

## Research Article

## Open Access

Praew Suppajariyawat, Ana Flavia Belchior de Andrade, Mathieu Elie, Mark Baron, Jose Gonzalez–Rodriguez\*

# The Use Of Chemical Composition And Additives To Classify Petrol And Diesel Using Gas Chromatography–Mass Spectrometry And Chemometric Analysis: A Uk Study

<https://doi.org/10.1515/chem-2019-0021>

received August 27, 2018; accepted November 26, 2018.

**Abstract:** The identification of the fuel found in a crime scene and establishing its source is important for forensic investigations. In this study petrol and diesel samples were analysed in order to identify chemical composition and additives with the aim of obtaining a set of markers or compounds that will allow forensic scientists to identify fuel sources. The study was performed using petrol and diesel sold in the city of Lincoln (UK) during four seasons (winter, spring, summer, autumn). Diesel samples, representing eight different brands were analyzed by gas chromatography–mass spectrometry (GC–MS) and principal component analysis (PCA). From the GC–MS analytical results, MTBE and ETBE were identified as the additive more preferably added in super unleaded petrol samples that can be used to aid in separation and identification. In diesel, the distribution of FAME contents showed the effect of seasonal variation as these were found in all spring, summer and autumn samples, but not found in all winter ones. The selection of a reduced number of key fuel compounds and additives was also shown

to be sufficient to achieve a high level of classification among the different fuel samples. Leave–one–out cross–validation was applied in order to validate the results.

**Keywords:** petrol; diesel; fuel additives; principal component analysis; gas chromatography–mass spectrometry.

## 1 Introduction

Petrol, sometimes referred to as gasoline, and diesel are the most important petroleum derived liquid products used in combustion engines. They are a complex mixture consisting mostly of organic hydrocarbons enhanced with a variety of special additives such as oxygenates, antiknock, antioxidants, anti–corrosion, anti–icing and cetane improvers. All of these additives are used to improve engine capacity [1] and can vary significantly depending on country and brand. Petrol and diesel have different chemical properties, boiling points and chemical components. Most fuel products vary depending on the original source of the crude oils used, the refining process, market demand, legislative requirements, and the particular product requirements with respect to the fuels use. Give this, the content of various types of additives or blending agents help to differentiate and characterize different fuel products.

The identification and classification of fuel is extremely important in a forensic context as it may help police forces to identify fuel sources in various fuel related offenses. These offenses can range from fuel fraud and spills through to arson and terrorism. For example, petrol and diesel provide powerful heat energy and are the choice for arsonists due to its high ignition power. Additionally, fuels can be combined with ammonium nitrate to make

\*Corresponding author: Jose Gonzalez–Rodriguez, School of Chemistry, University of Lincoln, Joseph Banks Laboratories, Green Lane, Lincoln, Lincolnshire LN6 7DL, United Kingdom, E-mail: [jgonzalezrodriguez@lincoln.ac.uk](mailto:jgonzalezrodriguez@lincoln.ac.uk)

Praew Suppajariyawat, Ana Flavia Belchior de Andrade, Mathieu Elie, Mark Baron: School of Chemistry, University of Lincoln, Joseph Banks Laboratories, Green Lane, Lincoln, Lincolnshire LN6 7DL, United Kingdom

Praew Suppajariyawat: Central Institute of Forensic Science, Ministry of Justice, 499 Sukprapute Building 16th Floor, Prachachuen Rd, Bang–Sue, Bangkok 10800, Thailand

Ana Flavia Belchior de Andrade: Criminalistic Institute, Federal District Civil Police, Secao de Pericia e Analises Laboratoriais–SPAL–SPO, Conjunto A, Lote 23, Bloco E, CEP 70.610–907 Brasilia, DF, Brazil

explosive substances such as ammonium nitrate fuel oil (ANFO), which is used by terrorist groups due to the ease of legally purchasing fuel and ammonium nitrate. Given this the ability to identify and classify the fuel samples from different sources is very useful for forensic investigation as it can potentially be used to find the link from sample to sample, sample to suspect and sample to scene.

Gas Chromatography-Mass Spectrometry (GC-MS) is widely used as a primary method for characterization of fuels. Typical analyses performed usually involve the identification of certain compounds [2-5], the use of ratios between chemicals [6,7,8] and identification and quantification of markers or additives [9-12]. Many previous studies successfully used GC-MS to determine the individual component composition and fuel additives. For example, the advantages and disadvantages of using one-dimensional detailed hydrocarbon analysis (DHA), two dimensional gas chromatography (2D-GC), two-dimensional gas chromatography mass spectrometry (2D-GC-MS) and multidimensional gas chromatography (MGC) in determining the individual component composition of oxygenates and benzene by retention times and mass spectra of compounds have been previously discussed in the literature [5]. Other examples in the literature offered the possibility of an accurate determination of the additive fraction in diesel fuel by using a two-step procedure, using high performance liquid chromatography (HPLC) in the first step followed by gas chromatography with flame ionization detection (GC-FID) and mass spectrometry (GC-MS) for the final determination. This use of a two-step procedure ensured reproducible results for determinations and reporting the limits of quantitation of diesel additives [6].

Over the past decades, there has been a growing interest in fuel source identification and classification using gas chromatography and chemometric methods. It has been shown in the literature that GC×GC-FID and parallel factor analysis can be successfully utilized to differentiate the chromatogram patterns between petrol (Type A) samples from Venezuela and Brazil [7]. Twenty-five diesel samples from 13 different US diesel brands were classified using GC-MS total ion chromatogram (TIC) and extracted ion chromatogram (EIC) data evaluated by Pearson product moment correlation (PPMC) and principal component analysis (PCA). The results showed that EICs provided higher discrimination power among samples than TICs. Aromatic components, typically  $m/z$  91, provided the greatest discrimination according to both statistical procedures [8]. The extracted ion chromatogram (EIC) of nine peak ratios identified by PCA, including the combinations of alkyl benzenes, sesquiterpanes, and alkyl

phenanthrenes, were used to classify 14 diesel samples from different US refineries [9]. Five unleaded petrol samples collected from 5 local gas stations in Seattle, USA were analyzed by GC-FID with PCA correctly classifying all chromatograms into 5 specific clusters [10]. The discrimination of 28 unevaporated petrol samples from New Zealand and Australia using GC-MS selected ion monitoring chromatograms (SIM) and PCA showed most of samples could be grouped based on their country of origin [11]. In Canada, forty-four samples of regular petrol (22 winter and 22 summer) and 44 samples of premium petrol (22 winter and 22 summer) sold throughout the country were analyzed by GC-MS. PCA correctly classified between 80–93% of the samples as premium and regular petrol while only 48–62% of the samples were correctly classified into their winter or summer sub-groups. The study also used artificial neural networks (ANN) for classification showing an approximately 97% correct classification of winter versus summer samples [12]. Five brands of petrol and diesel fuels collected from fifteen different local petrol stations in Madrid, Spain were successfully identified and discriminated using GC and ANN [13].

As mentioned previously, many of the previous studies mostly used the whole chemical composition as raw data for further chemometric or identification analyses. This study presents an example of separation of UK petrol and diesel samples from different brands in four different seasons using gas chromatography mass spectrometry and chemometric procedures to determine this minimum number of key compounds required to successfully separate samples from different brands and also take into account the seasonal variation effect in fuels. The novelty of this study lies in the selection of a small number of specific markers that can aid in the chemometric separation. The selection of a small number of markers has several practical advantages, namely faster chemometric analysis as a consequence of the use of a reduced number of variables, and the potential of using spectroscopic or non-chromatographic techniques to identify them in forensic samples.

