ELSEVIER

Forensic Science International 132 (2003) 26–39

Forensic Science International

www.elsevier.com/locate/forsciint

Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks

Philip Doble[a,*], Mark Sandercock[a], Eric Du Pasquier[a], Peter Petocz[b], Claude Roux[a], Michael Dawson[a]

[a]*Department of Chemistry, Materials and Forensic Science, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia*
[b]*Department of Mathematical Sciences, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia*

Received 16 January 2002; received in revised form 10 August 2002; accepted 8 November 2002

Abstract

Detection and correct classification of gasoline is important for both arson and fuel spill investigation. Principal component analysis (PCA) was used to classify premium and regular gasolines from gas chromatography–mass spectrometry (GC–MS) spectral data obtained from gasoline sold in Canada over one calendar year. Depending upon the dataset used for training and tests, around 80–93% of the samples were correctly classified as either premium or regular gasoline using the Mahalanobis distances calculated from the principal components scores. Only 48–62% of the samples were correctly classified when the premium and regular gasoline samples were divided further into their winter/summer sub-groups. Artificial neural networks (ANNs) were trained to recognise premium and regular gasolines from the same GC–MS data. The best-performing ANN correctly identified all samples as either a premium or regular grade. Approximately 97% of the premium and regular samples were correctly classified according to their winter or summer sub-group.

Keywords: Gasoline; Artificial neural networks; Gas chromatography–mass spectrometry; Principal component analysis

1. Introduction

The identification of gasoline is of forensic significance for the investigation of arson and environmental spills. In arson, petroleum-based accelerants such as gasoline, kerosene, and paint thinners are often used to accelerate a fire. In some cases, liquid accelerant is left at the scene, which may be matched to samples that are associated with the suspect. In the environment, gasoline spills are commonplace, but identification of the source is not always straightforward.

The identification of gasoline is crucial for the successful prosecution of an offending individual and/or company. Current methods of identification usually involve analysis of the gasoline by gas chromatography–mass spectrometry (GC–MS) and sample matching performed by visual comparison of target compounds and/or inspection of the profile of the gas chromatogram [1]. One problem with this technique is that the interpretation and classification of the data is limited by the skill and experience of the analyst.

One approach to overcome the problems associated with the subjectivity of the analyst for interpretation of GC–MS data is the principal component analysis (PCA). PCA is a multivariate statistical technique that reduces the dimensionality of a dataset, whilst retaining as much as the variability of the original data as possible. PCA is particularly helpful in the analysis of datasets containing a large number of variables, such as GC–MS profiles of gasoline. This reduction is achieved by sequential linear transformations of the data, where the first few transformations retain most of the variation of the original variables. Computation of the principal components reduces to the solution of an eigenvalue–eigenvector matrix. This solution allows organisation of results in lower dimensional space, leading to classification through

* Corresponding author.
URL: http://www.forensics.edu.au

0379-0738/03/$ – see front matter
doi:10.1016/S0379-0738(03)00002-1

visual interpretation of scatter plots of the principal components. Such scatter plots can be interpreted through examination of clusters, each cluster representing a particular class. Observations may also be classified into a group using linear discriminant analysis (LDA). LDA calculates the squared distance (also called the Mahalanobis distance) of an observation from a group to decide the class membership. In this case, an assumption is made that covariance matrices are equal for all groups.

Tan et al. [2] analysed five classes of petroleum-based accelerants by GC–MS and performed PCA on the GC–MS data. The resulting PCA scatter plots were interpreted using a soft independent model classification analogy (SIMCA) model. SIMCA is a method in which confidence envelopes are constructed to contain the data points. Unknowns falling within the confidence envelope are identified [3]. With this approach, the authors successfully classified light petroleum distillates, gasoline, medium petroleum distillates, kerosene and heavy petroleum distillates.

Another method to overcome the subjectivity of the analyst is to utilise the pattern recognition capabilities of artificial neural networks (ANNs). An ANN is loosely modelled on the architecture of the brain and learns by example [4]. Representative data is collected and delivered to the ANN, where training algorithms are called to learn the structure of the data. The user does not have to know any relationships between the inputs and the outputs, but follows a heuristic process to discover the underlying structure of the data. An ANN usually consists of neurons arranged in a layered topology. Fig. 1 shows an example of a typical ANN consisting of five inputs and three outputs, each represented by a neuron. A hidden layer of seven neurons is also present. The hidden and output neurons are all connected to the preceding layer. When the ANN is executed, it attempts to learn the structure of the data by a feed forward iterative process that continually adjusts the weights of each of the neurons in the hidden and output layer to minimise the error of the response surface. Each iteration is known as an epoch. After a suitable number of epochs, the neural network will arrive at a minimum error. At this point, the ANN may describe the response very well or very poorly, or somewhere in between. If the ANN does not adequately describe the response surface, the user may improve the performance by changing the architecture of the ANN.

The architecture of the ANN has a very large effect upon its ability to accurately model the response surface. The best architecture is usually determined by trial and error where the numbers of neurons in the hidden layer are varied. Increasing or decreasing the number of neurons in the hidden layer will often lead to an increase or decrease in the overall error of the response surface. If the user continually moves in the direction of minimising the error, a global minimum is often found. For example, increasing the number of neurons in the hidden layer of the ANN in Fig. 1 from seven to eight may decrease the overall error on the response surface. Adding another neuron may decrease the error further. The user continues to increase the number of neurons in the hidden layer until the error starts to increase, at which point the preceding architecture is deemed to be the most suitable.

Back propagation is the best known of the neural network training algorithms in which the gradient vector of the error surface is calculated. This vector points along the direction of the steepest descent. Moving along the vector a short distance will decrease the error. Repeating this process and moving along the vector in shorter distances will eventually find a minimum. Conjugate gradient is a more sophisticated training technique in which the line of descent directions are selected to maintain the second derivative of the error surface at zero. Conjugate gradient typically requires fewer epochs than back propagation, and usually converges to a lower minimum.

One major problem with these two types of training algorithms is the tendency of over-learning, or over-fitting. Over-learning can be illustrated by the following example. Consider a non-linear dataset containing two variables. One may wish to model the data by fitting a polynomial function as shown in Fig. 2. The dashed line represents a poor-fitting, low-order polynomial. The solid line represents a good-fitting, high-order polynomial. Although the solid line goes

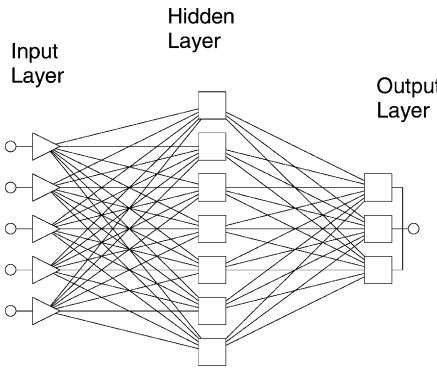

Fig. 1. Typical ANN architecture.

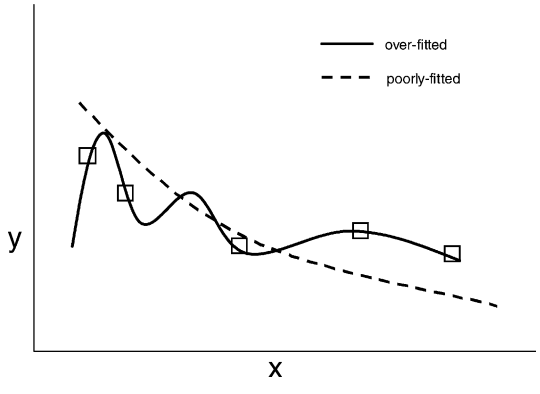

Fig. 2. Representation of over-learning.

through all the points, its predictive power is poor. The low-order polynomial does not fit the data well, but its predictive power is better than the high-order polynomial. An ideal solution would have a good fit and good predictive power. ANNs trained with back propagation and conjugate gradients can suffer from over-learning in the same way. The training data may fit well, but the description of the whole response surface may be poor. This problem is overcome by cross-verification in which some of the training set data is reserved for the purposes of verifying that the ANN is accurately modelling the response surface. As the learning progresses, the training error of the response surface will reduce. Assuming that the true error of the response surface is being modelled, the verification error will also reduce. If the training error still continues to decrease and the verification error remains large relative to the training error, or if the verification error starts to increase, it is clear that over-learning has occurred. Therefore, the key indicator of an ANN performance is the verification error. When the training error is significantly less than the verification error, over-learning has occurred. An ANN will best describe the response surface when the verification is similar in magnitude or less than the training error, assuming that that the overall error is sufficiently minimised.

Andrews and Lieberman analysed seven classes of petroleum-based products using laser-induced fluorescence spectroscopy [5]. They compared the classification capability of PCA of the fluorescence spectra against direct ANN analysis of the normalised fluorescence spectra. The authors conclude that an ANN performed considerably better than PCA. The ANNs had a correct classification rate of approximately 96% compared with PCA with approximately 90%.

Lavine et al. employed a genetic algorithm in combination with PCA to aid in identification of spilled aviation fuels [6]. The genetic algorithm was used to uncover features in each class of the fuels by sampling key features and scoring their principal components. This approach led to the identification of 20 GC peaks that produced principal components that showed significant clustering, which allowed classification of six types of jet fuels.

Ichikawa et al. analysed 31 premium and 30 regular grade automotive gasolines by GC–MS without separation of the components of the samples [7]. The resulting mass spectra were broken down into nine fragment ions related to aromatic, naphthenic, olefinic and branched paraffinic compounds. LDA was performed on these nine ion fragments, allowing complete discrimination of the samples. PCA was also performed which allowed discrimination based on scatter plots of the first two principal component scores.

