University of Strathclyde

Department of Pure and Applied Chemistry

The Application of Pattern Recognition Techniques to Data Derived from the Chemical Analysis of Common Wax Based Products and Ignitable Liquids

By

Ismail Dzulkiflee

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

2010

Posters, Presentations and Publications

- *Application of Pattern Recognition Procedures for the Classification of Wax-Based Products,* (Poster Presentation). Mediterranean Academy of Forensic Sciences (MAFS), 2nd Workshop, 2006, Malta.

- *Pattern Recognition Procedures for the Classification and Prediction of Stains Originating from Wax-Based Products,* (Oral Presentation). California Association of Criminalists (CAC), Fall Seminar, 2006, Temecula, California.

- *Artificial Neural Networks (ANNs) in the Development of Predictive Tools for Wax-Based Products: Preliminary Studies,* (Poster Presentation). Forensic Science Society AGM, 2008, Wyboston, England.

- *Comparison of Smears of Wax-Based Products: Thin Layer Chromatography (TLC) and Microspectrophotometric (MSP) Detection,* (Poster Presentation). 5th European Academy of Forensic Science Institute (EAFS), 2009, Glasgow, Scotland.

- *Clustering of Wax-Based Products: A Self Organising Feature Maps (SOFM) Approach,* (Poster Presentation). 5th European Academy of Forensic Science Institute (ENFSI), 2009, Glasgow, Scotland.

- *Application of Unsupervised Chemometric Analysis and Self Organising Feature Maps (SOFM) for the Classification of Lighter Fuels.* Paper published in Analytical Chemistry, 2010, 82, p. 6395 - 6400

- *Comparison of Smears of Wax-Based Products using Thin Layer Chromatography (TLC) and Microspectrophotometric Detection (MSP).* Paper published in Journal of Forensic Identification, 2011.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Dr. Niamh Nic Daeid for her supervision, encouragement, motivation and fruitful advice during the period of my study and above all for giving me the opportunity to complete my PhD journey. Thanks are also extended to all staff at the Centre for Forensic Science, both academic and technical especially to Ms. Lynn Curran and the phenomenal guy, Mr. James Christie.

I would also like to thanks the Friendly Forensic Buddies (F.F.Bs) especially to Miss Lucy Hill, Miss Anika Ludwig, Mr. Kevin Farrugia, Mr. Graham Reed, Mr. Saravana Kumar, Ms. Yuvaneswari, Miss Alaia Lopez, Mr. Phuvadol Thanakiatkrai (Ice), Miss Thitika Kitpipit (Four), Miss Aude Mazollier, Dr. Ainsley Dominick, Dr. Shanan Tobe and Mr. Greg Owens for making my PhD journey packed with fun.

Above all, I am truly indebted to my nearest and dearest, my beloved wife, Ms. Wan Nur Syuhaila for being very supportive and understanding throughout the years and to my pride and joy, my two beautiful daughters, Miss Wan Nur Aisyah and Miss Wan Nur Aein. Thanks also go to my dear friends who are always there for me and my family through thick and thin.

Last but not least to Prof. Peter White and Dr. James Thorpe who had thought me about life and to the University of Strathclyde, a place of useful learning and so it is.

R647, Royal College Building
Summer 2010
Glasgow

ABSTRACT

Pattern recognition is a term that can be used to cover various stages of the investigation of characterising data sets including contributing to problem formulation and data collection through to discrimination, assessment and interpretation of results. Chemometrics techniques and Artificial Neural Networks (ANNs) are pattern recognition techniques commonly used to visualise and gather useful information from multidimensional datasets i.e. datasets with *n*-samples with *m*- variables.

Of the many chemometric techniques available, Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) are the most commonly used in the evaluation of dataset(s) generated from the analysis of samples which have relevance to forensic science. By contrast, Artificial Neural Networks (ANNs) and in particular Self Organising Feature Maps (SOFM) and Multi Layer Perceptron (MLP) have had limited application in forensic science eventhough these pattern recognition techniques have been known for almost 30 years.

This study focuses on the applicability of the Artificial Neural Networks (ANNs) to specific datasets of forensic science interest and compares these with 'conventional' PCA and HCA techniques. Datasets generated from the analysis of wax based products and lighter fuels were used. The wax based product data set contained information obtained from Thin Layer Chromatography (TLC), Microspectrophotometry (MSP), Ultra-Violet and Visible Spectroscopy (UV/Vis) and Gas Chromatography with Flame Ionisation Detector (GC-FID) analysis of a variety of products from multiple sources where discrimination by brand was the objective. The data provided for the lighter fuel samples was obtained from analysis of a number of brands, both unevaporated and evaporated by Gas Chromatography-Mass Spectroscopy (GC-MS) and the objective was to discriminate the samples by brand as well as link degraded samples from the same brand together.

The wax based product analysis provided simple, straight forward data whilst the lighter fuel analysis provided a more complicated and challenging dataset to investigate in terms of facilitating sample discrimination and/or linkage. In all cases, the 'conventional' Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) failed to provide any meaningful discrimination of the samples by product type regardless of the nature of the datasets. In contrast, the Artificial Neural Networks (ANNs) techniques provided full discrimination of the samples by product type even when the samples had undergone considerable ageing and weathering.

This work has demonstrated the potential use of Self Organising Feature Maps (SOFM) and Multi Layer Perceptron (MLP) to datasets of forensic science relevance. The findings of this work provide avenues for further exploration of Artificial Neural Networks (ANNs) in forensic science.

Abbreviations

ASTM	American Standard Testing Methods
ANN	Artificial Neural Network
BP	Back Propagation
BP-NN	Back Propagation Neural Network
CG	Conjugate Gradient
CSV	Comma Separated Values
CVA	Canonical Variate Analysis
DCM	Dichloromethane
GC-FID	Gas Chromatography Flame Ionisation Detector
GC-MS	Gas Chromatography Mass Spectrometry
GPC	Gel Permeation Chromatography
HCA	Hierarchical Cluster Analysis
HPLC	High Performance Liquid Chromatography
IEC	Ion Exchange Chromatography
IR	Infra Red
IUPAC	International Union of Pure and Applied Chemistry
LB	Lip Balm
LDA	Linear Discriminant Analysis
LLC	Liquid-Liquid Chromatography
LM	Lavenberg-Marquardt
LSC	Liquid Solid Chromatography
MLP	Multi Layer Perceptron
MS	Mass Spectroscopy
MSE	Mean Squared Error
MSP	Microspectrophotometry
OCVA	Orthogonal Canonical Variate Analysis
PCA	Principal Component Analysis

