.010	0.000
thiozole	0.000	0.000	0.000	0.000	0.000

Compound Type	MISC	MPD	NA	NP	OXY
alcohol	0.008	0.001	0.000	0.000	0.064
aldehyde	0.005	0.000	0.000	0.000	0.031
alkane	0.145	0.157	0.776	0.167	0.108
alkene	0.004	0.008	0.000	0.004	0.002
alkyl halide	0.000	0.000	0.000	0.000	0.004
amide	0.000	0.000	0.000	0.000	0.000
anhydride	0.002	0.002	0.000	0.000	0.000
aromatic alcohol	0.001	0.000	0.000	0.000	0.002
aromatic aldehyde	0.000	0.000	0.000	0.000	0.016
aromatic amide	0.000	0.000	0.000	0.000	0.000
aromatic carboxylic acid	0.000	0.000	0.000	0.000	0.000
aromatic ester	0.005	0.000	0.000	0.000	0.002
aromatic hydrocarbon	0.275	0.256	0.066	0.162	0.268
aromatic ketone	0.000	0.000	0.000	0.000	0.000
aromatic nitrile	0.000	0.000	0.000	0.000	0.000
aryl halide	0.000	0.000	0.000	0.000	0.001
aryl halide ester	0.000	0.000	0.000	0.000	0.000
benzimidazoles	0.000	0.000	0.000	0.000	0.000
bicyclic sesquiterpene	0.000	0.000	0.000	0.000	0.002
branched alkane	0.233	0.195	0.079	0.282	0.166
branched alkene	0.021	0.032	0.013	0.000	0.011
carboxylic acid	0.000	0.000	0.000	0.000	0.000
cycloalkane	0.202	0.301	0.000	0.312	0.143
cycloalkene	0.000	0.000	0.000	0.000	0.000
ester	0.007	0.000	0.000	0.000	0.051
ether	0.007	0.001	0.000	0.000	0.032
furan	0.000	0.000	0.000	0.000	0.002
halagonated alcohol	0.000	0.000	0.000	0.000	0.000
halagonated aromatic hydrocarbon	0.001	0.000	0.000	0.000	0.000
halogenated branched alkane	0.000	0.000	0.000	0.000	0.000

Compound Type	MISC	MPD	NA	NP	OXY
indane	0.022	0.037	0.053	0.060	0.014
indene	0.000	0.000	0.000	0.000	0.000
ketone	0.003	0.000	0.000	0.000	0.012
nitrile	0.000	0.000	0.000	0.000	0.000
nitro alkane	0.000	0.000	0.000	0.000	0.004
organo phosphate	0.000	0.000	0.000	0.000	0.000
PAH	0.037	0.009	0.000	0.013	0.040
phenol	0.000	0.000	0.000	0.000	0.001
phenyl propanoids	0.000	0.000	0.000	0.000	0.000
phenyl propene	0.000	0.000	0.000	0.000	0.000
pyrrole	0.000	0.000	0.000	0.000	0.000
sesquiterpene alcohol	0.000	0.000	0.000	0.000	0.001
siloxane	0.000	0.000	0.000	0.000	0.000
terpene	0.025	0.001	0.013	0.000	0.025
thiozole	0.000	0.000	0.000	0.000	0.000

APPENDIX B: PUBLICATIONS AND COPY RIGHTS

Forensic Chemistry 5 (2017) 91–108

Contents lists available at ScienceDirect

Forensic Chemistry

journal homepage: www.elsevier.com/locate/forc

Major chemical compounds in the Ignitable Liquids Reference Collection and Substrate databases

Anuradha Akmeemana [a,b], Mary R. Williams [b,*], Michael E. Sigman [a,b]

[a] Department of Chemistry, University of Central Florida, Orlando, FL 32816-2366, United States
[b] National Center for Forensic Science, University of Central Florida, P.O. Box 162367, Orlando, FL 32816-2367, United States

ARTICLE INFO

Article history:
Received 28 April 2017
Received in revised form 17 July 2017
Accepted 22 July 2017
Available online 23 July 2017

Keywords:
Fire debris
Ignitable liquids
Substrates
And matrix

ABSTRACT

Major peaks identified in samples within the Ignitable Liquids Reference Collection (ILRC) and Substrate databases were tabulated to ascertain information on the chemical compound types present in ignitable liquids and pyrolytic decomposition products from substrate materials. There were 221 major compounds identified in the ignitable liquid and substrate records in the databases. Thirty-six of these major compounds were identified in both. The major compounds were grouped by compound type to provide a general characterization of ignitable liquids and substrates within the databases. ASTM E1618-14 designated classes were characterized by how often particular types of compounds were identified within each class. Identification of major compounds and how often they occur within a specified ASTM E1618-14 class and as substrate decomposition products may provide useful chemical characterization information.