## 2 Materials and methods

### 2.1 Sample collection

Ninety-one samples were collected from eight different petrol stations within Lincoln, Lincolnshire throughout the year in four different seasons – winter, spring, summer and

autumn. To preserve confidentiality all brands have been anonymized. These samples consisted of twenty-four samples of super unleaded petrol from six different brands, thirty-two samples of unleaded petrol from eight different brands and thirty-five samples of diesel from eight different brands. Samples were collected in 60 mL amber wide mouth bottles with PTFE liner polypropylene cap from Sigma Aldrich (Gillingham, UK). Sample bottles were sealed and stored at 20–25°C before analysis.

## 2.2 Sample preparation

Prior to GC–MS analysis, all samples were prepared in an appropriate solvent with constant concentration and a specific internal standard. For petrol analysis, each sample was diluted 1:100 in dichloromethane (99.6% reagent grade) from Fisher Scientific (Loughborough, UK). Each solution was then spiked with 1  $\mu\text{L}$  of 1,2-dimethoxyethane (analytical standard) from Sigma Aldrich (Gillingham, UK) and 3  $\mu\text{L}$  of ASTM D5769 deuterated internal standard mixture (Benzene- $\text{d}_6$ , ethylbenzene- $\text{d}_{10}$ , toluene- $\text{d}_8$  and naphthalene- $\text{d}_8$ ) from Spectrum Quality Standards (Texas, USA). For diesel, each sample was diluted 1:100 in dichloromethane (99.6% reagent grade) from Fisher Scientific (Loughborough, UK) and spiked with 100 ppm of internal standard *p*-terphenyl- $\text{d}_{14}$  (analytical standard) from Sigma Aldrich (Gillingham, UK).

## 2.3 GC-MS analysis

All samples were analysed and performed in sextuplicates on a Perkin Elmer<sup>a</sup> Clarus<sup>a</sup>500 Gas Chromatograph - Mass Spectrometer (GC-MS). The Clarus<sup>a</sup>500 GC-MS was equipped with a SLB<sup>a</sup>-5MS fused-silica capillary GC column (30 m  $\times$  0.25 mm  $\times$  0.25  $\mu\text{m}$ ) purchased from Sigma Aldrich (Gillingham, UK). The sample injection volume was 1  $\mu\text{L}$  at a split ratio of 10:1. Helium was used as a carrier gas at 1 mL min<sup>-1</sup>.

For petrol samples, the inlet line temperature was set at 280°C and the injector temperature was maintained at 220°C. The GC oven temperature program started with an initial temperature of 35°C and initial hold time of 7 min. Then the temperature then increased to 90°C at a rate of 4°C min<sup>-1</sup> before ramping to 220°C at a rate of 35°C min<sup>-1</sup>. The maximum temperature was held for 2 min. An electron impact ionization source was utilized at 70 eV with a mass analyzer operated in full scan mode ( $m/z$  35–260).

For diesel samples, the inlet line temperature was set at 300°C and the injector temperature was maintained at 280°C. The oven temperature was initially programed at 100°C and held for 5 min then ramped to 280°C at a rate of 10°C min<sup>-1</sup> and kept at that temperature for another 5 min. An electron impact ionization source was applied at 70 eV with a full scan mass range of  $m/z$  35–400.

All chromatography data were processed with the PerkinElmer TurboMass<sup>TM</sup> GC–MS Software (version 5.4.2.1617). All compound mass spectra were identified and compared using the Automated Mass Spectral Deconvolution and Identification System (AMDIS) (version 2.66) and the National Institute of Standards and Technology (NIST) mass spectral search program for the NIST/EPA/NIH mass spectral library (version 2.0).

## 2.4 Principal component analysis (PCA)

Multivariate analysis (MVA) was used for data analysis. The objective was to find features and characteristics among different compound ratios in order to identify the minimum number of fuel markers required to be able to distinguishing different brands of petrol and diesel samples. Principal component analysis (PCA) was used to analyze the structure of the data by rearranging the larger number of possibly correlated variables and reducing them to a smaller number of uncorrelated variables, the principal components (PCs) without any loss of chemical information [14]. In this study, the differences among various brands were obtained using the ratio between peak area ratios (PAR) from extracted ion chromatogram (EIC) of each analyte to their associated internal standards (IS). The typical highest mass ions ( $m/z$ ) of each individual compound used to provide peak area response are shown in Table 1. The ratios of oxygenated compounds in petrol were calculated using a 1,2-dimethoxyethane (1,2-DME) internal standard. The deuterated standards (benzene- $\text{d}_6$ , toluene- $\text{d}_8$ , ethylbenzene- $\text{d}_{10}$ , and naphthalene- $\text{d}_8$ ) were used to calculate the ratios for benzene, toluene,  $\text{C}_2$ – $\text{C}_3$  alkyl benzenes, and naphthalene, respectively. In diesel, *p*-terphenyl- $\text{d}_{14}$  was used to obtain the ratios of all components.

The dataset containing ratios of different compounds of each fuel type from different brands were evaluated by principal component analysis (PCA) and linear discriminant analysis (LDA) following by leave-one-out cross-validation (LOOCV) algorithms were performed using TANAGRA free data mining software version 1.4.50 (Ricco Rakotomalala, Lyon, France).

**Table 1:** List of petrol and diesel additives and compounds including internal standards and target ions.

Compounds	Ions ( <i>m/z</i> )	Compounds	Ions ( <i>m/z</i> )
<b>Petrol</b>		<b>Diesel</b>	
Methyl- <i>tert</i> -butyl-ether	73	Farnesane	57, 71
Ethyl- <i>tert</i> -butyl-ether	59, 87	Tetradecane	57, 43
1,2-dimethoxyethane*	45	Pentadecane	57, 43
Benzene- $d_6^*$	84	Hexadecane	57, 43
Benzene	78	Norpristane	57, 43
2,2,4-trimethylpentane	57	Octadecane	57, 43
Heptane	43	Phytane	57, 71
Toluene- $d_8^*$	98, 100	Nonadecane	57, 43
Toluene	91, 92	Palmitic acid, methyl ester	74, 87
Ethylbenzene- $d_{10}^*$	98, 116	Eicosane	57, 43
Ethylbenzene	91, 106	Linoleic acid, methyl ester	67, 81
<i>m</i> -xylene	91, 120	Oleic acid, methyl ester	55, 69
<i>p</i> -xylene	91, 120	Stearic acid, methyl ester	74, 87
<i>o</i> -xylene	91, 120	Docosane	57, 43
Propylbenzene	91, 120	<i>p</i> -terphenyl- $d_{14}^*$	244
3-ethyltoluene	105, 120		
4-ethyltoluene	105, 120		
1,3,5-trimethylbenzene	105, 120		
2-ethyltoluene	105, 120		
1,2,4-trimethylbenzene	105, 120		
1,2,3-trimethylbenzene	105, 120		
Naphthalene- $d_8^*$	136		
Naphthalene	128		

**Note:** \*Internal standard used in petrol and diesel analysis by GC-MS

Ethical approval: The conducted research is not related to either human or animal use.

### 3 Results and Discussion

At present, there are six major operating oil refineries in UK. Four of the oil refineries provide fuel throughout England, with one in Scotland and one in Wales. Each refinery has its own particular design characteristics depending on several factors such as market demand and production capacity, and this will cause the fuel products

from different refineries to vary. The fuel distributor can supply fuel from different refineries to the petrol station and hence the original source of the fuel in each sample in this study is unknown. In addition to differences in fuels due to the refining process, the chemical composition of fuel can also vary based on the additives added, transportation, the weather and the storage facilities. Apparently, some petrol retailing supermarkets also offer their own fuels which contain specific additive packages unique to their band. Therefore, it is expected that each fuel will contain its own specific chemical composition. Similar arrangements occur in other countries and hence it

is important to highlight that this study is only applicable to the UK and would need to be replicated in other places to obtain the specific markers and ratios for that region.