This study presents a comparison of PCA and ANNs for the classification of liquid premium and regular gasolines from their GC–MS chromatograms. ANNs have been used in the past for many different types of pattern recognition problems, but this is the first report of applying ANNs for the classification of summer and winter, premium and regular grade gasolines from GC–MS chromatograms.

2. Experimental

The experimental data was taken from a Canadian Petroleum Products Institute Report of the composition of unleaded summer and winter gasolines in 1993 [8]. In this report, 44 samples of regular gasoline (22 winter, 22 summer), and 44 samples of premium gasoline (22 winter, 22 summer) were analysed by GC–MS. The gasolines were collected over the course of 1 year from different regions across the country. Forty-four compounds that may be present in automotive gasoline in concentrations of greater than 1% were reported. The percent peak area of each of the target 44 compounds was extracted from the report and is

Table 1
GC–MS percent peak area of target compounds [1]

Compound	Sample										
	1	2	3	4	5	6	7	8	9	10	11
Isobutane	2.77	3.53	4.17	1.92	2.31	3.37	2.00	0.28	2.65	2.67	2.85
Butane	6.91	4.71	6.19	6.51	3.59	5.47	11.34	12.79	6.19	5.48	5.39
c-2-Butene	0.52	0.28	0.17	0.51	0.71	0.00	0.04	0.04	0.02	0.35	0.22
2-Methyl-1-butene	0.45	0.58	0.20	0.62	0.57	0.01	0.18	0.67	0.02	0.62	0.63
t-2-Pentene	0.55	0.77	0.27	0.78	0.80	0.00	0.22	0.86	0.02	0.72	0.73
2-Methyl-2-butene	0.85	1.19	0.46	1.15	1.03	0.01	0.34	1.35	0.04	1.14	1.18
2,2-Dimethylbutane	0.31	0.12	0.67	0.20	0.28	0.74	0.20	0.05	0.31	1.15	0.19
2-Methylpentane	2.21	4.03	4.18	2.16	2.66	5.26	1.49	2.15	6.49	3.86	3.18
Hexane	1.52	2.20	3.65	1.50	2.01	2.05	0.62	1.08	2.11	2.46	2.07
2,4-Dimethylpentane	0.77	0.51	0.32	1.14	0.38	0.40	5.91	2.77	0.49	0.36	0.33
Cyclohexane	0.07	0.20	0.17	0.11	0.11	0.61	0.18	0.15	0.64	0.40	0.27
3-Methylhexane	1.87	2.02	1.50	1.77	1.97	2.35	0.76	0.85	1.43	1.59	1.20
Heptane	1.24	1.16	0.99	1.20	1.28	1.72	0.53	0.35	0.64	1.01	0.54
2,4-Dimethylhexane	0.59	0.26	0.10	0.60	0.22	0.23	1.73	1.08	0.07	0.17	0.16

Table 1 (*Continued*)

Compound	Sample										
	1	2	3	4	5	6	7	8	9	10	11
Toluene	11.56	10.68	17.94	11.58	14.81	9.01	1.32	2.68	7.99	11.18	10.46
2,3-Dimethylhexane	0.64	0.15	0.07	0.65	0.14	0.19	1.40	1.11	0.05	0.11	0.12
Ethylbenzene	2.07	1.80	1.91	2.01	2.49	1.09	0.36	0.64	1.65	2.07	1.66
p-Xylene	2.79	2.53	2.50	2.36	3.29	2.20	0.39	0.72	2.37	2.62	1.66
1-Methyl-3-ethylbenzene	1.95	1.97	1.66	1.16	1.52	3.75	0.43	0.81	3.12	2.23	2.97
1,3,5-Trimethylbenzene	1.16	1.31	1.03	0.67	0.99	2.21	0.27	0.47	1.67	1.27	1.49
1,2,4-Trimethylbenzene	3.33	3.34	2.82	2.15	2.67	7.42	0.80	1.31	6.13	3.33	5.24
2-Methylnaphthalene	0.07	0.08	0.13	0.02	0.24	0.16	0.13	0.03	0.04	0.08	0.18
1,3-Butadiene	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
t-2-Butene	0.66	0.41	0.25	0.55	0.87	0.00	0.05	0.03	0.03	0.37	0.23
Isopentane	4.82	10.20	6.77	5.02	5.12	10.79	5.10	5.64	14.81	8.22	7.25
Pentane	1.76	4.21	5.31	2.27	2.84	2.27	0.97	1.86	3.55	3.31	4.26
c-2-Pentene	0.31	0.42	0.15	0.44	0.45	0.00	0.12	0.48	0.01	0.40	0.41
Cyclopentane	0.10	0.37	0.32	0.15	0.16	0.61	0.14	0.23	0.77	0.36	0.37
2,3-Dimethylbutane	1.05	0.79	0.84	1.14	0.54	0.93	2.00	1.04	1.27	0.98	0.58
3-Methylpentane	1.57	2.56	3.00	1.50	1.92	3.35	0.92	1.33	3.80	2.61	2.00
Methylcyclopentane	0.52	2.01	0.99	0.99	0.96	1.66	0.63	1.15	5.68	1.30	1.54
Benzene	2.10	2.33	5.62	2.66	3.31	2.68	0.34	0.56	2.82	3.10	0.62
2-Methylhexane	2.47	2.47	1.76	3.49	2.31	2.71	11.12	8.04	2.04	1.90	1.59
2,2,4-Trimethylpentane	3.71	0.23	0.00	3.93	0.00	0.00	17.09	13.74	0.00	0.00	0.00
Methylcyclohexane	0.26	0.62	0.21	0.38	0.41	0.14	0.46	0.46	0.29	0.50	0.55
2,3,4-Trimethylpentane	1.74	0.04	0.02	2.21	0.03	0.00	5.12	4.29	0.00	0.03	0.00
2,3,3-Trimethylpentane	1.50	0.00	0.00	1.57	0.00	0.00	3.80	2.29	0.00	0.00	0.00
3-Methylheptane	0.65	0.67	0.29	0.69	0.71	0.67	0.44	0.55	0.17	0.69	0.52
m-Xylene	6.44	5.68	5.79	5.84	8.23	5.39	1.04	1.99	5.26	6.07	4.26
o-Xylene	3.43	2.82	3.20	3.07	3.95	4.80	0.64	1.12	4.55	3.01	2.44
1-Methyl-4-ethylbenzene	0.86	0.90	0.71	0.51	0.64	1.72	0.18	0.32	1.43	0.99	1.35
1-Methyl-2-ethylbenzene	0.69	0.59	0.61	0.39	0.57	1.40	0.17	0.32	1.19	0.76	1.14
Naphthalene	0.14	0.17	0.20	0.04	0.20	0.20	0.08	0.16	0.10	0.19	0.27
1-Methylnaphthalene	0.02	0.04	0.05	0.01	0.08	0.08	0.04	0.01	0.02	0.04	0.08
Type	PU	PU	PU	PU	PU	PU	PU	PU	PU	PU	PU
	12	13	14	15	16	17	18	19	20	21	22
Isobutane	0.57	1.84	1.78	2.45	3.07	4.12	3.43	2.75	3.59	2.43	0.29
Butane	11.01	8.85	8.63	4.76	4.91	4.55	4.33	3.55	5.35	6.54	13.47
c-2-Butene	0.09	0.14	0.39	0.68	0.60	0.34	0.38	0.99	0.62	0.89	0.04
2-Methyl-1-butene	0.42	0.46	0.46	0.89	0.62	0.38	0.34	0.97	0.85	0.71	0.68
t-2-Pentene	0.52	0.70	0.59	1.14	0.80	0.54	0.43	1.24	1.05	0.91	0.87
2-Methyl-2-butene	0.82	1.07	0.93	1.66	1.16	0.74	0.68	1.85	1.75	1.28	1.35
2,2-Dimethylbutane	0.27	0.59	0.10	0.21	0.45	0.79	0.88	0.46	0.37	0.14	0.05
2-Methylpentane	2.66	3.12	2.06	3.68	2.83	3.06	3.98	3.79	3.09	2.87	2.08
Hexane	1.16	1.37	0.74	2.58	1.90	1.60	3.06	2.47	2.23	2.15	1.08
2,4-Dimethylpentane	0.73	1.05	1.21	0.77	0.53	0.44	0.37	0.41	0.40	0.35	2.61
Cyclohexane	0.05	0.16	0.08	0.14	0.17	0.48	0.47	0.20	0.33	0.28	0.16
3-Methylhexane	0.90	0.80	1.28	2.30	2.70	2.19	1.88	1.61	1.68	1.70	0.77
Heptane	0.62	0.52	0.75	1.55	1.92	1.31	1.27	0.96	1.13	1.28	0.32
2,4-Dimethylhexane	1.08	0.64	1.10	0.63	0.45	0.25	0.20	0.13	0.18	0.27	0.97
Toluene	11.08	7.57	4.28	4.57	14.54	14.29	14.92	14.80	13.08	10.93	2.36
2,3-Dimethylhexane	0.99	0.61	1.17	0.70	0.36	0.16	0.17	0.09	0.12	0.26	0.99
Ethylbenzene	2.14	2.50	0.94	2.42	2.98	2.80	3.20	2.63	2.42	2.39	0.55
p-Xylene	2.63	1.10	1.09	2.19	2.67	2.96	3.14	3.04	2.81	2.55	0.61
1-Methyl-3-ethylbenzene	1.60	2.60	1.12	2.55	1.81	1.77	1.99	1.82	1.67	2.41	0.66
1,3,5-Trimethylbenzene	0.94	1.33	0.66	1.10	0.79	1.00	0.89	0.90	0.90	1.13	0.37
1,2,4-Trimethylbenzene	2.73	4.05	1.98	3.27	2.69	2.58	2.78	2.37	2.45	4.22	1.08
2-Methylnaphthalene	0.08	0.36	0.11	0.03	0.03	0.02	0.08	0.04	0.05	0.06	0.02
1,3-Butadiene	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.01	0.02	0.00