Forensic Chemistry 7 (2018) 38–46

Contents lists available at ScienceDirect

Forensic Chemistry

journal homepage: www.elsevier.com/locate/forc

Model-effects on likelihood ratios for fire debris analysis

Richard Coulson [a,b], Mary R. Williams [b], Alyssa Allen [a,b], Anuradha Akmeemana [a,b], Liqiang Ni [a,c], Michael E. Sigman [a,b,*]

[a] National Center for Forensic Science, University of Central Florida, P.O. Box 162367, Orlando, FL 32816, United States
[b] Department of Chemistry, University of Central Florida, P.O. Box 162367, Orlando, FL 32826, United States
[c] Department of Statistics, University of Central Florida, P.O. Box 162367, Orlando, FL 32826, United States

ARTICLE INFO

Article history:
Received 4 August 2017
Received in revised form 18 December 2017
Accepted 18 December 2017
Available online 19 December 2017

Keywords:
Fire debris
Likelihood ratio
Principal components
Population distribution
Evidentiary value

ABSTRACT

A simple method is introduced for assessing the evidentiary value of fire debris samples. The method relies on models built by random draws from a database of ignitable liquid and substrate pyrolysis samples. A stratified random draw from database records belonging to each ASTM E1618 ignitable liquid class and the substrate class is taken in proportion to a prescribed distribution. The likelihood ratios are estimated by direct calculation from a one-level Gaussian kernel density model based on the covariance structure and multivariate means resulting from each random draw. Multiple draws result in model-related variation in the calculated likelihood ratios. The method is demonstrated for 500 random draws with replacement based on three distributions. Ten-fold cross-validation is performed for each model, with subsequent testing on laboratory burn samples. The average cross-validated receiver operating characteristic areas under the curve were 0.959, 0.956 and 0.947, for the three distributions. The mean log (base 10) of the likelihood ratio (LLR) values for the laboratory test burns of gasoline on polyester carpet and carpet padding were in the range of 1.5 to −3.1 and were qualitatively consistent with the observed gasoline contributions to the total ion chromatograms for the samples. The LLR values for laboratory test burns of a petroleum distillate on polyester carpet and carpet padding, 1.7 to −3.3, were also in qualitative agreement with the observed chromatographic patterns. The calculated LLR values for laboratory burns of a set of polyester carpet and padding samples were in the range −0.9 to −6.7.

REFERENCES

1. Stauffer, E.; Dolan, J. A.; Newman, R., CHAPTER 1 - Introduction. In *Fire Debris Analysis*, Academic Press: Burlington, 2008; pp 1-17.
2. Keto, R. O.; Wineman, P. L., Detection of petroleum-based accelerants in fire debris by target compound gas chromatography/mass spectrometry. *Analytical Chemistry* **1991,** *63* (18), 1964-1971.
3. Keto, R. O., GC/MS Data Interpretation for Petroleum Distillate Identification in Contaminated Arson Debris. *J. Forensic. Sci. 40*.
4. Cerdan-Calero, M.; Sendra, J. M.; Sentandreu, E., Gas chromatography coupled to mass spectrometry analysis of volatiles, sugars, organic acids and aminoacids in Valencia Late orange juice and reliability of the Automated Mass Spectral Deconvolution and Identification System for their automatic identification and quantification. *Journal of chromatography. A* **2012,** *1241*, 84-95.
5. ASTM Standard E1618-14, "Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography-Mass Spectrometry", ASTM International, West Conshohocken, PA, 2014.
6. National Center for Forensic Science, Substrate Database, 2010, http://ilrc.ucf.edu/substrate/

7. Castelbuono, Joseph, "The Identification Of Ignitable Liquids In The Presence Of Pyrolysis Products: Generation Of A Pyrolysis Product Database" (2008). Electronic Theses and Dissertations. 3626.
8. National Center for Forensic Science, Ignitable Liquids Reference Collection Database, 2010, http://ilrc.ucf.edu/index.php