In The United Kingdom, the manufacture and use of petrol and diesel have been regulated under The Motor Fuel (Composition and Content) (Amendment) Regulations 2015 (S.I. 2015/1796) [15]. This regulation sets the guidelines and restrictions related to the sale and distribution of petrol and diesel including leaded petrol. Automotive fuels sold in UK markets must comply with the environmental fuel quality specification set out in Annex I (Petrol) and Annex II (Diesel) of Directive 2009/30/EC of the European Parliament and of the Council of 23 April 2009 [16]. Nowadays, the use of biodiesel is gradually increasing, with many countries significantly exceeding their capacity in producing it. Each has its own independent policies regarding geographic areas, climates, economics, use and production. Rapeseed oil is the important biodiesel feedstock in EU and the UK [17]. Apparently, it has required the blending up to 7% v/v of FAME in transport diesel in Europe and United Kingdom according to the European standard for automotive diesel (EN 590) [18] and the European FAME standard (EN 14214) [19].

### 3.1 Characterization by GC-MS

GC-MS data is often displayed as TIC. The information obtained from TIC can be used to identify or classify individual components contained in the sample. Figure 1 illustrates TICs and EICs of petrol, diesel and biodiesel blend samples.

For petrol, the presence and allocation of BTEX (Benzene, toluene, ethylbenzene, *m*-,*p*-,*o*-xylenes), heptane, C3-alkylbenzenes (propylbenzene, 3-ethyltoluene, 4-ethyltoluene, 1,3,5-trimethylbenzene, 2-ethyltoluene, 1,2,4-trimethylbenzene, 1,2,3-trimethylbenzene), naphthalene, and 1- and 2-methylnaphthaene were consistent between super unleaded and unleaded petrol samples in all seasons. These are distinctive compounds frequently found in petrol and are very important compounds used in petrol identification [20],[21].

Two oxygenates – methyl-*tert*-butyl ether (MTBE) and ethyl-*tert*-butyl ether (ETBE) – were identified and found irregularly in the samples. MTBE was found added in three out of six brands of super unleaded samples in all seasons. ETBE was found added in two out of six brands in spring and summer, three out of six brands in autumn, but not observed in any winter samples from any brand. For

one autumn sample of super unleaded petrol both MTBE and ETBE were detected. Moreover, one brand of the super unleaded petrol sample revealed the lack of oxygenates in every seasons.

Not all brands of unleaded petrol samples contained an oxygenate compound. For those that did, MTBE was the only oxygenate compound found, with one out of eight brands in winter and three out of eight brands in summer and autumn. All spring unleaded petrol samples showed a lack of oxygenates.

Isooctane is an important component of petrol, frequently used as an antiknock agent, and was observed in almost all samples. Exceptionally, only one brand of super unleaded petrol sample exhibited the lack of isooctane which occurred during winter.

According to the standard specifications of petrol in the UK (BS EN228: 2012 Automotive fuels unleaded petrol requirement and test method) [22], oxygenates or any other oxygen-rich additives can be added to petrol in order to increase the octane rating. Our results confirmed that two types of oxygenates, MTBE and ETBE, were found more in super unleaded petrol suggesting that these additives were used to achieve the higher octane rating associated with premium unleaded fuels.

For diesel, the distribution of major peaks of hydrocarbon compounds varies between  $C_{10}$ – $C_{21}$ . This was constant between the four sets (winter, spring, summer and autumn) of diesel samples. Also, the intensities of these dominant peaks varied among different brands, explaining the dissimilarity of ratios.

Most of diesel compounds observed in this study consisted of undecane, dodecane, tridecane, 2,6,10-trimethyldodecane(farnesane),tetradecane,pentadecane, hexadecane, 2,6,10-trimethylpentadecane (Norpristane), heptadecane, 2,6,10,14-tetramethylpentadecane (pristane), octadecane, 2,6,10,14-tetramethylhexadecane (phytane), nonadecane, eicosane, heneicosane, docosane, tricosane, tetracosane and pentacosane. These are target compounds frequently found in diesel with some used for diesel identification according to the ASTM E1618 standard [21]. Additionally, fatty acid methyl esters (FAMES) were also evidently observed. Four identified contents of FAME – palmitic acid methyl ester (methyl palmitate, C16:0), linoleic acid methyl ester (methyl linoleate, C18:2), oleic acid methyl ester (methyl oleate, C18:1) and stearic acid methyl ester (methyl stearate, C18:0) – were detected.

**FAMES are a major component found in biodiesel, a renewable energy source that is less polluting than petroleum diesel** and often blended with petroleum diesel. Such blends are known as biodiesel blends and are found in many different concentrations for specific



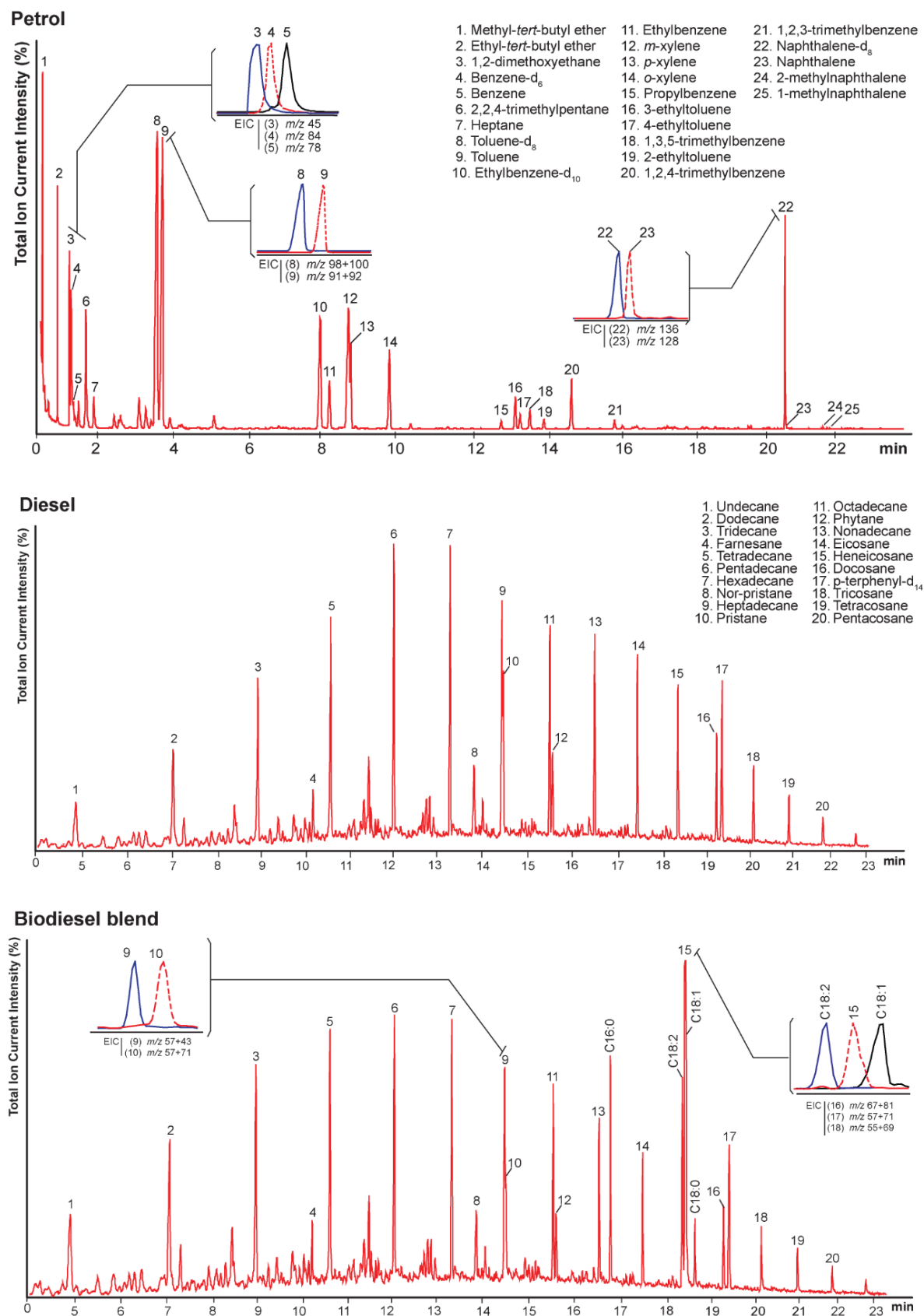


Figure 1: Total ion chromatograms and extracted ion chromatograms of petrol, diesel and biodiesel blend fuel samples.