Table 1 (*Continued*)

Compound	Sample										
	12	13	14	15	16	17	18	19	20	21	22
t-2-Butene	0.09	0.16	0.46	0.79	0.76	0.43	0.50	1.21	0.76	1.15	0.02
Isopentane	7.20	7.02	7.38	6.76	4.67	6.46	6.88	6.78	6.77	5.78	5.68
Pentane	3.87	3.97	1.11	2.63	1.92	2.44	3.96	3.06	3.91	3.20	1.88
c-2-Pentene	0.29	0.28	0.33	0.63	0.45	0.30	0.24	0.69	0.59	0.50	0.48
Cyclopentane	0.46	0.76	0.11	0.36	0.24	0.24	0.43	0.33	0.28	0.35	0.22
2,3-Dimethylbutane	1.01	1.10	1.54	1.13	0.66	0.75	0.93	0.83	0.67	0.56	0.99
3-Methylpentane	1.55	1.75	1.33	2.41	1.98	2.07	2.74	2.60	2.13	1.81	1.27
Methylcyclopentane	0.55	0.87	0.95	1.72	1.42	1.40	1.35	1.52	1.24	1.72	1.12
Benzene	0.35	0.44	0.78	1.53	3.67	3.05	3.98	4.75	3.01	1.70	0.52
2-Methylhexane	1.36	2.80	2.43	2.90	3.16	2.64	2.26	1.98	2.04	2.02	7.37
2,2,4-Trimethylpentane	8.41	4.80	9.13	4.06	0.58	0.19	0.00	0.00	0.00	0.00	12.74
Methylcyclohexane	0.21	0.30	0.49	0.29	0.39	0.60	0.44	0.41	0.52	0.36	0.42
2,3,4-Trimethylpentane	3.25	1.76	4.55	2.21	0.16	0.04	0.03	0.01	0.00	0.05	3.85
2,3,3-Trimethylpentane	1.43	0.66	4.96	1.90	0.00	0.00	0.00	0.00	0.00	0.00	2.08
3-Methylheptane	0.60	0.38	0.78	0.76	1.08	0.81	0.60	0.39	0.58	0.94	0.47
m-Xylene	6.08	2.83	2.83	5.39	6.34	7.05	7.46	6.94	6.76	6.37	1.70
o-Xylene	3.27	1.46	1.57	3.13	3.52	3.71	4.00	3.51	3.47	3.69	0.95
1-Methyl-4-ethylbenzene	0.69	1.18	0.47	1.16	0.81	0.76	0.89	0.81	0.71	1.10	0.26
1-Methyl-2-ethylbenzene	0.59	0.95	0.47	0.87	0.71	0.65	0.69	0.58	0.62	1.01	0.26
Naphthalene	0.17	0.51	0.22	0.06	0.07	0.05	0.12	0.05	0.07	0.15	0.10
1-Methylnaphthalene	0.03	0.16	0.04	0.01	0.01	0.01	0.04	0.02	0.02	0.02	0.01
Type	PU	PU	PU	PU	PU	PU	PU	PU	PU	PU	PU
	23	24	25	26	27	28	29	30	31	32	33
Isobutane	2.12	2.58	3.42	1.83	2.77	2.67	2.12	0.36	3.33	3.48	2.82
Butane	5.43	3.70	5.11	7.59	5.12	5.47	9.06	12.99	7.30	4.09	5.79
c-2-Butene	0.66	0.21	0.41	0.49	0.88	0.00	0.12	0.03	0.02	0.52	0.27
2-Methyl-1-butene	0.70	0.39	0.59	0.75	0.67	0.00	0.55	0.55	0.02	0.76	0.79
t-2-Pentene	0.87	0.52	0.84	0.98	0.94	0.00	0.68	0.71	0.02	0.89	0.89
2-Methyl-2-butene	1.32	0.80	1.34	1.39	1.20	0.01	1.02	1.17	0.05	1.40	1.45
2,2-Dimethylbutane	0.25	0.15	0.23	0.19	0.83	1.09	0.59	0.14	0.69	1.49	0.35
2-Methylpentane	3.99	7.00	3.39	3.67	5.03	6.80	3.60	3.03	6.55	3.97	5.35
Hexane	4.42	3.90	2.70	3.29	5.19	4.33	1.63	3.15	3.12	2.05	3.92
2,4-Dimethylpentane	0.43	0.45	0.37	0.81	0.38	0.43	1.24	1.32	0.57	0.45	0.51
Cyclohexane	1.11	0.67	0.35	0.67	1.12	0.75	0.55	0.41	0.41	0.60	0.47
3-Methylhexane	1.74	1.19	0.93	1.71	1.61	2.49	1.73	1.83	2.57	1.72	1.92
Heptane	1.51	0.58	0.44	1.25	1.24	1.83	1.52	1.69	1.58	1.09	1.05
2,4-Dimethylhexane	0.20	0.15	0.14	0.49	0.16	0.24	0.51	0.67	0.20	0.25	0.19
Toluene	7.42	4.73	5.32	6.84	7.20	9.31	4.82	6.67	6.28	6.24	2.41
2,3-Dimethylhexane	0.15	0.10	0.10	0.50	0.10	0.19	0.44	0.64	0.17	0.17	0.13
Ethylbenzene	1.39	1.07	0.70	1.33	1.30	0.97	1.11	1.27	1.25	1.43	1.10
p-Xylene	1.76	1.41	0.74	1.52	1.65	1.88	1.19	1.43	1.82	1.66	1.22
1-Methyl-3-ethylbenzene	1.50	1.45	1.05	1.26	1.02	1.70	1.30	0.98	2.52	1.67	2.14
1,3,5-Trimethylbenzene	0.88	0.94	0.64	0.72	0.62	1.13	0.75	0.52	1.28	0.98	1.09
1,2,4-Trimethylbenzene	2.51	2.39	1.62	2.16	1.64	3.42	2.22	1.81	4.83	2.23	3.76
2-Methylnaphthalene	0.09	0.13	0.33	0.02	0.21	0.14	0.36	0.09	0.03	0.02	0.11
1,3-Butadiene	0.02	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
t-2-Butene	0.75	0.31	0.53	0.51	1.08	0.00	0.11	0.02	0.03	0.59	0.28
Isopentane	6.70	14.14	7.80	6.05	7.88	14.25	8.13	5.33	13.65	10.07	9.97
Pentane	5.47	5.59	5.97	4.11	6.26	7.69	2.33	3.85	7.09	3.28	6.24
c-2-Pentene	0.49	0.28	0.47	0.55	0.52	0.00	0.38	0.39	0.01	0.49	0.50
Cyclopentane	0.58	0.83	0.53	0.42	0.66	0.74	0.43	0.46	1.26	0.42	0.59
2,3-Dimethylbutane	0.71	1.26	0.64	1.03	1.04	1.33	1.14	0.73	1.27	1.10	0.95
3-Methylpentane	2.52	4.05	2.10	2.36	3.29	4.36	2.38	1.92	3.99	2.51	3.45
Methylcyclopentane	2.29	4.92	1.63	2.00	2.76	2.11	1.93	1.97	4.21	1.73	2.49
Benzene	1.96	1.88	1.68	1.97	2.19	2.97	1.08	1.53	1.59	1.47	0.76

Table 1 (*Continued*)