PLOT	Porous Layer Open Tubular
PMT	Photo Multiplier Tube
R	Regression Coefficients
RSD	Relative Standard Deviation
SCOT	Support Coated Open Tubular
SEC	Size Exclusion Chromatography
SOFM	Self Organising Feature Map
SP	Shoe Polish
TCC	Target Compound Chromatogram
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
TMAH	Tetra Methyl Ammonium Hydroxide
UVA	Ultra-violet A
UVB	Ultra-violet B
UV/Vis	Ultra-Violet and Visible
WCOT	Wall Coated Open Tubular

Glossary

Algorithm	Step by step process or sequence of instructions used to solve a problem or for completing a given task. Similar to recipe in laymen term.
Artificial Neural Network	A group of interconnected artificial neurons. Used to model underlying mathematical function in a given dataset. Also used for data or information processing and data mining for finding linear and non-linear relationships between variables in a given dataset
Chemometric	The application of mathematical or statistical techniques to data derived from chemical analysis for finding structure or relationships within a given dataset.
Hidden Layer	One of the layer in ANN that is responsible for processing the input vectors (patterns) introduced to the ANN. Hidden in the sense that it does not have 'interactions' with the outside environment.
Input Layer	The first layer in ANN. A 'fan-out' layer whose neurons do not involve in any mathematical processing of the input vectors (patterns).
Iterations	The number of time the input vectors (patterns) introduced into the ANN. 1000 iterations mean that the input vectors (patterns) are introduced 1000 times. Also known as epochs.
Ignitable Liquids	Flammable liquids commonly used by arsonists to initiate, promote and increase the rate of fire
Lanolin	Fatty substances that are extracted from sheep wool. Commonly used in soaps, cosmetics and ointments.
Neuron	In neural network sense is the basic building block of ANN.

Output Layer	The final layer in ANN that produces the output from ‘mathematical’ processing of the input vectors (patterns) by the hidden neurons in the hidden layer.
Over Learning	Undesirable condition or situation in neural network where network instead of modelling the underlying ‘mathematical’ function of a given dataset also model ‘noise’ within the dataset which jeopardise the network ability to generalise. Best described by the curve fitting process. Also known as over-fitting.
Pattern Recognition – Supervised Learning	Learning process in pattern recognition techniques that requires both the input and output vectors (patterns) pairs. The output patterns act as ‘teacher’ that direct the outcomes of a neural network.
Pattern Recognition – Unsupervised Learning	In contradict to the unsupervised learning. Do not require any output vectors (patterns) in the learning process. Pattern recognition techniques for examples PCA and SOFM are left on their own to find any structure or natural groupings within a given dataset.
Performance	In ANN sense especially with the Multi Layer Perceptron (MLP), performance is measured in term of mean squared error (MSE), the difference between the actual output and target vectors (patterns).
Test Set	A set of data that is used to check on the generalisation ability of a neural network. Introduced after training of neural network has stopped.
Training	In neural network sense is the process of updating the weights between neurons connections.
Training Set	A set of data that is used to train a neural network.

Validation Set	A set of data that is used to validate a neural network whose error is used to monitor the training progress of a neural network in other words to check on possible network over-learning or over-fitting.
Wax-Based Products	In this study means products that constitute mainly of waxes with other ingredients but in much smaller proportions for examples dyes, pigments and oils.

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 3

Chapter 4

Chapter 5

THESIS OVERVIEW

Chapter 1 presents the underlying principles of pattern recognition techniques. It also discusses some principles behind Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) which are the 'conventional' multivariate pattern recognition techniques used in this study. It also gives some insight on the applications of PCA and HCA in some areas of forensic science.

In Chapter 2, the general ideas behind Artificial Neural Networks (ANNs) are presented. This includes some underlying principles of Self Organising Feature Maps (SOFM), an unsupervised neural network and Multi-Layer Perceptron (MLP), a supervised neural network. These were the two artificial neural network methods used in this work.

Chapter 3 and Chapter 4 present the application of the pattern recognition techniques i.e. both the conventional i.e. PCA and HCA and ANNs particularly SOFM to datasets generated from wax-based products and lighter fuels analysis respectively. Chapter 5 discusses the application of the supervised MLP to datasets generated from wax-based products analysis which also discusses and demonstrates its ability as predictive tool.

Finally, Chapter 6 provides some overall conclusions to the work presented and suggests areas relevant to forensic science in which the data analysis approach proposed in this work could be further carried and explored.

A number of appendices are also included which provide the detailed methodology used to generate the wax based product data set in particular. The lighter fluid data set was supplied as a set of raw data from chromatographic analysis of the samples using a validated methodology.

CHAPTER 1: INTRODUCTION TO MULTIVARIATE PATTERN RECOGNITION TECHNIQUES

1.1 INTRODUCTION

This chapter describes the theoretical background of the 'conventional' pattern recognition techniques used in this study namely Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA).

1.2 Pattern Recognition

Pattern Recognition as a field of study was developed around the 1960's [1]. It was very much an interdisciplinary subject, covering developments, among others, in the areas of statistics, engineering, artificial intelligence, computer science, psychology and physiology [1]. A pattern recognition investigation, as outlined by Webb [1], consists of the following stages:

1. Formulation of the problem i.e. gaining a clear understanding of the aims of the investigation and planning the remaining stages.
2. Data collection i.e. making measurements on appropriate variables and recording details of the data collection procedure.
3. Initial examination of the data i.e. checking the data, calculating summary statistics and producing plots in order to explore the structure of the data.
4. Feature selection or feature extraction: selecting variables from the measured set that are appropriate for the task. These new variables may be obtained by a linear or non-linear transformation of the original variables.
5. Unsupervised pattern classification or clustering. This may be viewed as exploratory data analysis (EDA) and it may provide a successful conclusion to a study. On the other hand, it may act as a means of pre-processing the data for a supervised classification procedure.

6. Apply discrimination or regression procedures as appropriate. The classifier is designed using a training set of exemplar patterns.
7. Assessment of the results. This may involve applying the trained classifier to an independent test set of labelled patterns.
8. Interpretation.

This pattern recognition investigation is an iterative process, which means that the analysis of the results may pose further hypotheses that require further data collection. The above cycle can be terminated at different stages, for example when the question posed can be answered by the initial examination or if it is determined that the data obtained cannot answer the question posed [1].