**Table 2:** The additives and compounds used for PCA.

Types	No. of compounds	Description
Petrol	18	MTBE, ETBE, benzene, isooctane, heptane, toluene, ethylbenzene, <i>m</i> -xylene, <i>p</i> -xylene, <i>o</i> -xylene, propylbenzene, 3-ethyltoluene, 4-ethyltoluene, 1,3,5-trimethylbenzene, 2-ethyltoluene, 1,2,4-trimethylbenzene, 1,2,3-trimethylbenzene, naphthalene
Diesel	14	farnesane, tetradecane, pentadecane, hexadecane, norpristane, octadecane, phytane, nonadecane, eicosane, docosane, methyl palmitate, methyl linoleate, methyl oleate, methyl stearate

purposes. **FAMES use vary on regional and seasonal basis, for example in cold climate regions where fuel quality is more critical FAMES content is reduced.** This is because FAMES can cause waxing and gelling, as their relative viscosity increases at low temperatures, and this can lead to filter clogging in winter. Correspondingly, FAMES are such a rich source of food for bacteria that microbiological growth in the tank can also occur and cause blockage in a diesel engine system [23], [24], [25]. This is the likely reason for seasonal variation in FAME content in samples. In this study, four FAMES species were found in four out of nine winter diesel samples, and in all spring, summer and autumn diesel samples.

### 3.2 Classification by PCA

In this study, PCA was used to establish a classification model. Cross-validation and leave-one-out analysis were further employed to evaluate the classification model and correct the misclassified samples from PC visual inspection. PCA was performed on three different datasets containing: the 18 petrol compounds and the 144 (24 samples  $\times$  6 repeats) super unleaded petrol samples representing 18 $\times$ 144 data matrix, the 18 petrol compounds in the 192 (32 samples  $\times$  6 repeats) unleaded petrol samples representing 18 $\times$ 192 data matrix and the 14 diesel compounds in the 210 (35 samples  $\times$  6 repeats) diesel samples representing 14 $\times$ 210 data matrix. Table 2 details the compounds used for PCA. In this study, the first three principal components (3PCs) were used to describe all data according to the percentage of total variance explained over 80 – 90% of all samples and 2-dimension scatterplots were used to establish visual classification of models.

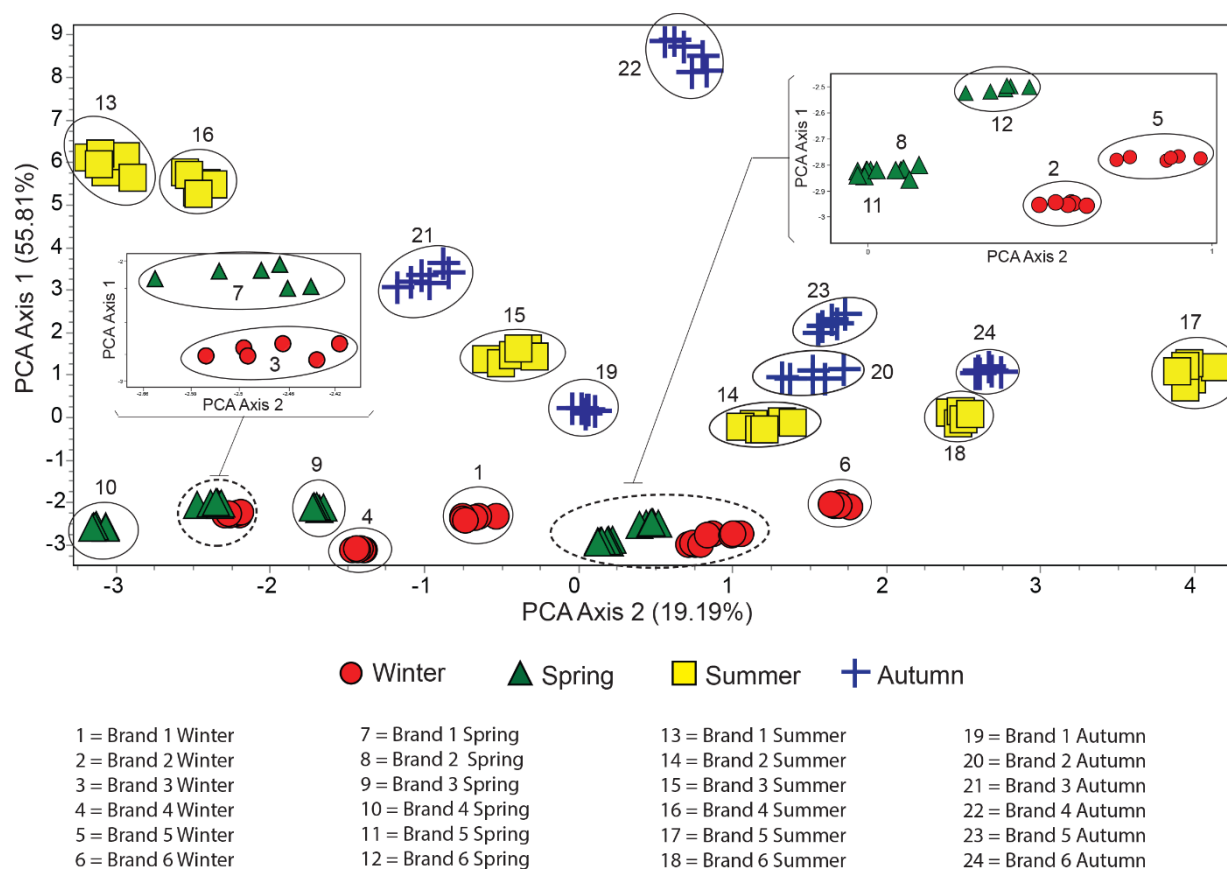
#### 3.2.1 Super unleaded petrol

PCA demonstrated that the most of the variability in the data can be explained by the first three principal components (approximately 83%). The first principal component is responsible for 55.81% of the entire information, the second accounts for 19.19%, and the third explain 8.27% of the total information. Figure 2 shows that clustering of 24 super unleaded petrol samples from six different brands collected in four seasons did emerge when PC1 was plotted against PC2. A good separation was visually made among each different sample when the figure is enlarged. Although there were two spring samples from two brands exhibit overlapping clusters, however, the majority of the samples were apparent and clearly classified. This can be explained that the samples from different brands and different time of collection most likely contain unique chemical compositions. From this, it can be concluded that 18 petrol components obviously provide the discrimination ability among super unleaded petrol samples from different brands and seasons.