Compound	Sample										
	23	24	25	26	27	28	29	30	31	32	33
2-Methylhexane	2.32	1.64	1.29	2.73	2.14	2.88	3.75	4.71	3.17	2.16	2.51
2,2,4-Trimethylpentane	0.37	0.13	0.29	2.58	0.00	0.00	2.69	5.18	0.00	0.34	0.00
Methylcyclohexane	0.86	0.58	0.62	0.79	0.91	0.18	1.37	0.46	0.19	0.69	0.53
2,3,4-Trimethylpentane	0.20	0.00	0.00	1.48	0.03	0.01	0.84	1.67	0.00	0.00	0.00
2,3,3-Trimethylpentane	0.00	0.00	0.00	1.16	0.00	0.00	0.40	0.71	0.00	0.00	0.00
3-Methylheptane	0.52	0.70	0.62	0.69	0.67	0.70	1.21	0.99	0.54	0.83	0.60
m-Xylene	4.38	3.44	1.96	3.88	4.14	4.40	3.12	3.70	4.04	3.90	3.06
o-Xylene	2.27	1.74	1.02	2.06	2.01	2.80	1.71	2.02	3.73	1.92	1.76
1-Methyl-4-ethylbenzene	0.65	0.61	0.42	0.51	0.41	0.78	0.56	0.41	1.15	0.73	0.95
1-Methyl-2-ethylbenzene	0.61	0.56	0.51	0.52	0.41	0.61	0.56	0.42	0.96	0.62	0.80
Naphthalene	0.17	0.20	0.35	0.05	0.19	0.13	0.25	0.29	0.07	0.05	0.18
1-Methylnaphthalene	0.03	0.05	0.10	0.01	0.07	0.07	0.13	0.03	0.02	0.01	0.05
Type	RU	RU	RU	RU	RU	RU	RU	RU	RU	RU	RU
	34	35	36	37	38	39	40	41	42	43	44
Isobutane	1.15	2.07	2.01	2.71	2.29	3.84	2.83	3.07	4.44	1.27	0.34
Butane	9.64	8.04	7.33	3.41	4.62	4.68	5.22	5.03	4.82	7.65	12.21
c-2-Butene	0.06	0.20	0.50	0.53	0.98	0.49	0.52	0.42	0.39	0.64	0.03
2-Methyl-1-butene	0.34	0.58	0.66	0.47	0.84	0.63	0.61	0.40	0.50	0.78	0.53
t-2-Pentene	0.47	0.79	0.84	0.61	1.09	0.83	0.88	0.51	0.73	0.99	0.69
2-Methyl-2-butene	0.77	1.23	1.33	0.94	1.62	1.44	1.25	0.74	0.96	1.42	1.13
2,2-Dimethylbutane	0.48	0.46	0.82	1.24	0.28	0.29	0.17	0.89	1.88	0.17	0.14
2-Methylpentane	4.56	4.26	3.56	4.81	4.02	3.40	5.59	4.25	4.44	3.27	2.85
Hexane	2.48	3.82	2.13	3.04	3.04	2.95	4.25	3.03	1.81	2.25	2.86
2,4-Dimethylpentane	0.53	0.67	0.45	0.40	0.44	0.43	0.35	0.44	0.45	0.58	1.20
Cyclohexane	0.09	0.38	0.38	0.86	0.64	0.78	0.12	0.60	1.30	0.28	0.39
3-Methylhexane	1.68	2.15	2.02	1.86	1.55	1.44	1.47	2.09	2.07	1.61	1.74
Heptane	1.10	1.50	1.37	1.24	0.75	1.11	0.83	1.50	1.17	0.84	1.51
2,4-Dimethylhexane	0.49	0.40	0.36	0.26	0.24	0.18	0.22	0.36	0.32	0.39	0.62
Toluene	7.14	4.38	7.00	8.90	9.04	8.01	3.46	10.72	6.54	4.78	6.61
2,3-Dimethylhexane	0.42	0.36	0.31	0.22	0.15	0.12	0.16	0.31	0.21	0.40	0.59
Ethylbenzene	1.75	1.55	1.50	2.03	1.80	1.62	1.48	2.37	1.43	1.06	1.28
p-Xylene	2.01	1.09	1.70	2.03	1.96	1.82	1.48	2.14	1.42	1.07	1.43
1-Methyl-3-ethylbenzene	1.57	2.26	1.71	1.44	0.97	1.53	1.90	1.71	1.13	1.45	1.00
1,3,5-Trimethylbenzene	0.93	1.06	1.01	0.69	0.48	0.85	0.95	0.78	0.68	0.76	0.50
1,2,4-Trimethylbenzene	2.62	3.24	2.99	2.04	1.11	2.19	2.78	2.47	1.74	2.25	1.72
2-Methylnaphthalene	0.05	0.25	0.14	0.24	0.02	0.04	0.33	0.05	0.21	0.30	0.10
1,3-Butadiene	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.00
t-2-Butene	0.06	0.22	0.58	0.69	1.23	0.61	0.62	0.53	0.46	0.75	0.02
Isopentane	8.59	8.77	7.24	9.50	7.21	5.99	10.02	7.39	9.69	7.13	4.99
Pentane	6.31	4.92	2.83	4.87	5.05	4.42	6.19	4.45	2.90	3.41	3.46
c-2-Pentene	0.27	0.36	0.47	0.34	0.60	0.47	0.44	0.29	0.40	0.55	0.39
Cyclopentane	0.77	0.79	0.35	0.60	0.67	0.46	0.83	0.63	0.46	0.37	0.42
2,3-Dimethylbutane	1.03	0.95	0.86	1.25	0.78	0.71	1.02	0.99	1.24	0.92	0.69
3-Methylpentane	2.85	3.09	2.28	3.11	2.57	2.31	3.52	2.82	2.77	2.08	1.82
Methylcyclopentane	1.11	2.16	1.76	1.91	2.15	1.82	2.21	2.62	2.30	1.97	1.85
Benzene	0.61	0.64	1.42	2.28	3.07	2.34	1.28	2.62	1.41	1.67	1.47
2-Methylhexane	2.11	3.12	2.59	2.27	2.06	1.89	1.81	2.56	2.58	2.17	4.39
2,2,4-Trimethylpentane	2.68	1.44	0.00	0.00	0.00	0.15	0.00	0.38	0.00	2.05	4.62
Methylcyclohexane	0.45	0.58	0.78	0.79	1.02	1.09	0.50	0.50	0.93	0.72	0.46
2,3,4-Trimethylpentane	1.06	0.54	0.10	0.03	0.00	0.06	0.05	0.13	0.00	1.20	1.47
2,3,3-Trimethylpentane	0.50	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.01	0.55
3-Methylheptane	0.65	0.78	1.17	1.04	1.04	0.77	0.74	0.97	1.10	0.84	0.91
m-Xylene	4.80	2.69	4.48	4.67	4.67	4.51	3.70	5.12	3.61	2.94	3.72
o-Xylene	2.52	1.44	2.33	2.58	2.36	2.33	2.04	2.88	1.88	1.62	2.02
1-Methyl-4-ethylbenzene	0.66	1.01	0.72	0.63	0.40	0.63	0.80	0.74	0.46	0.59	0.42

Table 1 (*Continued*)

Compound	Sample										
	34	35	36	37	38	39	40	41	42	43	44
1-Methyl-2-ethylbenzene	0.60	0.86	0.74	0.60	0.32	0.67	0.73	0.73	0.51	0.63	0.42
Naphthalene	0.31	0.44	0.27	0.17	0.02	0.07	0.44	0.09	0.36	0.51	0.25
1-Methylnaphthalene	0.02	0.11	0.06	0.09	0.01	0.02	0.12	0.02	0.06	0.09	0.03
Type	RU	RU	RU	RU	RU	RU	RU	RU	RU	RU	RU
	45	46	47	48	49	50	51	52	53	54	55
Isobutane	0.51	0.35	0.57	0.33	1.19	0.85	1.09	0.17	0.68	0.86	1.43
Butane	2.47	2.86	2.32	2.11	2.43	3.11	7.34	7.18	3.45	2.38	3.25
c-2-Butene	0.14	0.14	0.08	0.08	0.37	0.00	0.02	0.03	0.01	0.24	0.06
2-Methyl-1-butene	0.73	0.54	0.26	0.44	0.52	0.01	0.06	0.50	0.02	0.63	0.59
t-2-Pentene	1.10	0.75	0.45	0.62	0.71	0.00	0.07	0.70	0.03	0.75	0.73
2-Methyl-2-butene	1.72	1.09	0.61	0.88	1.01	0.01	0.11	0.99	0.05	1.16	1.10
2,2-Dimethylbutane	0.38	0.18	0.55	0.17	0.30	0.85	0.33	0.09	0.56	1.16	0.25
2-Methylpentane	3.48	2.37	3.34	2.14	2.79	5.66	1.83	1.80	6.31	4.19	3.49
Hexane	2.66	1.38	2.11	1.22	2.11	2.54	0.56	0.81	3.12	2.68	1.86
2,4-Dimethylpentane	0.49	1.76	1.16	2.09	0.63	0.40	4.59	2.83	0.60	0.38	0.37
Cyclohexane	0.44	0.11	0.33	0.09	0.08	0.62	0.13	0.07	0.50	0.49	0.16
3-Methylhexane	1.36	2.06	1.44	1.98	2.13	2.44	0.94	1.07	2.75	1.77	1.52
Heptane	1.25	1.32	0.92	1.28	1.40	1.78	0.82	0.57	1.59	1.14	0.82
2,4-Dimethylhexane	0.33	0.58	0.61	0.67	0.30	0.23	1.63	1.17	0.30	0.17	0.20
Toluene	12.29	12.63	10.05	10.58	14.36	9.59	2.71	3.95	3.26	12.85	19.84
2,3-Dimethylhexane	0.30	0.60	0.50	0.69	0.25	0.20	1.57	1.21	0.19	0.11	0.15
Ethylbenzene	2.38	2.00	1.95	1.76	2.59	1.16	0.57	0.86	1.10	2.27	1.20
p-Xylene	2.78	2.55	2.31	2.06	3.40	2.19	0.59	0.96	1.70	2.89	1.43
1-Methyl-3-ethylbenzene	2.34	1.68	2.71	1.41	2.02	3.51	0.58	1.04	3.74	2.29	1.89
1,3,5-Trimethylbenzene	1.23	1.01	1.40	0.77	1.17	1.96	0.33	0.57	1.84	1.30	1.01
1,2,4-Trimethylbenzene	3.81	3.06	4.27	2.48	3.23	6.63	1.01	1.66	6.37	3.41	3.40
2-Methylnaphthalene	0.21	0.11	0.35	0.13	0.28	0.23	0.05	0.18	0.09	0.07	0.30
1,3-Butadiene	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
t-2-Butene	0.14	0.15	0.08	0.09	0.43	0.00	0.02	0.02	0.01	0.25	0.06
Isopentane	7.09	6.33	7.94	5.88	4.92	11.94	6.05	4.89	11.83	9.01	6.61
Pentane	3.86	2.62	3.96	2.42	2.69	3.30	0.53	0.91	6.71	3.93	3.68
c-2-Pentene	0.62	0.42	0.25	0.34	0.40	0.00	0.04	0.39	0.01	0.42	0.41
Cyclopentane	0.43	0.16	0.41	0.13	0.16	0.70	0.20	0.11	1.05	0.40	0.33
2,3-Dimethylbutane	0.91	1.31	1.31	1.55	0.63	0.99	1.78	1.06	1.18	1.05	0.66
3-Methylpentane	2.24	1.60	2.21	1.46	2.03	3.56	1.16	1.17	3.83	2.84	2.23
Methylcyclopentane	1.55	1.06	1.32	0.89	0.92	1.81	0.76	0.84	4.37	1.47	1.38
Benzene	1.65	2.19	1.96	2.05	3.37	3.16	0.69	0.88	1.32	3.57	0.41
2-Methylhexane	1.75	5.00	3.51	5.87	3.07	2.82	13.16	8.56	3.54	2.13	1.93
2,2,4-Trimethylpentane	1.58	2.34	2.88	3.19	1.07	0.00	18.61	13.87	0.00	0.00	0.00
Methylcyclohexane	0.00	0.42	0.29	0.39	0.37	0.14	0.27	0.40	0.84	0.54	0.47
2,3,4-Trimethylpentane	0.61	1.40	1.44	1.95	0.25	0.01	6.09	4.54	0.04	0.03	0.06
2,3,3-Trimethylpentane	0.00	0.00	1.09	1.05	0.00	0.00	3.49	2.41	0.00	0.00	0.00
3-Methylheptane	0.64	0.74	0.48	0.80	0.76	0.70	0.47	0.55	0.74	0.57	0.65
m-Xylene	6.46	6.06	5.60	5.18	8.21	5.40	1.41	2.47	3.90	6.59	3.49
o-Xylene	3.63	3.29	3.11	2.82	3.96	3.97	0.94	1.41	3.37	3.26	2.00
1-Methyl-4-ethylbenzene	1.05	0.75	1.23	0.62	0.87	1.63	0.26	0.43	1.75	1.02	0.84
1-Methyl-2-ethylbenzene	0.93	0.59	0.96	0.50	0.69	1.32	0.25	0.44	1.37	0.74	0.79
Naphthalene	0.34	0.15	0.49	0.12	0.21	0.29	0.06	0.41	0.29	0.15	0.32
1-Methylnaphthalene	0.09	0.05	0.14	0.06	0.11	0.11	0.02	0.05	0.05	0.03	0.15
Type	PU	PU	PU	PU	PU	PU	PU	PU	PU	PU	PU
	56	57	58	59	60	61	62	63	64	65	66
Isobutane	0.45	1.05	1.06	1.08	1.11	1.91	0.82	0.90	1.47	0.74	1.00
Butane	3.88	3.92	3.25	2.63	2.60	2.83	4.75	1.97	2.33	4.91	7.39
c-2-Butene	0.08	0.09	0.38	0.47	0.46	0.23	0.10	0.72	0.37	0.17	0.02