The term 'pattern' does not only mean structure within an image. In its broader sense, it covers any aspects where measurements can be made for example, measurements made on patients in order to identify a disease or measurements on weather variables in order to predict or forecast future weather outcomes. In a mathematical sense, the term pattern can be described as $\boldsymbol{p}$-dimensional data vector $\boldsymbol{x}$, where $\boldsymbol{x} = (x_1,\ldots, x_p)$ where x_i represents the measurements made on features of an object or sample which are deemed to be important to describe the object or sample.

The overall goal of pattern recognition is classification [2], to classify an object or set of objects given a set of measurements associated with the object(s). The pattern recognition techniques can be either unsupervised (searching for natural groupings within a dataset) or supervised (assigning group to the object beforehand). Two of the most commonly used mechanisms for the application of pattern recognition techniques are Chemometrics and Artificial Neural Networks (ANNs) which will be discussed in the following sections.

1.3 Chemometrics

Chemometrics is an approach to data analysis that uses mathematical and statistical methods to determine the properties of substances that otherwise would be very difficult to measure directly [2]. Measurements related to the chemical composition of a substance are usually taken, and the value of some property of interest is inferred from these measurements, through an appropriate mathematical relationship [2].

1.3.1 Multivariate Data

Most chemical analyses produce datasets of a number of measurements of variables over a range of objects or samples. Such multivariate datasets carry more information than uni-variate or bi-variate datasets which are based on measurements of one or two variables only. Multivariate datasets are often presented in the form of a matrix, Figure 1.1, where the row represents the number of measurements taken for a particular sample and the column represents the number of measurements taken for a particular variable.

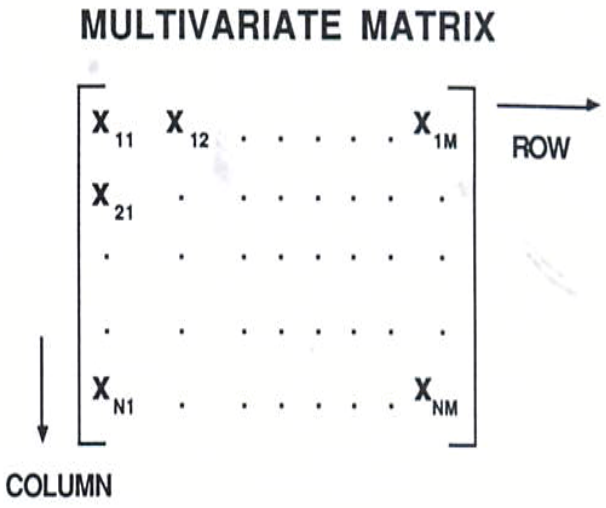


Figure 1.1 Multivariate data matrix – The row represents the samples measurements and the column represents the variables measurements. Reproduced from Everitt [3]

Multivariate datasets have high dimensionality and once the matrix exceeds three dimensions it becomes very difficult to analyse without the use of appropriate mathematical data analysis systems. Commonly used techniques to process multivariate datasets are Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA). These techniques fall under the general framework of pattern recognition using an unsupervised pattern recognition scheme.

The main objective of these techniques is dimensionality reduction, in other words, processing of a high dimensional multivariate dataset into a more manageable low dimensional one, most commonly into a two or sometimes three dimensional form. The final outcome of these techniques are a series of clusters of the data, a process defined by Hong *et al.* [4] as a methodology for finding a natural structure within a dataset based on the partitioning of a collection of data points into a number of subgroups (clusters) where objects inside a cluster show a certain degree of similarity to each other. The important aspect of these techniques is that the clusters can be visualised graphically by means of a score plot (in PCA) or a dendrogram (in HCA).

1.3.2 Principal Component Analysis (PCA)

PCA is used to identify patterns in data in such a way as to highlight their similarities and differences [3]. The original dataset is described using new variables known as the principal components (PCs) derived by linear combinations of the original variables with specific loadings for each principal component as in Equation 1.1 [3];

$$PC_n = \alpha_{n1}X_1 + \alpha_{n2}X_2 + ... + \alpha_{np}X_p \quad \text{.....equation 1.1}$$

Where α is the loading value and X_i is the standardised value of the original variables.

If, for example, there are 38 variables in a given dataset, then it would be expected that 38 PCs could be derived however only 2 or 3 of these PCs may effectively describe or explain most of the variability in the dataset thus reducing the dimensionality of the data set. Associated with each PC is an eigenvalue which measures the 'size' of each PC such that the larger the eigenvalues, the more significant the PC is.

In PCA, the PC's are ordered in decreasing order of importance where the first PC accounts for the largest variability in the dataset followed by the second and so on. In choosing the number of PCs to describe the original dataset, the conventional approach is to retain just enough components to explain as large a percentage of the total variability of the original dataset as possible. Values of 70 – 90% variation are normally suggested [3]. PCA generates two smaller matrices known as loading and score matrices where the former contains information about the PCs and the latter contains information about the samples which are described in term of their ordination onto the PCs [5].

Graphical representations of the PCs and data can be constructed from the loading and score matrices. A score plot of the first two PCs is most commonly used to describe the cluster outcomes. This plot provides an important means of visualising and summarising the original dataset and often reveals patterns that were previously elusive. The score plot shows the relative position of the samples where samples having similar scores are positioned closely together. An example of a typical PCA score plot of tea samples from Tagori *et al*. [6] is presented in Figure 1.2.

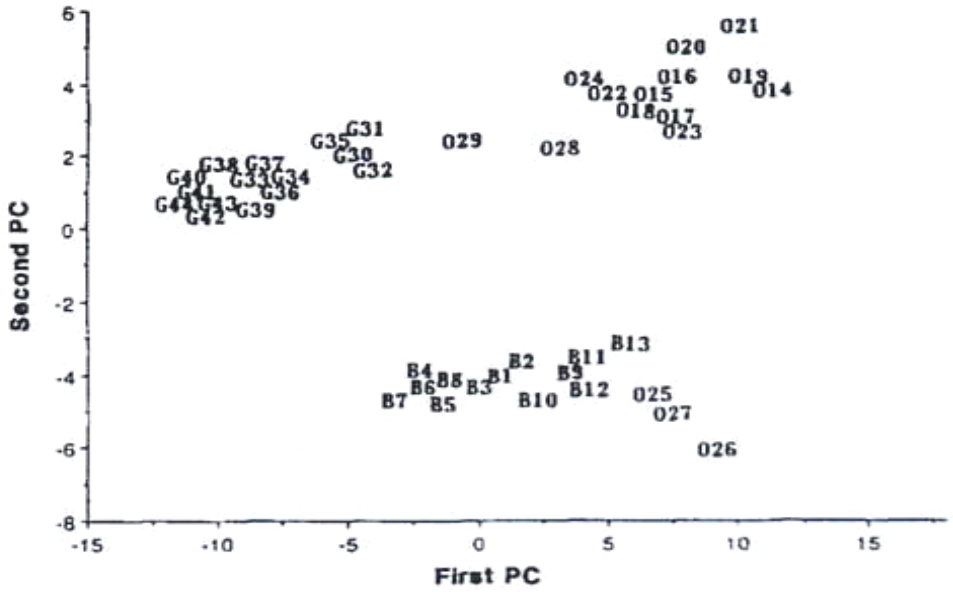


Figure 1.2 Score plot of first and second principal component. Reproduced from Tagori *et al* [6] based on study on tea samples.