#### 3.2.2 Unleaded petrol

**PCA was also performed on the whole dataset of 32 unleaded petrol samples from eight different brands in four different seasons using entire 17 petrol (without the presence of ETBE) compounds as mentioned in Table 2.**

Figure 3 shows the ability of PCA to classify all 32 unleaded petrol samples from 8 different brands collected in 4 seasons. The majority of the samples were visually clustered when PC1 was plotted against PC2 as they accounted for 87.90% of the total data variability. From the plot, a better classification was visibly made after zooming the figure scale. However, there was still an overlap among winter samples of brand 5, 7 and 8. Also, the spring samples of brand 1 and brand 5 were not clearly separated. Two autumn samples of brand 3 and brand



**Figure 2:** A scatterplot of PC1 vs PC2 showing the classification of 24 super unleaded petrol samples using all 18 petrol compounds. The correctly classified clusters are shown in circles.

8 were not separated. From this, the samples which were misclustered or not separated may contain similar chemical compositions. On the contrary, the samples which were well-separated more likely have distinctive compositions. Hence, the use of 17 different petrol compounds are able to offer the ability to differentiate unleaded petrol samples from different sources.

### 3.2.3 Diesel

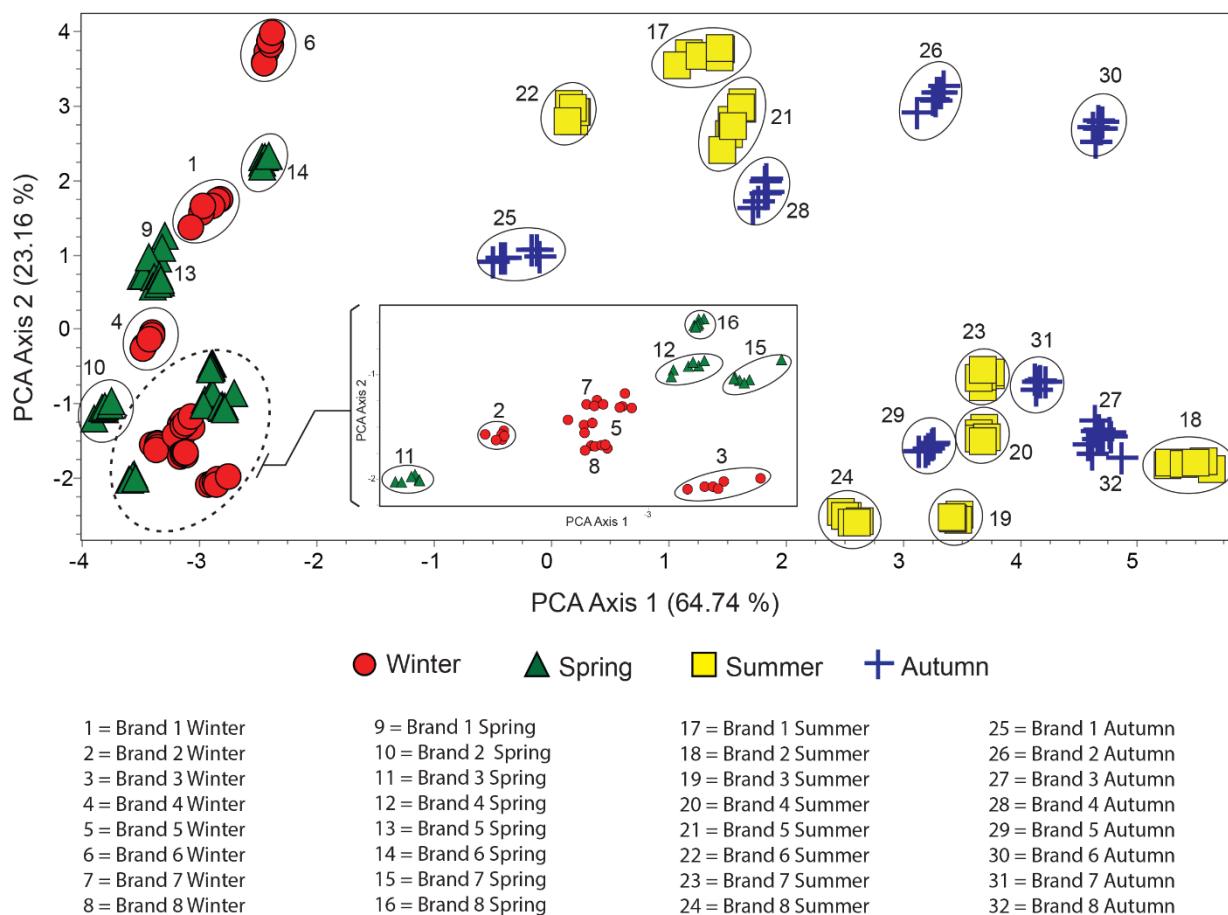
The ability of PCA to classify all 35 diesel samples from 8 different brands collected from 4 different seasons is shown in Figure 4. More than half of the samples were visually clustered when PC1 was plotted against PC2 which contributed for 89.70% of the total data variability. A better separation was visibly made when the figure is enlarged. However, the samples from brand 6 autumn and brand 1A summer were not clearly separated. Also, there are some overlaps among brand 7 winter, brand 5 winter and brand 4 spring samples. In the same way, the spring

samples of brand 1, 2 and 5 exhibit overlapping clusters. From this result, it demonstrates that most samples which were clustered particularly contain different chemical compositions and the samples which were unclassified likely contain similar chemical compositions. Overall, the 14 different diesel compounds can be potentially used to distinguish among diesel samples from different sources.

### 3.3 PCA classification using key fuel compounds

**Although PCA scatterplots can provide a visual classification, the difficulty of classification might appear when the view of data is limited by a two-dimensional plot. For this reason, linear discriminant analysis (LDA) following by leave-one-out cross-validation (LOOCV) are further required to evaluate the classification model and correct the misclassified samples from PC visual inspection.**





**Figure 3:** A scatterplot of PC1' PC2 showing the classification of 32 unleaded petrol samples using all 18 petrol compounds. The correctly classified clusters are shown in circles.

Correlation to the principal components was also evaluated by using factor loading scores. Factor loading is the correlation coefficient between variables and principle components. Factor loadings can range from  $-1$  to  $1$ , the closer to  $-1$  or  $1$  the stronger the effect of the variable on the PCs. Figure 5 shows the correlation coefficient of factor loadings in super unleaded petrol, unleaded petrol and diesel samples.