Table 1 (*Continued*)

Compound	Sample										
	56	57	58	59	60	61	62	63	64	65	66
2-Methyl-1-butene	0.37	0.66	0.48	0.93	1.08	0.37	0.34	0.90	0.57	0.04	0.08
t-2-Pentene	0.47	1.08	0.68	1.10	1.37	0.54	0.46	1.16	0.74	0.05	0.10
2-Methyl-2-butene	0.73	1.58	0.97	1.69	2.06	0.75	0.72	1.75	1.23	0.07	0.15
2,2-Dimethylbutane	0.31	0.39	0.15	0.24	0.29	0.39	0.74	0.39	0.22	0.40	0.30
2-Methylpentane	2.79	3.43	2.21	3.69	3.48	2.55	3.03	4.12	2.71	2.50	1.76
Hexane	0.74	1.66	0.84	2.10	1.79	1.66	1.45	2.68	2.15	2.10	0.53
2,4-Dimethylpentane	0.62	0.96	1.29	0.90	0.66	0.41	3.26	0.50	0.39	0.93	4.40
Cyclohexane	0.06	0.23	0.10	0.13	0.15	0.09	0.51	0.17	0.39	0.04	0.12
3-Methylhexane	0.70	1.10	1.28	1.85	2.33	1.93	1.09	2.19	1.87	1.56	0.86
Heptane	0.40	0.69	0.72	1.24	1.44	1.18	0.60	1.34	1.38	1.31	0.73
2,4-Dimethylhexane	0.86	0.60	1.12	0.71	0.45	0.23	1.23	6.26	0.33	0.85	1.51
Toluene	10.59	8.89	4.10	2.07	11.77	13.73	3.94	12.04	11.97	8.69	2.55
2,3-Dimethylhexane	0.75	0.61	1.18	0.69	0.38	0.16	1.21	0.20	0.23	0.99	1.45
Ethylbenzene	1.91	2.23	0.99	2.88	2.47	2.53	0.80	2.32	3.53	1.82	0.51
p-Xylene	2.30	1.17	1.05	2.61	2.27	2.86	0.97	2.66	3.09	1.80	0.51
1-Methyl-3-ethylbenzene	2.05	2.16	1.18	2.66	1.97	2.00	0.68	2.04	3.04	1.84	0.52
1,3,5-Trimethylbenzene	1.18	1.03	0.66	1.30	0.85	1.18	0.38	1.14	1.34	0.91	0.30
1,2,4-Trimethylbenzene	3.36	3.11	1.90	3.97	2.79	3.03	1.04	2.99	4.04	3.60	0.91
2-Methylnaphthalene	0.41	0.27	0.19	0.28	0.10	0.37	0.09	0.20	0.10	0.15	0.06
1,3-Butadiene	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00
t-2-Butene	0.08	0.09	0.44	0.51	0.50	0.27	0.09	0.81	0.43	0.32	0.02
Isopentane	9.21	7.08	7.26	6.20	6.55	4.33	8.97	6.96	5.40	4.57	5.91
Pentane	6.32	4.88	1.21	2.63	2.40	1.71	2.35	2.85	3.33	3.40	0.57
c-2-Pentene	0.26	0.43	0.38	0.61	0.76	0.30	0.25	0.64	0.42	0.03	0.05
Cyclopentane	0.77	1.08	0.14	0.24	0.22	0.16	0.30	0.29	0.32	0.37	0.19
2,3-Dimethylbutane	0.98	1.02	1.53	1.34	0.81	0.56	2.59	0.90	0.55	1.28	1.72
3-Methylpentane	1.33	1.97	1.43	2.38	2.36	1.83	1.90	2.88	1.83	1.66	1.11
Methylcyclopentane	0.70	1.23	1.05	1.30	1.62	1.02	1.21	1.74	1.31	0.80	0.72
Benzene	0.40	0.36	0.88	0.74	3.13	3.65	1.00	3.68	1.76	2.71	0.64
2-Methylhexane	1.11	2.85	2.62	2.48	2.89	2.34	5.81	2.66	2.22	2.55	12.55
2,2,4-Trimethylpentane	7.12	4.17	11.03	3.48	1.23	0.33	10.49	0.00	0.00	7.21	17.55
Methylcyclohexane	0.32	0.37	0.56	0.32	0.47	0.46	0.62	0.55	0.47	0.10	0.27
2,3,4-Trimethylpentane	2.83	1.55	5.32	1.73	0.38	0.05	4.85	0.03	0.03	4.16	5.68
2,3,3-Trimethylpentane	1.09	0.68	6.00	1.61	0.00	0.00	5.53	0.00	0.00	3.96	3.13
3-Methylheptane	0.54	0.37	0.90	0.65	0.78	0.76	0.57	0.75	0.98	0.79	0.42
m-Xylene	5.43	2.95	2.72	6.21	5.33	6.92	2.27	6.47	7.56	4.42	1.29
o-Xylene	2.90	1.44	1.49	3.46	2.97	3.61	1.17	3.21	4.23	2.87	0.79
1-Methyl-4-ethylbenzene	0.90	0.96	0.49	1.20	0.89	0.86	0.30	0.92	1.32	0.84	0.23
1-Methyl-2-ethylbenzene	0.77	0.79	0.50	0.94	0.78	0.78	0.23	0.64	1.37	0.89	0.23
Naphthalene	0.40	0.47	0.29	0.34	0.19	0.32	0.09	0.29	0.33	0.36	0.06
1-Methylnaphthalene	0.15	0.12	0.06	0.13	0.05	0.13	0.03	0.09	0.04	0.07	0.02
Type	PU	PU	PU	PU	PU	PU	PU	PU	PU	PU	PU
	67	68	69	70	71	72	73	74	75	76	77
Isobutane	0.40	0.55	0.46	0.45	0.89	0.54	0.92	0.30	0.90	1.27	1.07
Butane	1.77	2.00	1.52	1.66	1.18	2.09	4.60	5.99	1.97	1.72	2.50
c-2-Butene	0.30	0.10	0.15	0.11	0.74	0.00	0.12	0.02	0.72	0.33	0.04
2-Methyl-1-butene	0.62	0.41	0.55	0.62	1.01	0.01	0.51	0.42	0.90	0.80	0.57
t-2-Pentene	1.02	0.58	0.97	0.86	1.51	0.01	0.64	0.58	1.16	0.97	0.71
2-Methyl-2-butene	1.41	0.86	1.36	1.24	1.97	0.02	0.95	0.85	1.75	1.51	1.09
2,2-Dimethylbutane	0.35	0.27	0.36	0.16	0.60	1.14	0.62	0.20	0.39	1.22	0.43
2-Methylpentane	4.03	5.29	4.20	3.99	3.53	6.82	4.05	3.23	4.12	4.16	5.66
Hexane	3.35	3.28	3.26	4.15	1.94	4.26	1.44	3.03	2.68	2.21	2.94
2,4-Dimethylpentane	0.51	0.63	0.49	0.63	0.42	0.41	1.22	1.05	0.50	0.45	0.53
Cyclohexane	0.62	0.55	0.50	0.61	0.26	0.78	0.30	0.37	0.17	0.62	0.28
3-Methylhexane	1.40	1.87	1.17	1.48	1.61	2.49	1.99	2.02	2.19	1.79	2.39