Associated with the principal component transformation is a scree plot which gives the measures of the variability of each principal component derived from a given dataset. An example of a scree plot is illustrated in Figure 1.3. A scree plot presents the percentage variability within the data set accounted for by each principal component. For example in Figure 1.3, the first principal component accounts for 68% of the variability while the second principal component accounts for approximately 18%. Both principal components account for 86% of the variability in the dataset.

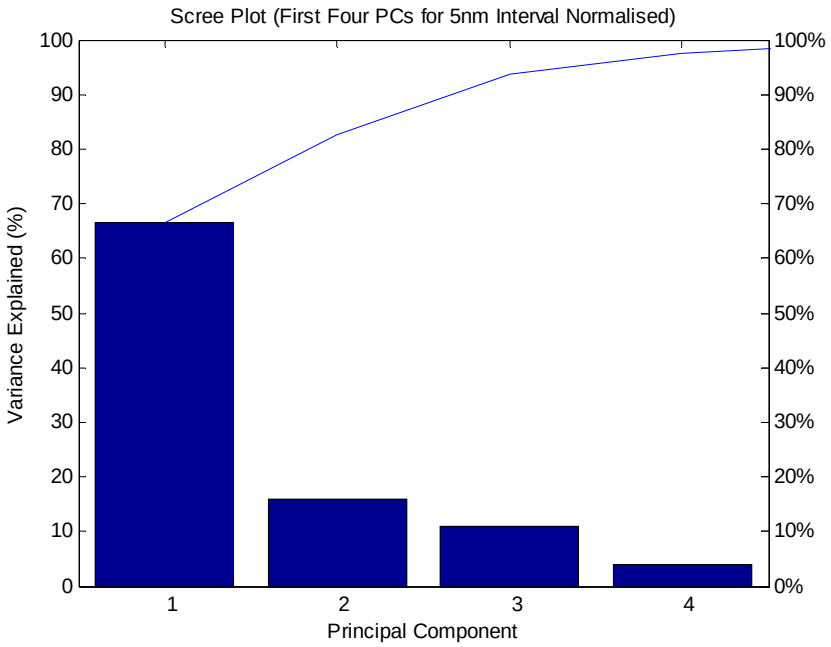


Figure 1.3 Example of a scree plot

1.3.3 Hierarchical Cluster Analysis (HCA)

HCA uses either agglomerative or divisive methods to identify clusters in a given dataset. In the agglomerative method [3, 7], clustering begins with all observations being separate, each forming its own cluster. In the first step, two samples or objects which are close (based on their proximity measure) are joined. In the next step, either a third object joins the first two objects, or the two other objects join together into a different cluster. This process continues until all clusters are joined into one large group. By contrast in the divisive method, all observations are considered as one large group and successively divided into smaller groups until each group contains only one sample [3, 7]. The outcomes of HCA can be visualised graphically by a two-dimensional tree diagram or dendrogram which illustrates the joining made at each stage of the clustering process and is depicted in Figure 1.4.

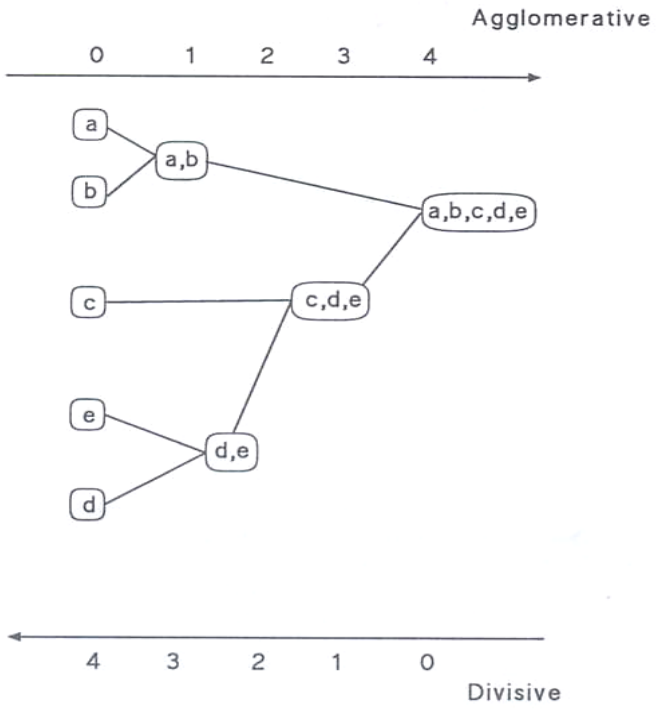


Figure 1.4 A typical dendrogram with arrows showing the clustering strategies. Agglomerative starts with all objects form one individual cluster while divisive starts with all objects in one cluster. Reproduced from Everitt [3]

In HCA, the similarity between two objects or cluster is measured by its distance such that the closer the distance is, the greater the similarity between the objects. There are a number of distances or proximity measures available for examples Euclidean, Manhattan and Mahalonobis distance [8] but for the purpose of this study only Euclidean distance was used which is given by Equation 1.2.

$$d_{AB} = \left[\sum_{j} (x_{1,j} - x_{2,j})^2 \right]^{1/2} \quad \text{.....equation 1.2}$$

Where d_{AB} is the distance between object A and B, $x_{i,j}$ is the value of the j^{th} variable measured on the i^{th} object.

The clustering strategy in a hierarchical procedure defines how similarity is determined between clusters containing multiple member [8]. There are numerous clustering strategies available where single, complete and average linkage are the most popular [8].

The single linkage method uses the minimum distance between samples to be clustered as the measure of relatedness. Complete linkage on the other hand, uses the maximum distance between samples while the average linkage uses the average distance between all pairs of samples. The various linkage mechanisms are depicted in Figures 1.5 to 1.7. Different clustering strategies may give different clustering solutions [7] and a number of clustering strategies are normally evaluated to determine which describes the data best.