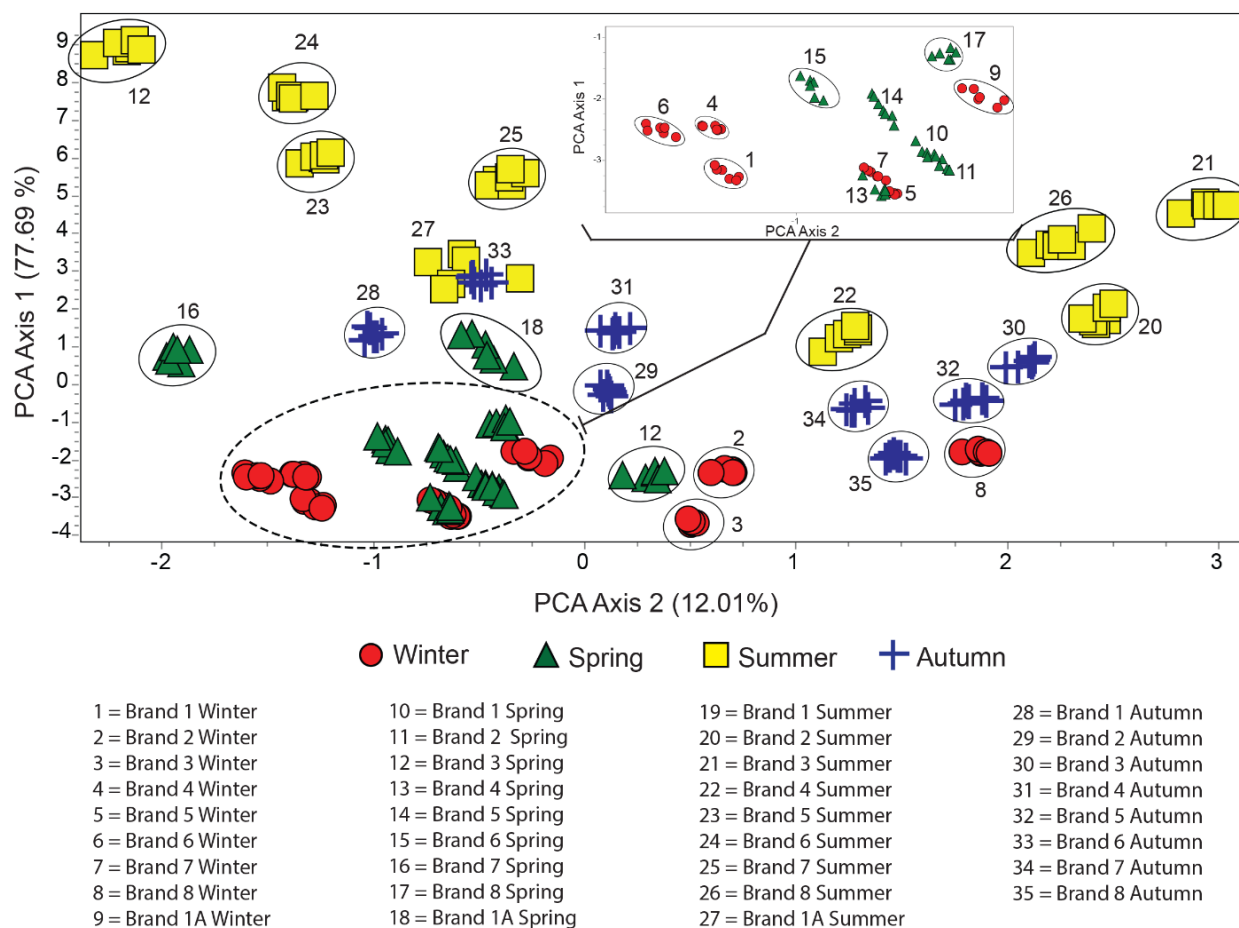
From Figure 5 it can be inferred that most of fuel compounds in the three types of fuel samples exhibited relatively high factor loading to PC1, with some to PC2 and PC3 as shown in unleaded petrol samples. These factor loadings can be used to indicate the weight of each of the variables on the different PCs under study (PC1, PC2 and PC3) and should be used to establish a classification model and also to identify key specific compounds for fuel classification of those PCs.

In an attempt to identify the key components for fuel analysis in the UK, the most valuable specific indicators should be the only ones considered. It was expected that

the reducing number of effective compounds would give reliable and significant discrimination result. In this study, a different set of components were selected based on their factor loading scores and individual chemical properties.

The percentages of cumulative variation in the data set are shown for the chosen petrol and diesel compounds in Table 3. This indicates that the reduction in the number of compounds can affect the PC model. The table also shows the percentage of correct classification on the first two and three principal components for each set of compounds using linear discriminant analysis (LDA) followed by a leave-one-out cross-validation algorithm.

From Table 3, it is apparent that for all fuel samples using all different compound sets, PCA gave poor classification performance with respect to their seasons sub-group was also considerably low for most compound sets ranging from 26.39% to 83.81% (with 2PCs) and 53.47 to 90.48% (with 3PCs). Similarly, the ability of PCA to acquire correct classification among fuel samples with respect to the



**Figure 4:** A scatterplot of PC1 vs PC2 showing the classification of 35 diesel samples using all 14 diesel compounds. The correctly classified clusters are shown in circles.

brand sup-group with correct classification ranging between 8.57% to 38.89% (with 2PCs) and 13.33% to 50% (with 3PCs). Classification performance of over 80% was obtained from diesel samples using a set of FAME compounds. This indicates that the use of FAME compounds is able to classify different diesel samples from different seasons.

The PCA model was able to correctly classify most samples with a close to 100% success rate for some sets of compounds. Nonetheless, while correct classification was generally high, different sets of compounds showed some differences in the degree of correct classification for the three types of fuel samples.

The maximum correct classification was 100% for super unleaded samples using the set of all 18 petrol compounds and the set of MTBE, ETBE, isooctane, toluene, 1,2,4-TMB, and 1,2,3-TMB with 3PCs; 99.48% for unleaded petrol samples using the set of all 17 petrol compounds with 3PCs; and 99.52% for diesel samples using the set of

all 14 diesel compounds with 3PCs. This indicates that the more compounds that are used the greater classification performance obtained.

The set of the top 5 highest factor loadings successfully distinguished different unleaded petrol and diesel samples obtaining 75% to 95% correct classification. However, reduced classification performance was obtained in super unleaded petrol samples resulting in only 64.58% to 75% correctly classified with 2PCs and 3PCs, respectively. Therefore, this set of compounds may not be useful for forensic investigations because of their inconsistency depending on the types of sample and slightly low classification performance compared to other sets.

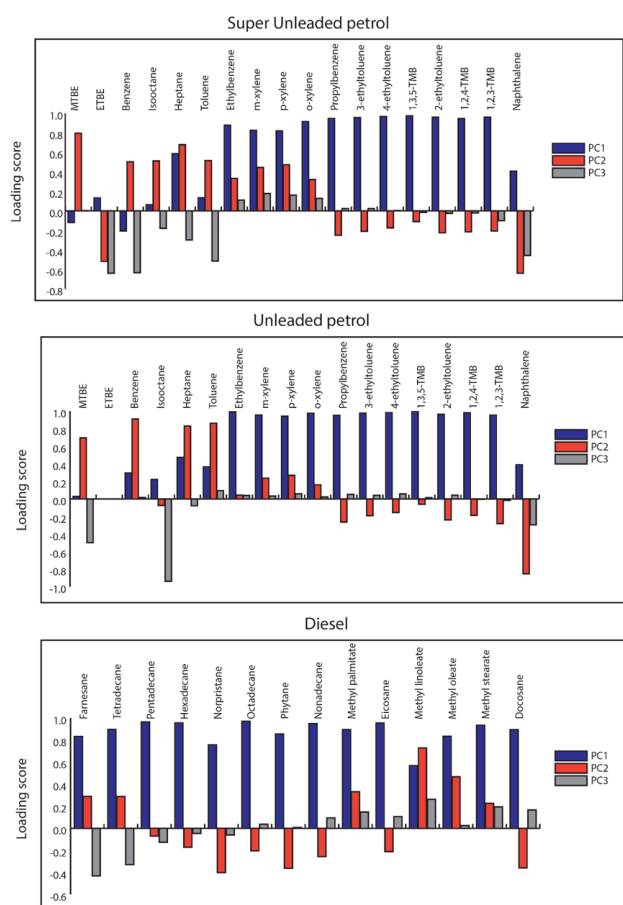
C3-alkylbenzene compounds showed high classification performance in unleaded petrol samples (over 90% success), but low classification performance in super unleaded petrol samples (less than 80% success). These are the most stable compounds from weathering in petrol [26], however, identification of the 8 types of

**Table 3:** The summary of PCA total data variability and leave-one-out cross-validation among different set of petrol and diesel compounds using factor loadings.