Table 1 (*Continued*)

Compound	Sample										
	67	68	69	70	71	72	73	74	75	76	77
Heptane	1.01	0.93	0.65	0.81	0.77	1.81	1.65	1.53	1.34	1.17	1.42
2,4-Dimethylhexane	0.28	0.22	0.24	0.27	0.25	0.23	0.58	0.62	6.26	0.25	0.28
Toluene	6.60	6.29	6.43	6.04	5.91	9.57	5.55	7.84	12.04	6.90	3.48
2,3-Dimethylhexane	0.27	0.20	0.19	0.26	0.16	0.20	0.55	0.60	0.20	0.16	0.19
Ethylbenzene	1.31	1.37	1.08	1.19	1.32	1.04	1.31	1.58	2.32	1.55	1.51
p-Xylene	1.57	1.74	1.18	1.40	1.52	1.88	1.35	1.73	2.66	1.74	1.93
1-Methyl-3-ethylbenzene	1.66	2.31	1.58	1.57	1.54	2.01	1.51	1.70	2.04	1.78	1.98
1,3,5-Trimethylbenzene	0.95	1.20	0.87	0.94	0.86	1.24	0.84	0.88	1.14	1.00	1.09
1,2,4-Trimethylbenzene	2.83	4.06	2.72	2.67	2.26	3.88	2.62	2.82	2.99	2.48	3.49
2-Methylnaphthalene	0.36	0.22	0.37	0.37	0.52	0.23	0.34	0.25	0.20	0.08	0.26
1,3-Butadiene	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
t-2-Butene	0.30	0.11	0.15	0.12	0.85	0.00	0.11	0.01	0.81	0.37	0.04
Isopentane	8.88	9.81	9.52	6.19	5.69	15.30	8.93	5.41	6.96	9.26	8.34
Pentane	6.15	5.03	6.92	4.52	2.74	8.27	1.52	3.74	2.85	3.35	5.14
c-2-Pentene	0.57	0.32	0.54	0.48	0.85	0.00	0.35	0.33	0.64	0.54	0.40
Cyclopentane	0.52	0.68	0.59	0.49	0.33	0.76	0.46	0.45	0.29	0.47	0.55
2,3-Dimethylbutane	0.93	1.04	0.92	0.82	0.83	1.30	1.16	0.75	0.90	1.12	1.03
3-Methylpentane	2.56	3.23	2.63	2.61	2.35	4.27	2.66	2.14	2.88	2.68	3.54
Methylcyclopentane	1.78	3.72	1.84	2.59	1.70	2.22	2.01	1.83	1.74	1.93	2.05
Benzene	1.85	1.96	1.85	1.70	1.35	3.02	1.53	1.98	3.68	1.62	0.48
2-Methylhexane	1.90	2.78	1.65	2.35	2.04	2.89	4.72	4.26	2.66	2.20	2.99
2,2,4-Trimethylpentane	1.25	0.25	0.84	0.42	0.00	0.00	3.73	3.84	0.00	0.26	0.00
Methylcyclohexane	0.68	0.98	0.61	0.56	0.76	0.17	0.88	0.47	0.55	0.80	0.66
2,3,4-Trimethylpentane	0.57	0.18	0.42	0.30	0.06	0.01	1.23	1.42	0.03	0.00	0.00
2,3,3-Trimethylpentane	0.58	0.00	0.00	0.00	0.00	0.00	0.80	0.60	0.00	0.00	0.00
3-Methylheptane	0.50	0.57	0.61	0.74	0.90	0.73	1.10	1.00	0.75	1.04	0.88
m-Xylene	3.88	4.24	2.98	3.68	4.05	4.61	3.56	4.50	6.47	4.14	4.61
o-Xylene	2.08	2.75	1.66	1.95	2.00	2.69	2.00	2.43	3.21	2.06	2.54
1-Methyl-4-ethylbenzene	0.73	1.03	0.68	0.65	0.60	0.93	0.67	0.74	0.92	0.78	0.88
1-Methyl-2-ethylbenzene	0.70	0.94	0.67	0.70	0.72	0.75	0.69	0.73	0.64	0.66	0.80
Naphthalene	0.41	0.29	0.40	0.36	0.39	0.20	0.27	0.46	0.29	0.21	0.30
1-Methylnaphthalene	0.12	0.09	0.13	0.14	0.20	0.11	0.13	0.07	0.09	0.03	0.12
Type	RU	RU	RU	RU	RU	RU	RU	RU	RU	RU	RU
	78	79	80	81	82	83	84	85	86	87	88
Isobutane	0.75	1.17	0.97	0.81	0.88	1.32	0.86	1.16	1.00	0.30	0.52
Butane	1.74	2.76	1.40	2.77	1.85	1.65	3.42	3.25	1.44	4.96	5.66
c-2-Butene	0.13	0.08	0.47	0.10	0.70	0.51	0.21	0.18	0.20	0.21	0.04
2-Methyl-1-butene	0.49	0.61	0.62	0.29	0.95	0.76	0.62	0.32	0.53	0.51	0.42
t-2-Pentene	0.62	0.81	0.84	0.38	1.17	1.00	0.84	0.40	0.74	0.70	0.57
2-Methyl-2-butene	1.01	1.21	1.21	0.61	1.81	1.68	1.20	0.61	1.02	0.94	0.81
2,2-Dimethylbutane	0.67	0.42	1.08	1.67	0.23	0.14	0.34	1.13	1.66	0.28	0.36
2-Methylpentane	4.47	5.45	4.02	5.43	3.42	3.06	5.74	4.93	3.70	2.92	3.58
Hexane	2.65	3.31	2.33	3.49	1.94	1.98	2.92	1.77	1.54	1.99	2.77
2,4-Dimethylpentane	0.37	0.61	0.45	1.14	0.45	0.44	0.46	0.43	0.43	0.80	0.94
Cyclohexane	0.13	0.48	0.48	1.41	0.37	0.56	0.18	0.66	1.08	0.10	0.34
3-Methylhexane	1.59	2.33	2.06	1.72	1.68	1.52	2.06	1.84	1.77	1.67	2.07
Heptane	1.06	1.51	1.31	1.21	0.85	0.88	1.46	1.55	0.93	1.17	1.64
2,4-Dimethylhexane	0.22	0.31	0.36	0.51	0.26	0.26	0.33	0.38	0.30	0.79	0.55
Toluene	7.48	5.85	6.75	7.79	6.45	5.91	4.77	8.19	6.01	6.90	7.80
2,3-Dimethylhexane	0.20	0.23	0.30	0.48	0.17	0.18	0.29	0.34	0.20	0.90	0.53
Ethylbenzene	2.24	1.36	1.59	1.31	1.56	1.70	1.88	2.06	1.43	1.32	1.58
p-Xylene	2.54	1.21	1.67	1.69	1.73	1.67	1.60	1.78	1.53	1.38	1.71
1-Methyl-3-ethylbenzene	2.26	1.85	1.82	1.10	1.73	1.97	2.08	1.97	1.59	1.48	1.71
1,3,5-Trimethylbenzene	1.34	0.97	1.01	0.66	1.01	1.07	0.98	0.92	0.93	0.74	0.89
1,2,4-Trimethylbenzene	3.60	3.08	2.87	1.75	2.71	2.79	2.97	2.76	2.35	2.65	2.84

Table 1 (*Continued*)

Compound	Sample										
	78	79	80	81	82	83	84	85	86	87	88
2-Methylnaphthalene	0.42	0.30	0.28	0.17	0.21	0.16	0.25	0.33	0.62	0.16	0.23
1,3-Butadiene	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
t-2-Butene	0.17	0.09	0.53	0.10	0.79	0.58	0.21	0.22	0.19	0.20	0.03
Isopentane	9.39	8.49	8.11	11.01	6.51	5.90	8.47	8.88	7.43	6.63	6.37
Pentane	6.97	5.34	3.05	5.52	3.26	3.21	6.47	3.11	2.00	3.39	3.43
c-2-Pentene	0.35	0.41	0.47	0.21	0.65	0.57	0.41	0.22	0.42	0.39	0.31
Cyclopentane	0.90	0.79	0.42	0.71	0.43	0.42	0.71	0.58	0.45	0.36	0.48
2,3-Dimethylbutane	1.03	1.05	0.98	1.92	0.74	0.65	1.07	1.20	1.03	1.20	0.83
3-Methylpentane	2.72	3.41	2.59	3.45	2.24	2.01	3.34	3.27	2.37	1.93	2.37
Methylcyclopenttane	1.20	2.58	1.99	2.63	1.82	1.91	1.57	2.35	2.07	1.48	1.88
Benzene	0.00	0.47	1.53	1.76	1.63	1.18	1.74	2.01	1.17	1.93	2.08
2-Methylhexane	1.94	3.16	2.64	3.28	2.18	1.98	2.45	2.35	2.28	2.39	4.02
2,2,4-Trimethylpentane	0.45	0.68	0.00	3.27	0.00	0.00	0.38	1.02	0.26	6.80	3.02
Methylcyclohexane	0.52	0.76	0.89	1.45	0.93	0.90	0.48	0.90	1.05	0.50	0.56
2,3,4-Trimethylpentane	0.20	0.29	0.07	1.54	0.05	0.06	0.23	0.32	0.06	3.69	1.13
2,3,3-Trimethylpentane	0.00	0.00	0.00	1.20	0.00	0.00	0.00	0.00	0.00	3.54	0.53
3-Methylheptane	0.60	0.78	1.21	0.76	1.10	1.05	0.89	0.98	1.27	0.84	1.02
m-Xylene	6.23	3.03	4.32	4.04	4.51	4.23	3.99	4.46	3.98	3.40	4.36
o-Xylene	3.22	1.57	2.24	2.03	2.30	2.27	2.20	2.45	2.06	2.06	2.40
1-Methyl-4-ethylbenzene	0.98	0.82	0.76	0.49	0.71	0.84	0.89	0.87	0.66	0.66	0.75
1-Methyl-2-ethylbenzene	0.87	0.81	0.79	0.40	0.73	0.90	0.86	0.84	0.73	0.68	0.76
Naphthalene	0.56	0.36	0.44	0.15	0.40	0.50	0.37	0.39	0.47	0.27	0.41
1-Methylnaphthalene	0.14	0.14	0.09	0.07	0.08	0.05	0.10	0.13	0.24	0.07	0.07
Type	RU	RU	RU	RU	RU	RU	RU	RU	RU	RU	RU

reproduced in Table 1. Samples numbered from 1 to 44 were sampled in the winter and samples 45–88 were sampled in the summer.