9. Brown, S. D., Chemometrics: A textbook. D. L. Massart. B. G. M. Vandeginste, S. N. Deming, Y. Michotte, and L. Kaufman, Elsevier, Amsterdam, 1988. ISBN 0-444-42660-4. Price Dfl 175.00. *Journal of Chemometrics* **1988,** *2* (4), 298-299.
10. Tan, B.; Hardy, J. K.; Snavely, R. E., Accelerant classification by gas chromatography/mass spectrometry and multivariate pattern recognition. *Analytica Chimica Acta* **2000,** *422* (1), 37-46.
11. Sinkov, N. A.; Sandercock, P. M. L.; Harynuk, J. J., Chemometric classification of casework arson samples based on gasoline content. *Forensic science international* **2014,** *235*, 24-31.
12. Doble, P.; Sandercock, M.; Du Pasquier, E.; Petocz, P.; Roux, C.; Dawson, M., Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks. *Forensic science international* **2003,** *132* (1), 26-39.
13. Sandercock, P. M. L.; Du Pasquier, E., Chemical fingerprinting of unevaporated automotive gasoline samples. *Forensic science international* **2003,** *134* (1), 1-10.
14. Waddell, E. E.; Williams, M. R.; Sigman, M. E., Progress toward the determination of correct classification rates in fire debris analysis II: utilizing soft independent modeling of class analogy (SIMCA). *Journal of forensic sciences* **2014,** *59* (4), 927-35.

15. Sigman, M. E.; Williams, M. R., Assessing evidentiary value in fire debris analysis by chemometric and likelihood ratio approaches. *Forensic science international* **2016,** *264*, 113-121.
16. Coulson, R.; Williams, M. R.; Allen, A.; Akmeemana, A.; Ni, L.; Sigman, M. E., Model-effects on likelihood ratios for fire debris analysis. *Forensic Chemistry* **2018,** *7*, 38-46.
17. Akmeemana, A.; Williams, M. R.; Sigman, M. E., Major chemical compounds in the Ignitable Liquids Reference Collection and Substrate databases. *Forensic Chemistry* **2017,** *5*, 91-108.
18. Thea, V.; Marie, C.; J., P. T., Verbal overshadowing: a sound theory in voice recognition? *Applied Cognitive Psychology* **2005,** *19* (9), 1127-1144.
19. Duray, S. M.; Morter, H. B.; Smith, F. J., Morphological variation in cervical spinous processes: potential applications in the forensic identification of race from the skeleton. *Journal of Forensic Science* **1999,** *44* (5), 937-944.
20. Delgado, R.; González, J.-L.; Sotoca, A.; Tibau, X.-A., Archetypes of Wildfire Arsonists: An Approach by Using Bayesian Networks. **2018**.
21. Stein, S. E., An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry* **1999,** *10* (8), 770-781.
22. Gale, W. A.; Sampson, G., Good-turing frequency estimation without tears. *Journal of quantitative linguistics* **1995,** *2* (3), 217-237.
23. ASTM Standard E1412-12, "Standard Practice for Separation of Ignitable Liquid Residues from Fire Debris Samples by Passive Headspace Concentration With Activated Charcoal", ASTM International, West Conshohocken, PA, 2012.
24. Stauffer, E., "Identification and characterization of interfering products in fire debris analysis" (2001). ProQuest ETD Collection for FIU. AAI1403692.

http://digitalcommons.fiu.edu/dissertations/AAI1403692.
25. Stauffer, E., Concept of pyrolysis for fire debris analysts. *Science & Justice* **2003,** *43* (1), 29-40.
26. Plasticizer for waterborne adhesives https://www.eastman.com/Literature_Center/L/L236.pdf (accessed 8/20/2018).
27. Why Choose Engineered Hardwood Flooring. https://www.builddirect.com/learning-center/flooring/engineered-hardwood-flooring/ (accessed 8/20/2018).
28. Innovative eco-efficient high fire performance wood products for demanding applications. http://virtual.vtt.fi/virtual/innofirewood/stateoftheart/database/burning/burning.html (accessed 8/20/2018).
29. Rowell, R. M., *Handbook of wood chemistry and wood composites*. CRC press: 2012.
30. White, M. A., The Chemistry behind Carbonless Copy Paper. *Journal of Chemical Education* **1998,** *75* (9), 1119.
31. Alsaleh, A.; Sattler, M. L., Waste Tire Pyrolysis: Influential Parameters and Product Properties. *Current Sustainable/Renewable Energy Reports* **2014,** *1* (4), 129-135.
32. Williams, P. T., Pyrolysis of waste tyres: A review. *Waste Management* **2013,** *33* (8), 1714-1728.
33. Kohler, M.; Künniger, T., Emissions of polycyclic aromatic hydrocarbons (PAH) from creosoted railroad ties and their relevance for life cycle assessment (LCA). *Holz als Roh- und Werkstoff* **2003,** *61* (2), 117-124.