Sets of compounds	Cumulative Variance (%)	Correct Classification (%)		
		Season	Brand	All Samples
Super unleaded petrol				
All 18 petrol compounds				
PC1ˆPC2	75.00	55.56	34.72	91.67
PC1ˆPC2ˆPC3	83.27	53.47	50.00	100
Top 5 highest factor loading				
PC1ˆPC2	99.68	51.39	31.94	64.58
PC1ˆPC2ˆPC3	99.92	54.17	46.53	75.00
BTEX compounds				
PC1ˆPC2	93.12	62.50	24.31	94.44
PC1ˆPC2ˆPC3	99.10	67.36	29.17	99.31
C3–alkylbenzene compounds				
PC1ˆPC2	99.58	26.39	27.78	63.89
PC1ˆPC2ˆPC3	99.84	54.17	45.83	77.78
MTBE, ETBE, isooctane, toluene, 1,2,4–TMB, 1,2,3–TMB				
PC1ˆPC2	62.12	50.00	38.89	93.75
PC1ˆPC2ˆPC3	79.38	57.64	43.06	100
Unleaded petrol				
All 17 petrol compounds (without ETBE)				
PC1ˆPC2	87.90	39.58	27.60	94.27
PC1ˆPC2ˆPC3	95.24	78.12	23.44	99.48
Top 5 highest factor loading				
PC1ˆPC2	99.70	57.84	13.54	84.37
PC1ˆPC2ˆPC3	99.92	64.06	26.56	95.83
BTEX compounds				
PC1ˆPC2	98.54	59.37	26.04	91.15
PC1ˆPC2ˆPC3	99.69	66.67	31.77	94.79
C3–alkylbenzene compounds				
PC1ˆPC2	99.78	31.25	17.71	90.10
PC1ˆPC2ˆPC3	99.96	81.25	19.27	92.19
MTBE, ETBE, isooctane, toluene, 1,2,4–TMB, 1,2,3–TMB				
PC1ˆPC2	73.05	45.31	18.75	91.67
PC1ˆPC2ˆPC3	93.13	75.00	15.62	98.96
Diesel				
All 14 diesel compounds				
PC1ˆPC2	89.70	74.29	10.00	94.29

**Table 3:** The summary of PCA total data variability and leave-one-out cross-validation among different set of petrol and diesel compounds using factor loadings.

Sets of compounds	Cumulative Variance (%)	Correct Classification (%)		
		Season	Brand	All Samples
PC1'PC2'PC3	93.23	77.62	22.86	99.52
<i>Top 5 highest factor loading</i>				
PC1'PC2	99.24	47.14	13.33	80.95
PC1'PC2'PC3	99.76	70.00	20.95	87.62
<i>All alkane compounds</i>				
PC1'PC2	92.84	66.19	20.00	96.19
PC1'PC2'PC3	96.25	66.67	22.86	96.67
<i>All 4 FAME compounds</i>				
PC1'PC2	95.18	83.81	8.57	84.76
PC1'PC2'PC3	98.83	90.48	13.33	85.24
<i>Farnesane, norpristane, phytane, and 4 FAMES</i>				
PC1'PC2	87.83	75.71	14.76	97.14
PC1'PC2'PC3	92.35	74.29	19.05	98.10



**Figure 5:** Loading scores of each individual compound for the first 3PCs of super unleaded petrol, unleaded petrol and diesel samples.

C<sub>3</sub>-alkylbenzene isomers using gas chromatography is relatively complex, unless a highly effective capillary column with good separation is used and all their reference standards can be provided [2].

BTEX compounds were able to be used to successfully distinguish different petrol samples, with 99.31% of super unleaded samples and 94.79% of unleaded petrol samples correctly classified. Nevertheless, BTEX are found in many chemicals such as paints, synthetic rubbers and agricultural chemicals [2, 27], and hence can contaminate forensic samples from these sources. Also, BTEX has less solubility and persistence than additives such as MTBE and ETBE in groundwater and soils [27]. This limits the use of BTEX as the first choice of forensic indicators to use for fuel source identification in applications such as fire debris analysis and oil spill identification.

The set of all alkane compounds showed relatively high classification performance with over 95% correctly classified. Theoretically, n-alkanes are more susceptible to microbial degradation than branched alkanes, small aromatics, cyclic alkanes and polycyclic aromatic hydrocarbons (PAHs) [28]. This could also restrain n-alkane compounds as the best choice of forensic markers used for diesel oil spill identification.

FAME compounds can be used to differentiate diesel samples from different seasons with over 80% correctly classified when samples were divided into season sub-

group. High classification performance of 84–85% success could be successfully obtained in all diesel samples. As a result, these markers could be used to distinguish diesel samples from different seasons due to the lower concentration or complete lack of FAMES in winter samples, as commented before.

The combination of oxygenates, and aliphatic and aromatic hydrocarbons consisting of MTBE, ETBE, isooctane, toluene, 1,2,4-trimethylbenzene and 1,2,3-trimethylbenzene were also tested in this study and successfully showed excellent classification performance of 100% (3PCs) and 98.96% (3PCs) in super unleaded and unleaded petrol samples, respectively. MTBE and ETBE are typical oxygenate blending additives in petrol used as octane booster and required by different fuel manufacturers and used as an indicator of long-term petrol contamination due to more solubility and persistence in environment [27]. Isooctane and toluene are typical anti-knock agents widely in use in petrol. Besides, MTBE, ETBE and isooctane showed the variation in different petrol samples in this study. Thus, these additive can also be beneficial as a key marker for fuel identification and classification. Moreover, 1,2,4-trimethylbenzene and 1,2,3-trimethylbenzene are  $C_3$ -alkylbenzenes which are known to be the most stable compounds from weathering in petrol [26] and more resistant to microbial degradation [29, 30]. This means that the individual characteristics, specific properties and the benefits of MTBE, ETBE, isooctane, toluene, 1,2,4-trimethylbenzene and 1,2,3-trimethylbenzene as key petrol specific indicators make them more interesting among other compounds to be used as markers. Thus, it can be concluded that the use of additives such as oxygenates and highly persistent petrol compounds offers more reliability for petrol identification and classification.

For diesel, the combination of the iso-alkanes (farnesane, norpristane, phytane) and FAME compounds (palmitic acid methyl ester, linoleic acid methyl ester, oleic acid methyl ester and stearic acid methyl ester) – showed good separation in diesel samples according to their high classification performance (over 95%). The reasons for selecting the set of iso-alkanes and FAME compounds to classify diesel samples was due to farnesane (2,6,10-trimethyldodecane) being a renewable hydrocarbon and also a diesel biomarker most commonly blended with petroleum diesel and jet fuel as well as FAME products [29]. These compounds are claimed to help reduce the problem of air pollution and also the consumption of imported fossil fuel. Most energy industries preferably use the renewable hydrocarbon diesel mixed with petroleum diesel as it is more environmentally friendly, cheaper and

easier to manufacture. Norpristane and phytane are more resistant to biodegradation compared to n-alkanes [29, 30]. Moreover, these compounds are a good biomarker and essential in classifying and identifying diesel [6, 32]. The seasonal variation effect is also found when FAMES contents are used. From these reasons, it can be indicated that the relative amount of iso-alkanes versus FAMES components provided the increased discrimination among diesel samples and is hence why this set of compounds is one of the best choices for diesel identification and classification.

## 4 Conclusions

GC-MS analytical results of petrol confirmed that MTBE and ETBE are commonly used as petrol additives in the UK and likely added in super unleaded petrol. In addition, the inconsistencies in the presence and concentration of isooctane can be associated to the different blending processes. In diesel, the distribution of FAME contents showed the effect of seasonal variation which presented in all spring, summer and autumn samples, but only some in some winter samples.