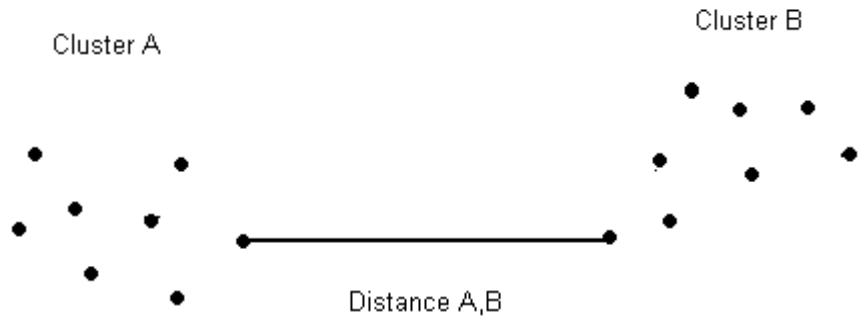


Figure 1.5 Single Linkage Cluster. Distance A, B is the smallest distance between objects in the two clusters.

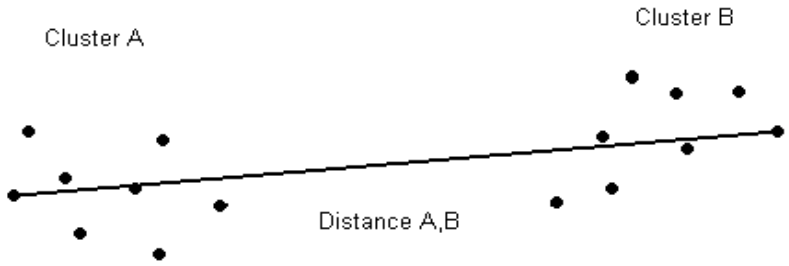


Figure 1.6 Complete Linkage Cluster. Distance A, B is the maximum distance between objects in the two clusters.

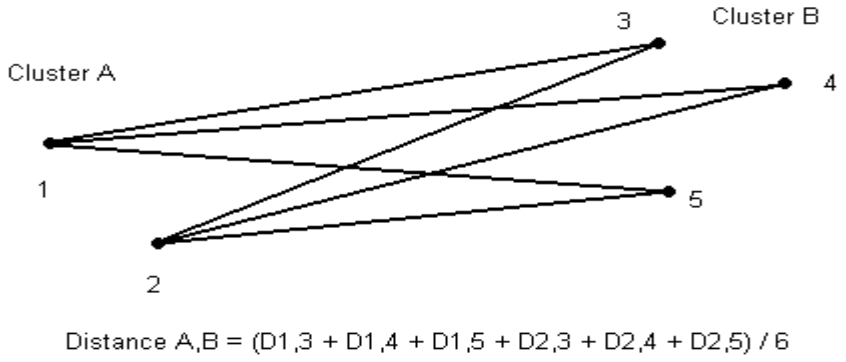


Figure 1.7 Average Linkage Cluster. Distance A, B is the average distance between all samples in the two clusters.

1.4 Applications of PCA and HCA in Forensic Science

In forensic science, PCA and HCA have been increasingly used in the interpretation of data sets of forensic science relevance which include datasets generated from drugs of abuse analysis [5, 9, 10], inks and dyes analysis [11, 12], accelerant analysis [13-15], footwear marks analysis [16], toxicology [17] and environmental forensic analysis [18, 19].

1.4.1 Applications in Drugs of Abuse

To combat the illicit heroin situation in Germany, the Federal Government of Germany initiated a program involving cooperation between forensic science laboratories within the states (Landeskriminalämter, LKA) and the Federal Criminal Police Office (Bundeskriminalamt, BKA) developing harmonised analytical methods for quantitation and comparison of heroin samples [9]. The quantitative determination was carried out using capillary gas chromatography with flame ionisation detector (GC-FID). The technique was chosen because of its ability to separate the principal opium alkaloids and derivatives as well as adulterants and diluents present in illicit heroin samples. This resulted in the generation of multi peak trace impurity profiles which required an objective means for comparison. From 1991 onwards, PCA and HCA were used to tackle both large and complicated datasets resulted from the impurity profiling. Figure 1.8 shows the dendrograms from the HCA for the analysis of heroin samples in Germany in 1991.

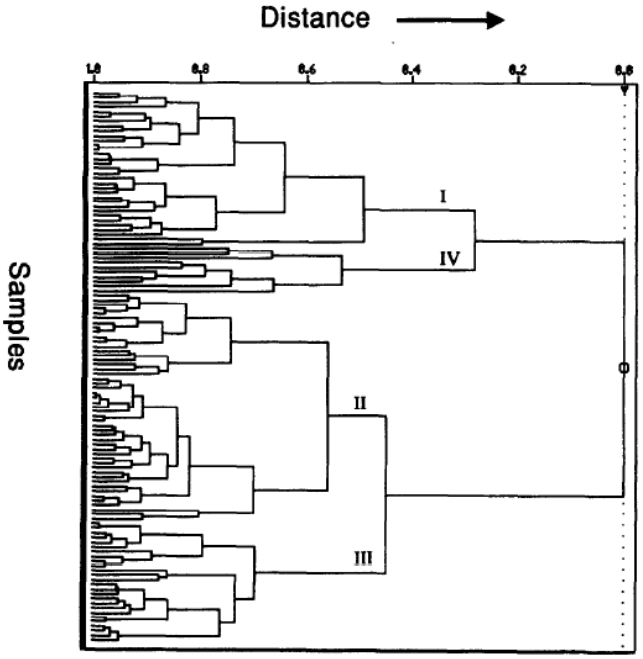


Figure 1.8 Dendrogram from HCA of heroin samples in Germany in 1991. Reproduced from Neumann [9]

A similar approach was taken in France, where a portion of all heroin seizures are sent to the Forensic Laboratory in Lyon (LPS Lyon) which performs chemical analyses and stores the analytical data in a national database [10]. In each case the sample is subjected to qualitative and quantitative analyses, analysis of trace level impurities and isotopic analysis of the major constituents present. The quantitative analysis undertaken using gas chromatography with flame ionisation detector (GC-FID), also involved the use of PCA in order reveal relationships between the samples. Three component ratios i.e. whole morphine/acetylcodeine (ratio 1), whole morphine/papaverine (ratio 2) and narcotine/papaverine (ratio 3) were used as the PCA variables. It has been reported that PCA successfully identified and distinguished different samples from different seizures using only the three component ratios investigated.