34. Mueller, J. G.; Chapman, P. J.; Pritchard, P. H., Creosote-contaminated sites. Their potential for bioremediation. *Environmental science & technology* **1989,** *23* (10), 1197-1201.
35. Dipple, A.; Slade, T. A., Structure and activity in chemical carcinogenesis: Reactivity and carcinogenicity of 7-bromomethylbenz[a]anthracene and 7-bromomethyl-12-methylbenz[a]anthracene. *European Journal of Cancer (1965)* **1970,** *6* (5), 417-423.
36. Dromey, R. G.; Stefik, M. J.; Rindfleisch, T. C.; Duffield, A. M., Extraction of mass spectra free of background and neighboring component contributions from gas chromatography/mass spectrometry data. *Analytical Chemistry* **1976,** *48* (9), 1368-1375.
37. Definition of Signal to Noise Ratio https://www.ssi.shimadzu.com/products/gas-chromatography-mass-spectrometers/definition_sn_ratio.html (accessed September 18th).
38. Mallard, W. G., AMDIS in the Chemical Weapons Convention. *Analytical and Bioanalytical Chemistry* **2014,** *406* (21), 5075-5086.
39. Koavi, R., A Study of Cross – Validation and Bootstrap for Accuracy Estimation and Model Selection. In International Joint Conference on Artificial Intelligence, 1995, http://robotics.stanford.edu/~ronnyk/accEst.pdf (accessed: February 08, 2018).
40. Stone, J. V., *Bayes Rule: A Tutorial Introduction to Bayesian Analysis*. 1st Edition ed.; Sebtel Press.
41. T Mitchell; B Buchanan; G DeJong; T Dietterich; P Rosenbloom, a.; Waibel, A., Machine Learning. *Annual Review of Computer Science* **1990,** *4* (1), 417-433.
42. The Identification of Isopar H in Vinyl Flooring.
43. Baerncopf, J. M.; McGuffin, V. L.; Smith, R. W., Association of ignitable liquid residues to neat ignitable liquids in the presence of matrix interferences using chemometric procedures. *Journal of forensic sciences* **2011,** *56* (1), 70-81.
44. https://onlinecourses.science.psu.edu accessed June 6/13/2018.
45. P. Mair, S. P. R., P. M. Bentler, IRT Goodness-of-Fit Using Approaches from Logistic Regression. *UCLA, Department of Statistics Papers*.
46. Fawcett, T., An introduction to ROC analysis. *Pattern Recognition Letters* **2006,** *27* (8), 861-874.
47. Vuk, M.; Curk, T., ROC curve, lift chart and calibration plot. *Metodoloski zvezki* **2006,** *3* (1), 89.
48. Choi, B. C. K., Slopes of a Receiver Operating Characteristic Curve and Likelihood Ratios for a Diagnostic Test. *American Journal of Epidemiology* **1998,** *148* (11), 1127-1132.
49. Zadora, G.; Martyna, A.; Ramos, D.; Aitken, C., *Statistical analysis in forensic science: evidential values of multivariate physicochemical data*. John Wiley & Sons: 2014.
50. Meuwly, D.; Ramos, D.; Haraksim, R., A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic science international* **2017,** *276*, 142-153.
51. Ramos, D.; Franco-Pedroso, J.; Lozano-Diez, A.; Gonzalez-Rodriguez, J., Deconstructing Cross-Entropy for Probabilistic Binary Classifiers. *Entropy* **2018,** *20* (3), 208.
52. Ramos, D.; Gonzalez-Rodriguez, J., Reliable support: Measuring calibration of likelihood ratios. *Forensic science international* **2013,** *230* (1), 156-169.
53. Hanley, J. A.; McNeil, B. J., The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982,** *143* (1), 29-36.