In this study, PCA following by LDA and leave-one-out cross-validation allowed the samples of petrol and diesel from different brands and different seasons to be classified with a high degree of accuracy. Unfortunately, the PCA was not able to differentiate samples into their season and brand sub-groups. It is expected that this is caused by the composition of fuel samples varying depending on various factors such as the original source of crude oils, the refining processes, the product specifications regarding its purposes and uses, the transportation, and the storage tank.

Moreover, the selection of subgroups of key fuel markers showed that using a number of them provided different degrees of discrimination of the fuel samples compared to using entire chromatogram or the full range of compounds in the samples. The selection of farnesane, norpristane, phytane and the four FAME compounds is the best option when trying to separate the diesel sample source. Due to high volatility, some compounds might not be useful for fire debris analysis or oil spill identification, yet these compounds are beneficial for forensic identification and comparison such as in fuels fraud and terrorism cases as its information can provide intelligence data for forensic investigators.

It is important to highlight that this study only accounted for the UK fuel samples in the area of study. Significant components such as additives and fuel



markers used in fuels depends on the type of fuel and varies between countries, regions, seasons, parent oil sources and different manufacturing processes. Thus, a similar methodology using GC–MS and chemometric analysis as described in this study can be applied to provide association of petrol and diesel samples for different sources in a specific country.

**Acknowledgement:** The authors gratefully acknowledge the Office of the Civil Service Commission (OCSC), Royal Thai Government and Central Institute of Forensic Science (CIFS), Ministry of Justice, Thailand for their financial support and cooperation of this research.

**Conflict of interest:** Authors declare no conflict of interest.

## References

- [1] Owen K., Gasoline and Diesel Fuel Additives. New York, NY: John Wiley and Sons Inc., 1989.
- [2] Wang Z., Fingas M., Landriault M., Sigouln L., Xu N., Identification of Alkylbenzenes and Direct Determination of BTEX and (BTEX +C3 Benzenes) in Oils by GC/MS, *Anal. Chem.*, 1995, 67(19), 3491–3500.
- [3] Xiong W., Bernesky R., Bechard R., Michaud G., Lang J., A tiered approach to distinguish sources of gasoline and diesel spills, *Sci. Total Environ.*, 2014, 487, 452–462.
- [4] Achten C. and Püttmann W., Method for determination of methyl-tert-butyl ether in gasoline by gas chromatography, *J. Chromatogr. A*, 2001, 910(2), 377–383.
- [5] Kanateva A.Y., Paleev A.V., Kurganov A.A., Gorshkov A.V., Gribanovkaya M.G., Determination of oxygenates and benzene in gasoline by various chromatographic techniques, *Pet. Chem.*, 2013, 53(5), 349–355.
- [6] Boczkaj G., Jaszczolt M., Przyjazny A., Kamiński M., Application of normal-phase high-performance liquid chromatography followed by gas chromatography for analytics of diesel fuel additives, *Anal. Bioanal. Chem.*, 2013, 405(18), 6095–103.
- [7] de Godoy L.A.F., Pedroso M.P., Hantao L.W., Augusto F., Poppi R.J., Determination of Fuel Origin by Comprehensive 2D GC-FID and Parallel Factor Analysis, *J. Braz. Chem. Soc.*, 2013, 24(4), 645–650.
- [8] Hupp A.M., Marshall L.J., Campbell D.I., Smith R.W., McGuffin V.L., Chemometric analysis of diesel fuel for forensic and environmental applications, *Anal. Chim. Acta*, 2008, 606(2), 159–171.
- [9] Gaines R.B., Hall G.J., Frysinger G.S., Gronlund W.R., Juare K.L., Chemometric Determination of Target Compounds Used to Fingerprint Unweathered Diesel Fuels, *Environ. Forensics*, 2006, 7(1), 77–87.
- [10] Pierce K.M., Hope J.L., Johnson K.J., Wright B.W., Synovec R.E., Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis, *J. Chromatogr. A*, 2005, 1096(1–2), 101–110.
- [11] Sandercock P.M.L. and Du Pasquier E., Chemical fingerprinting of gasoline: 3. Comparison of unevaporated automotive gasoline samples from Australia and New Zealand, *Forensic Sci. Int.*, 2004, 140(1), 43–59.
- [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 Sci. Int.*, 2003, 132(1), 26–39.
- [13] Ugena L., Moncayo S., Manzoor S., Rosales D., Cáceres J.O., Identification and Discrimination of Brands of Fuels by Gas Chromatography and Neural Networks Algorithm in Forensic Research, *J. Anal. Methods Chem.*, 2016.
- [14] Zadora G., Martyna A., Ramos D., Aitken C.G.G., Statistical Analysis in Forensic Science: Evidential Values of Multivariate Physicochemical Data. John Wiley & Sons, 2013.
- [15] Statutory Instruments, The Motor Fuel (Composition and Content) (Amendment) Regulations 2015: SI 2015/1796). UK, 2015.
- [16] European Commission, Directive 2009/30/EC of the European Parliament and of the Council of 23 April 2009, Dir. 2009/30/EC, 2009, pp. 88–113.
- [17] Worldwatch Institute, Part I: Status and global trends, in Biofuels for Transport - Global Potential and Implications for Sustainable Energy and Agriculture, Earthscan, 2012, 2007, p. 481.
- [18] European Committee for Standardization, BS EN 590:2013+A1:2017 Automotive fuels - Diesel - Requirements and test methods, 2013, p. 26.
- [19] European Committee for Standardization, BS EN 14214:2012+A1:2014 Liquid petroleum products - Fatty acid methyl esters (FAME) for use in diesel engines and heating applications - Requirements and test methods, 2013, p. 28.
- [20] Stauffer E., Dolan J.A., Newman R., Flammable and Combustible Liquids, in Fire Debris Analysis, pp. 199–233, 2008.
- [21] American Society for Testing and Materials, ASTM E1618-14: Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography– Mass Spectrometry, ASTM Int., 2014.
- [22] European Committee for Standardization, BS EN 228:2012 Automotive fuels - Unleaded petrol - Requirements and test methods, 2012, p. 26.
- [23] Passman F.J., Microbial contamination and its control in fuels and fuel systems since 1980 – a review, *Int. Biodeterior. Biodegradation*, 2013, 81, 88–104.
- [24] Sørensen G., Pedersen D.V., Nørgaard A.K., Sørensen K.B., Nygaard S.D., Microbial growth studies in biodiesel blends, *Bioresour. Technol.*, 2011, 102(8), 5259–5264.
- [25] Siegert W., Microbial Contamination in Diesel Fuel – Are new Problems Arising from Biodiesel Blends ?, proceeding IASH 2009, 11th Int. Conf. Stability, Handl. Use Liq. Fuels Prague, Czech Repub. Oct., 2009, 18–22.
- [26] Wang Z. and Stout S.A., Oil spill environmental forensics : fingerprinting and source identification. Elsevier/Academic Press, 2007.
- [27] Rosell M., Lacorte S., Ginebreda A., Barceló D., Simultaneous determination of methyl-tert-butyl ether and its degradation products, other gasoline oxygenates and benzene, toluene,

ethylbenzene and xylenes in Catalanian groundwater by purge-and-trap-gas chromatography–mass spectrometry, 2003, 995, 1–26.

- [28] Rojo F., Degradation of alkanes by bacteria, *Environ. Microbiol.*, 2009, 11(10), 2477–2490.
- [29] Amyris, Farnesane: a Breakthrough Renewable Hydrocarbon for Use as Diesel and Jet Fuel, Presidential Green Chemistry Challenge: 2014 Small Business Award, 2014. [Online]. Available: <https://www.epa.gov/greenchemistry/presidential-green-chemistry-challenge-2014-small-business-award>. [Accessed: 23-May-2018].

---

**Supplemental Material:** The online version of this article offers supplementary material (<https://doi.org/10.1515/chem-2019-0021>).