1.4.2 Applications in Inks and Dyes Analysis

Thanasoulias *et al* [11] applied chemometrics techniques i.e. cluster analysis (CA) and linear discriminant analysis (LDA) including PCA for discrimination of blue ballpoint pen inks using their visible spectra. Fifty blue ball-point pen inks from five different brands were examined where their spectra were measured in the range of 400 – 750 nm. In doing so, 351 variables were made available each representing a 1 nm wavelength interval. The major problem associated with the study was that the STATISTICA software used did not allow more than 300 variables for its analyses. In order to overcome this issue, cluster analysis was performed prior to PCA. The latter was used to detect any outliers. Figure 1.8 shows the PCA score plot generated from the study for detection of outliers. The cluster analysis was used to calculate the classification functions used for classification of new items.

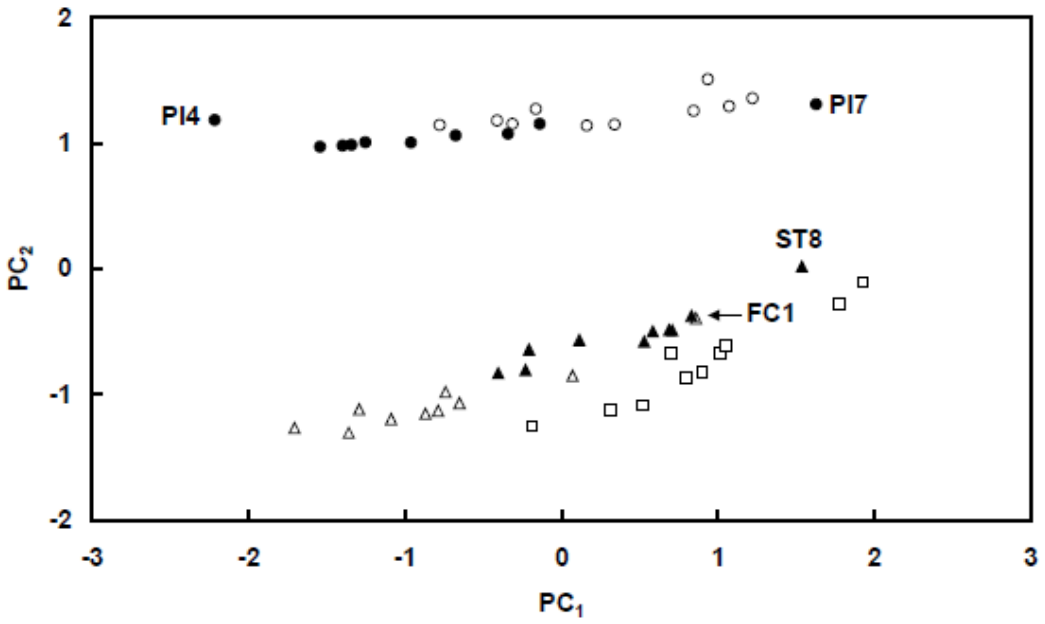


Figure 1.8 PCA score plot for detection of outliers. P14, P17, ST8 and FC1 are outliers. Reproduced from Thanasoulias *et al* [11].

In another study, Adam *et al* [12] applied PCA for classification and attempted discrimination of black ball-point pen inks available in the UK market using UV/Vis absorption spectra measured from 300 – 700 nm. Numerical loadings were attached to the first few principal components i.e. PC1, PC2 and PC3 were used to characterise each spectrum within each pen group. For the complete set of 25 pens studied, it was found that, interpretation of the loadings for the first few principal components were able to classify both pen inks and their respective ink lines in an objective manner. Figure 1.9 shows the score plot for the complete 25 pen inks and ink lines under study. Having established the potential of PCA of UV/Vis spectroscopy data for discriminating the pen inks, PCA was later used to identify questioned spectra from questioned inks. The idea was that, inks i.e. reference inks which are spectrally similar to the questioned spectra would be assigned loadings which were numerically close to each other thus potentially identifying the ink type [12].

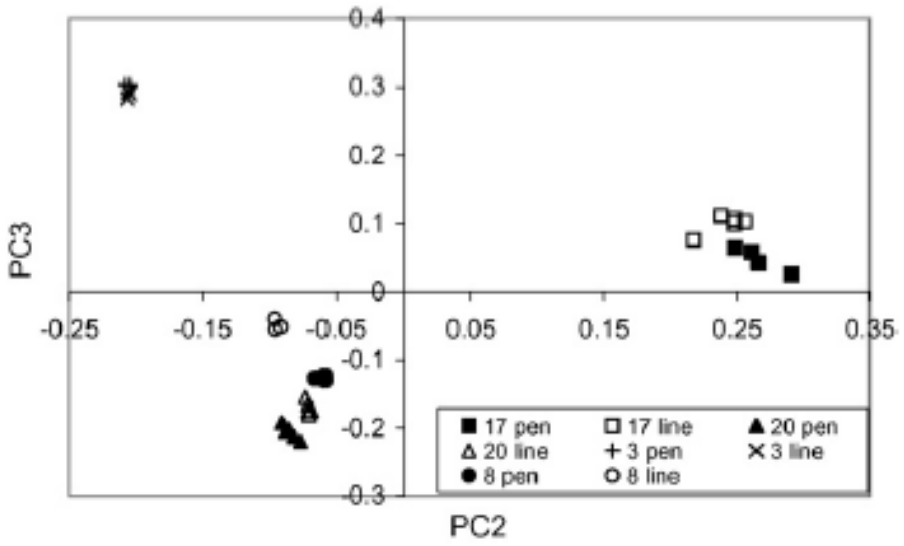


Figure 1.9 PCA score plot for the complete pen inks and their respective ink lines. Reproduced from Adam *et al* [12].

1.4.3 Application in Footwear Marks Analysis

In the field of forensic footwear examination, it is widely held belief that patterns of accidental marks found on footwear and footwear impressions possess a high degree of 'uniqueness' [16]. This belief however, has not been thoroughly studied in a numerical way using controlled experiments [16]. As a result of this, such physical evidence has been the subject of admissibility challenges. To numerically measure how similar footwear mark patterns are to one another, Petraco *et al* [16] used PCA where inputs (for the PCA) were the accidental patterns measured from the top portion of the shoe soles. In the study, five pairs of shoes of same make and sole pattern and worn by same subject for 30 days were used. During the study, the shoe prints were recorded on days 1 through 7, 14, 16, 18, 20, 24, 28 and 30. Figure 1.10 shows the three-dimensional (3D) PCA score plot involving PC1, PC2 and PC3 for all the shoe pairs recorded at the initial stages of the study where some clustering of the shoes are evident.