PCA was performed using Minitab statistical software (Minitab Inc., PA, USA). The ANNs were trained with Trajan Neural Network software (Trajan Software Ltd., Durham, UK).

3. Results and discussion

3.1. Classification of premium and regular gasolines by PCA

The data was loaded into Minitab statistical software as a 44×88 data matrix, representing the 44 variables (%mass of each compound) and the 88 samples (44 regular, 44 premium).

PCA was performed on the whole dataset. The Scree plot (Fig. 3) gives the eigenvalue of each of the principal component transformations. The eigenvalue is a measure of the variability of the data, the greater the eigenvalue the greater the variability. From this plot, it is clear that most of the variability in the data is explained by the first four principal components (approximately 75%). Therefore, the data matrix has effectively been reduced from 44 dimensions to 4 while still maintaining 75% of the information.

Fig. 4 details score scatter plots of each of the first four principal components (PC). The following general observations can be made from visual inspection of Fig. 4. There is no obvious relationship between PC1 and PC2 (plot (a)). The data points from the two types of gasoline show considerable overlap, and no discrimination can be made visually. Clustering does emerge when PC1 is plotted against PC3 (plot (b)). The majority of the regular gasolines lie above the PC3 axis, while the majority of the premium gasolines lie below it. In plot PC2 versus PC3 (plot (c)), clustering is more apparent and clear groupings can be established. Such a plot would allow classification where an unknown is entirely below (regular) or entirely above (premium) the PC3 axis. Similarly, PC3 versus PC4 (plot (f)) shows clear clustering and again would allow some classification from visual inspection of the plots. Clusters are difficult to distinguish for PC1 versus PC4 (plot (d)) and PC2 versus PC4 (plot (e)).

Thus, calculation of the principal components and interpretation of the scatter plots can provide a simple way of determining which group a gasoline sample belongs to. However, there is considerable overlap for a significant number of the premium and the regular gasoline samples when the data is viewed using two-dimensional plots, resulting in classification difficulties. Therefore, calculation of the Mahalanobis distances is required to objectively classify the samples where there is uncertainty from visual inspection.

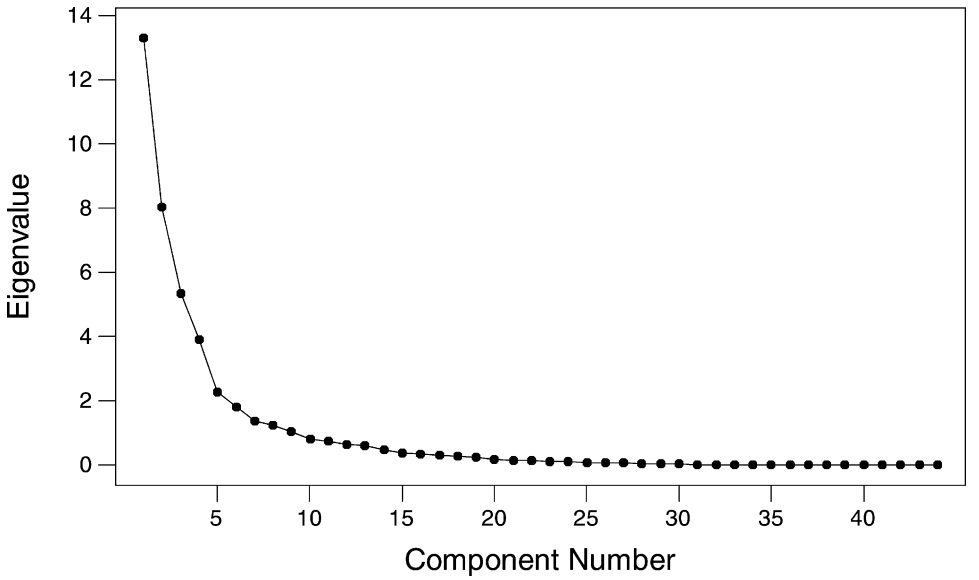

Fig. 3. Scree plot constructed using Minitab software.

The Mahalanobis distances were calculated for the whole dataset using Minitab software. Linear discriminant analysis was performed, resulting in the correct classification of 93.2% (82 from 88 samples) of the gasoline samples.

Rather than classifying the entire dataset using PCA and LDA, it would be beneficial to be able to construct a model from which class predictions could be made. Accordingly, half of the original dataset was used as a ''predictor set'' to construct the model, while the remaining data, the ''test set'', was used to test the model. Each set was chosen randomly by generating random numbers using Minitab software and ordering the dataset by an ascending random number. The first 44 samples were assigned as a predictor set and the second 44, the test set. PCA was performed on the predictor set, and the resulting Mahalanobis distances calculated from the first four principle component scores using Minitab software. The principle component scores of the test set were calculated using the eigenvalues obtained from principal component reduction of the predictor set. Each gasoline was assigned to a group using linear discriminant analysis based on the first four principal component scores. The gasolines were initially classified into regular unleaded

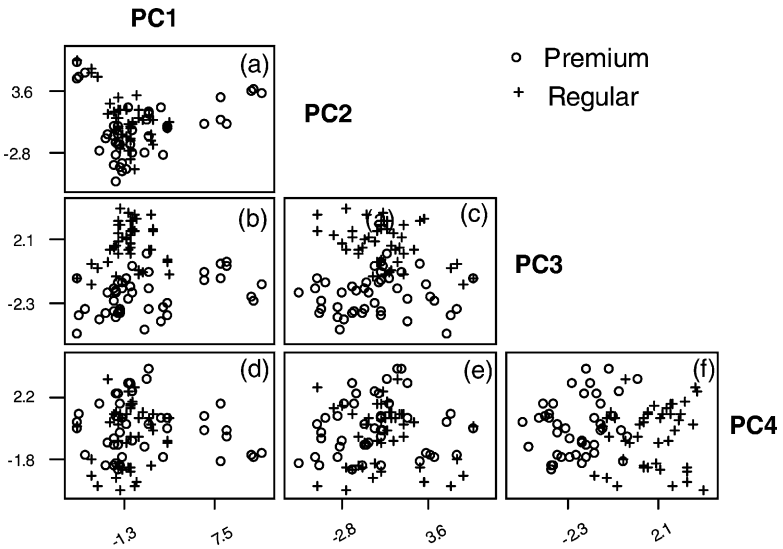

Fig. 4. Scatter plots of principal components.

Table 2
Classification of premium and regular gasoline by PCA

Dataset	Predictor				Test				Total correct
	PU actual	PU correct	RU actual	RU correct	PU actual	PU correct	RU actual	RU correct	
1	22	21	22	21	22	22	22	6	79.5
2	22	21	22	20	22	21	22	18	90.9
3	18	17	26	25	26	25	18	15	93.2
4	21	21	23	22	23	22	21	12	87.5
5	23	23	21	21	21	19	23	15	88.6
6	26	24	18	17	18	18	26	22	92.0
7	17	15	27	23	27	27	17	10	85.2
8	23	21	21	19	21	20	23	18	88.6
9	20	19	24	22	24	24	20	13	88.6
10	24	23	20	17	20	20	24	9	78.4

PU: premium unleaded; RU: regular unleaded.

and premium unleaded. This process was repeated 10 times to ensure that representative predictor and test sets were obtained. The results are presented in Table 2.

The number of samples correctly classified ranged from 79.5 to 93.2%. The variability is likely due to unevenness of the dataset when split into the predictor and the test set. That is, some predictor sets would have been more representative of the data than others due to the randomisation method used to construct each predictor set. This problem would be overcome by performing PCA and discriminant analysis on a larger dataset. Nevertheless, classification of liquid premium and regular gasolines using this technique is possible, and could be used as a routine tool to aid the analyst in the identification of accelerants and gasoline spills, assuming the localised variations of regular and premium fuels are similar to that of Canada.

The ability of PCA to classify samples was further tested by grouping each of the gasolines into its winter/summer sub-group, that is, premium unleaded winter (PUW), regular unleaded winter (RUW), premium unleaded summer (PUS) and regular unleaded summer (RUS). The first 44 samples listed in Table 1 were sampled in winter and the second 44 samples were sampled in summer. In a climate such as the one in Canada, there is a requirement for a higher fuel vapour pressure for cold engine-starting in the winter and a lower fuel vapour pressure to prevent vapour lock in the fuel line in the summer. Therefore, it was expected that the winter gasolines would tend to have more volatile compounds than the summer gasolines and that significant variations between winter and summer gasolines would be detectable. The dataset was spilt into a predictor set and a test set and a prediction model was constructed as described above. The results are presented in Table 3.

The number of samples correctly classified ranged from 47.7 to 61.4%. This is a relatively poor classification performance. Linear discriminate analysis of the principal component scores could not adequately identify the time of year in which the gasoline was sampled and the grade. It is expected that a larger dataset would improve the classification performance.