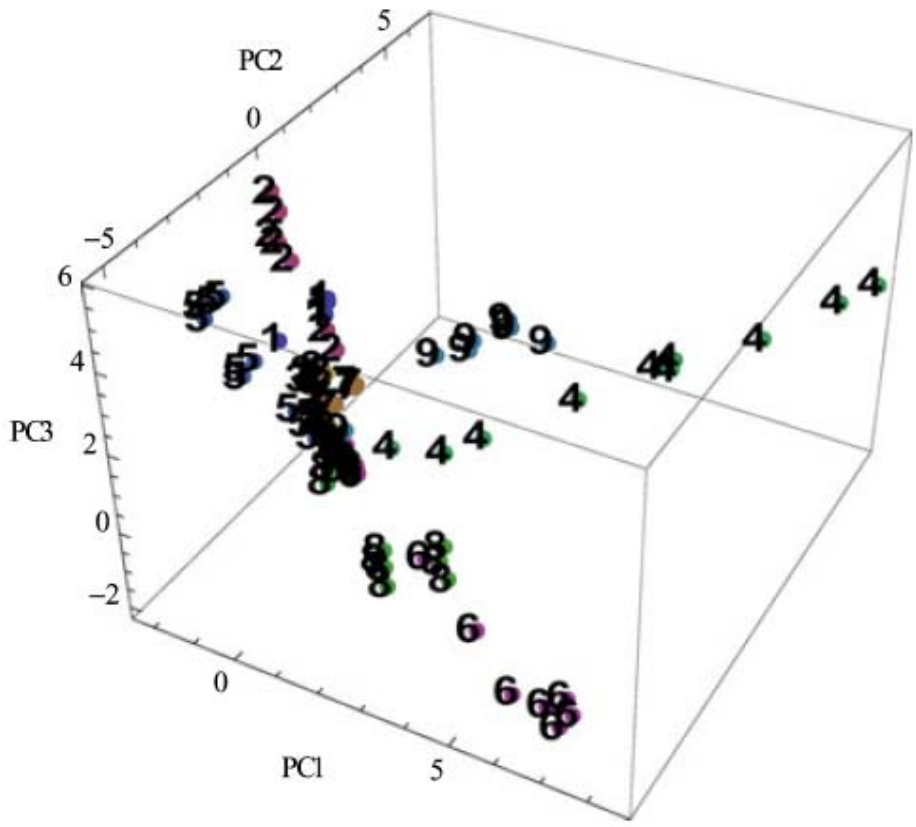


Figure 1.10 3D PCA score plot for the shoe under study at the initial stages of the study. The numbers label the shoes (n=10) used in the study. Reproduced from Petraco *et al* [16].

The best clustering of the shoes (based on their accidental patterns) were obtained at the later stages of the study i.e. from the period of 20 – 30 days of wear. Figure 1.11 shows the 3D PCA score plot for the shoes at the later stages of the study where distinguish clusters are evident. The finding of this study shows the capability of PCA to numerically measure the similarity of patterns promoting the proposition of individuality of footwear impressions [16].

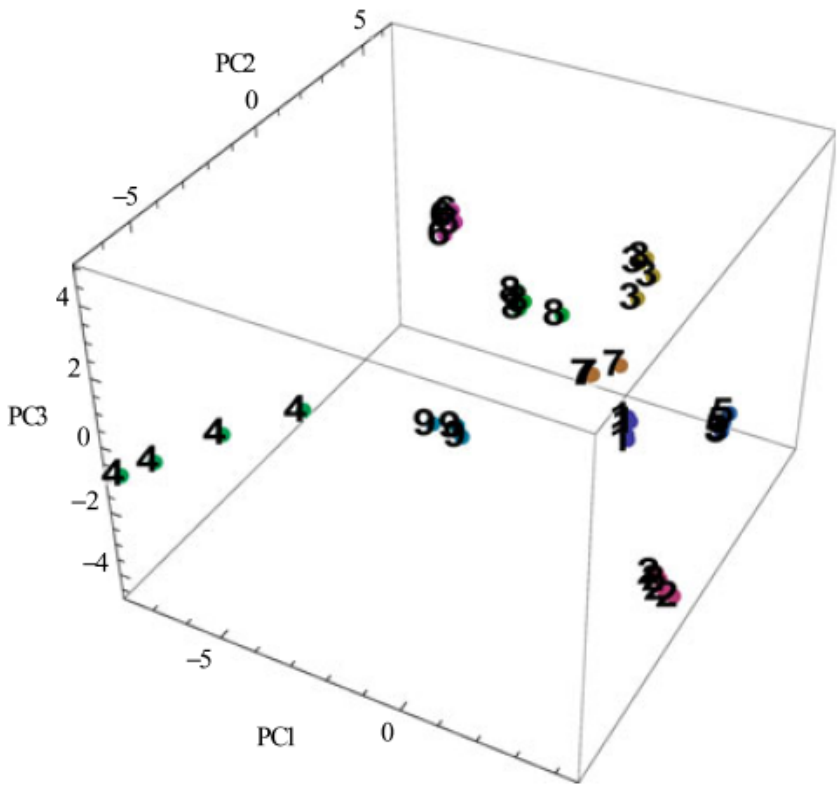


Figure 1.11 3D PCA score plot for the shoe at the later stages of the study. The numbers label the shoes (n=10) used in the study. Reproduced from Petraco *et al* [16].

1.4.4 Application in Environmental Forensic Analysis

In the environment, oil spills can cause great concern for human health in addition to contamination of coasts and estuaries [18]. In a 2004 report by the United States Coast Guard, over 3800 oil spills accounted for more than 1.4 million gallons of oil or fuel released into the environment [18]. Analysis and characterisation of the spilled fuel is therefore important not only to compare the spill sample with known sources but also to identify the responsible party [18]. Statistical and chemometric methods can and has been used to group fuel samples based on chemical similarities by performing pattern matching of the analytical results using statistical means. Such techniques allow a more objective comparison to be conducted.

Hupp *et al* [18] investigated diesel fuels and used PCA to identify natural clusters of similar diesels and determining their most discriminatory chemical compounds. Figure 1.12 illustrates the 3D PCA score plot of the Total Ion Chromatogram (TIC) for the diesel samples used in the study.