Table 3
Classification of premium winter, regular winter, premium summer and regular summer gasoline by PCA

Dataset	Predictor								Test								Total correct (%)
	PUW actual	PUW correct	PUS actual	PUS correct	RUW actual	RUW correct	RUS actual	RUS correct	PUW actual	PUW correct	PUS actual	PUS correct	RUW actual	RUW correct	RUS actual	RUS correct	
1	13	7	9	5	9	7	13	10	9	6	13	4	13	1	9	2	47.7
2	12	10	10	7	11	8	11	7	10	7	12	3	11	9	11	0	58.0
3	10	9	8	5	15	10	11	8	12	11	14	3	7	6	11	0	59.1
4	11	8	10	8	11	9	12	10	11	11	12	0	11	5	10	0	58.0
5	12	7	10	9	11	9	10	9	10	9	11	0	11	6	12	0	55.7
6	12	9	14	10	9	7	9	7	10	10	8	0	13	11	13	0	61.4
7	10	8	7	4	13	9	14	10	12	10	15	1	9	3	8	0	51.1
8	10	7	13	9	13	10	8	6	12	9	9	4	9	6	14	0	58.0
9	12	9	8	5	13	8	11	7	10	10	14	0	9	5	11	0	50.0
10	13	7	11	7	11	7	9	6	9	9	11	0	11	6	13	0	47.7

PUW: premium unleaded winter; PUS: premium unleaded summer; RU: regular unleaded winter; RUS: regular unleaded summer.

Quadratic discriminant analysis (QDA) was also performed on the above sets of data. As the name suggests, the squared distances do not simplify into a linear function as is the case with LDA. In addition, QDA does not assume that the groups have equal covariance matrices. QDA was approximately as effective in the classification of the gasolines into their respective grades and winter/summer subgroup as LDA. QDA was much less effective than LDA for the grade classification of the gasolines (70.5–88.6% correct total classification).

3.2. Classification of premium and regular gasolines by ANNs

The data was initially formatted as a data matrix in which the input variables (44 in total) of the neural network were the percent peak area of each of the compounds given in Table 1. The output of the network was simply its classification as either regular unleaded (RU) or premium unleaded (PU). Training and using a neural network requires three datasets: a training set, a cross-verification set, and a test set. The training set is used to train the network. The verification set is used to indicate possible over-learning. The test set is treated as unknowns, the correct classification of which indicates that the neural network is performing well. An ideal network would classify all three sets correctly. In this study, the 88 samples were subdivided into three datasets in a ratio of 2:1:1, that is, 44 training samples (the predictive set), 22 verification samples, and 22 test samples. This subdivision was performed randomly using the Trajan software. In addition, the order of the sample input matrix was randomized to prevent any bias in the order of the original dataset. This was achieved by generating random numbers in Microsoft Excel and sorting the data in ascending random number order before the data matrix was imported into the Trajan software. This ensures that the neural network has the best possible chance of describing the data well, as underlying structure in the input data order can adversely affect the weights on each of the nodes of the network, which diminishes performance.

A thorough search of a variety of network configurations was performed using Trajan Neural Network software. This software has an algorithm designed to mimic the heuristic process. It does this by investigating a number of configurations of the three-layered topology of the neural network as previously described. The algorithm will search through the possible combinations by sequentially changing the number of nodes in the hidden layer. Each network is trained to describe the data using a combination of back propagation and conjugate gradient training algorithms. The back propagation algorithm is used first, followed by conjugated gradient. Some network configurations are trained a number of times because each training run starts from a random selection of weights on the nodes of the neural network. This has the effect of the network reaching a different error depending upon the initial conditions. It is not possible to "know" the error surface, so one is never sure that the minimum error reached is the actual minimum or a local minimum.

Applying this technique, to find a suitable network configuration to model the GC–MS data of the gasolines resulted in excellent performance. Table 4 lists the best 10 networks (as determined by the verification error) found using the previously described process. Each quoted percent classification is based on a 95% confidence limit, and takes into account the three different datasets of training, verification and test. It is clear that each of the best 10 networks found describe the data exceptionally well. Of the best 10 networks, the worst performing network was able to correctly classify 94.5% of the gasoline samples as either premium or regular. The best network was found to be $44 \times 18 \times 2$ topology, which correctly classified all of the gasoline samples according to their grade.

The ability of the ANNs to classify the gasolines was further tested by attempting to group each of the gasolines into its winter/summer sub-group in the same manner attempted with PCA. A thorough search for suitable ANNs was performed as before, except that the network was now considered to have four classification outputs, each

Table 4
Best 10 networks

Hidden	Training error	Verification error	Testing error	Overall error	Training[a]	Correct (%)
14	5.87E−2	4.15E−7	4.97E−3	2.12E−2	BP50, CG136	97.7
18	8.31E−4	1.15E−23	1.14E−23	2.77E−4	BP50, CG035	100.0
19	3.73E−3	9.15E−14	2.13E−1	7.23E−2	BP50, CG030	98.9
19	2.65E−4	1.04E−12	2.13E−1	7.12E−2	BP50, CG029	98.9
21	3.95E−2	1.29E−8	1.46E−2	1.81E−2	BP50, CG045	98.9
29	6.76E−2	7.19E−7	2.86E−3	2.35E−2	BP50, CG124	97.7
32	8.44E−2	1.23E−7	7.04E−2	5.16E−2	BP50, CG057	96.6
35	8.60E−4	7.65E−21	3.02E−1	1.01E−1	BP50, CG036	97.7
35	2.76E−2	8.20E−11	3.01E−1	1.09E−1	BP50, CG032	94.3
35	1.01E−1	1.24E−8	2.67E−1	1.23E−1	BP50, CG032	95.5

[a] Indicates how the network was trained, e.g. BP50, CG136 means the network was trained initially with back propagation for 50 epochs, and then conjugated gradient for 136 epochs.

Table 5
Classification of premium winter, regular winter, premium summer and regular summer gasoline by ANNs

	Training				Verification				Test			
	PUW	RUS	PUS	RUW	PUW	RUS	PUS	RUW	PUW	RUS	PUS	RUW
Total	11	11	13	9	6	5	5	6	5	6	4	7
Correct	11	11	13	9	6	5	5	6	5	5	3	6
Wrong	0	0	0	0	0	0	0	0	0	0	1	0
Unknown	0	0	0	0	0	0	0	0	0	1	0	1

PUW: premium unleaded winter; PUS: premium unleaded summer; RU: regular unleaded winter; RUS: regular unleaded summer.

representing the gasolines' sub-group, that is, premium unleaded winter, regular unleaded winter, premium unleaded summer and regular unleaded summer. As before, the dataset consisted of 44 training, 22 verification and 22 test points. The best network was found to be $44 \times 53 \times 4$ topology. This network had an overall error of 5.1×10^{-1} rms. Table 5 details the classification for each of the datasets within a 95% confidence limit. This ANN correctly classified all of the training and verification samples. The ANN misclassified one premium unleaded summer gasoline, and could not classify one regular unleaded summer and one regular unleaded winter in the test sample set. Overall, the ANN correctly classified 96.6% of the samples. This is slightly less than the correct classification of all the samples (100% success) achieved with the previous ANN in which the data was classified into the two classes, premium and regular unleaded. This result is superior to the classification performance obtained from PCA and the discriminant analysis utilising Mahalanobis distances.

4. Conclusions

PCA is a very powerful dimensionality reduction technique, which allows highly correlated, multivariate data to be plotted and classified in a very simple way by the interpretation of scatter plots of the principal components. Calculation of the Mahalanobis distances of the principal components of the GC–MS data allowed the gasoline samples to be classified as either regular or premium grade with a high degree of accuracy. PCA and LDA performed poorly when it was attempted to further sub-divide the gasoline samples into their respective summer/winter groupings.

Training artificial neural networks to recognise the structures within the GC–MS data correctly classified premium and regular unleaded gasoline samples. It has been shown that all the gasoline samples could be correctly classified when an ANN was trained with a combination of back propagation and conjugate gradient algorithms. Furthermore, approximately 97% of the gasoline samples could be further sub-classified according to their seasonal formulation (winter or summer).

The ANNs were a valuable tool for the classification of the gasolines due to their simplicity, and excellent predictive ability.

References

[1] ASTM Method E1618-97, Standard Guide for Identification of Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography–Mass Spectrometry.
[2] B. Tan, J. Hardy, R. Snavely, Accelerant classification by gas chromatography/mass spectrometry and multivariate pattern recognition, Anal. Chim. Acta 422 (2000) 37–46.
[3] M.A. Sharaf, D.L. Illman, B.R. Kowalski, Chemometrics, Wiley, New York, 1986.
[4] C. Bishop, Neural networks and their applications, Rev. Sci. Instrum. 65 (6) (1994) 1803–1832.
[5] J.M. Andrews, S.H. Lieberman, Neural network approach to qualitative identification of fuels and oils from laser induced fluorescence spectra, Anal. Chim. Acta 285 (1994) 237–246.
[6] B.K. Lavine, A.J. Moores, H.T. Mayfield, A. Faraugue, Fuel spill identification by gas chromatography–genetic algorithms/pattern recognition techniques, Anal. Lett. 31 (15) (1998) 2805–2822.
[7] M. Ichikawa, N. Nonaka, I. Takada, S. Ishimori, Mass spectrometric analysis for distinction between regular and premium motor gasolines, Anal. Sci. 9 (1993) 261–266.
[8] Composition of Canadian Summer and Winter Gasolines, Canadian Petroleum Products Institute, Report No. 945, 1993.