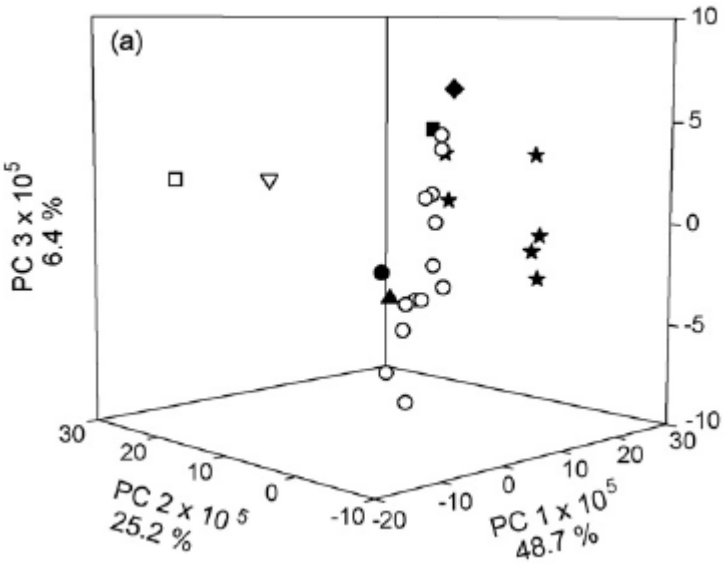


Figure 1.12 Score plot for the TIC. Symbols represent diesels used in the study. Reproduced from Hupp *et al* [18]

Figure 1.13 illustrates the loading plots for this data showing the variances contributed by each compound identified in the samples. Compounds that contribute most are those with large variances (either positive or negative). The PCA facilitates the clustering of chemically similar diesels where qualities of a sample and cluster of samples can be identified based on common traits [18]. It also provides a measure of the association of diesels of similar origin and discrimination of diesels of different origin [18].

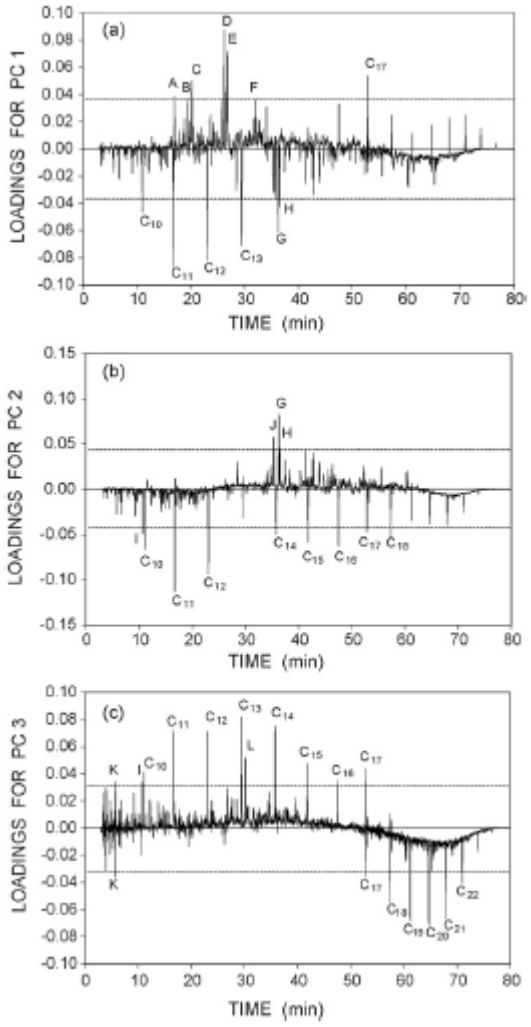


Figure 1.13 Loading plot for PC1, PC2 and PC3 for the TIC. Characters indicate compounds identified in the study whilst carbon number indicates linear alkanes. Reproduced from Hupp et al [18]

1.5 REFERENCES

1. Webb, A., *Statistical Pattern Recognition*. 2nd Ed. 2002, West Sussex: John Wiley and Sons Ltd.

2. Lavine, B.K., *Chemometrics.* Analytical Chemistry, 2000. **72**(12): p. 91R - 97R.

3. Everitt, B.S. and Dunn, G., *Applied Multivariate Data Analysis*. 2001, London: Arnold.

4. Hong, Y.S.T., Rosen, M.R. and Bhamidimarri, R., *Analysis of Municipal Wastewater Treatment Plant using a Neural Network-Based Pattern Analysis.* Water Research, 2003. **37**: p. 1608 - 1618.

5. Brereton, R.G., *Chemometrics for Pattern Recognition*. 2009, West Sussex: John Wiley and Sons Ltd.

6. Tagori, N., Kobayashi, A. and Aishima, T., *Pattern Recognition Applied to Gas Chromatographic Profiles of Volatile Components in Three Tea Categories.* Food Research International, 1995. **28**(5): p. 495 - 502.

7. Everitt, B., *Cluster Analysis*. 3rd Ed. 1993, London: Arnold.

8. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L., *Multivariate Data Analysis*. 6th Ed. 2006, New Jersey: Pearson Prentice Hall.

9. Neumann, H., *Comparison of Heroin by Capillary Gas Chromatography in Germany.* Forensic Science International, 1994. **69**: p. 7 - 16.

10. Besacier, F., Thozet, H.G., Rousseau-Tsangaris, M., Girard, J. and Lamotte, A., *Comparative Chemical Analyses of Drug Samples: General Approach and Application to Heroin.* Forensic Science International, 1997. **85**: p. 113 - 125.

11. Thanasoulias, N.C., Parisis, N.A. and Evmiridis, N.P., *Multivariate Chemometrics for the Forensic Discrimination of Blue Ball-Point Inks Based on their Vis Spectra.* Forensic Science International, 2003. **138**: p. 75 - 84.

12. Adam, C.D., Sheratt, S.L. and Zholobenko, V.L., *Classification and Individualisation of Black Ballpoint Pen Inks using Principal Component Analysis of UV-Vis Absorption Spectra.* Forensic Science International, 2008. **174**: p. 16 - 25.

13. Tan, B., Hardy, J.K. and Snavely, R.E., *Accelerant Classification by Gas Chromatography/Mass Spectrometry and Multivariate Pattern Recognition.* Analytica Chimica Acta, 2000. **422**: p. 37 - 46.

14. Bodle, E.S. and Hardy, J.K., *Multivariate Pattern Recognition of Petroleum Based Accelerants by Solid Phase Microextraction Gas Chromatography with Flame Ionisation Detector.* Analytica Chimica Acta, 2007. **59**: p. 247 - 254.

15. Petraco, N.D.K., Gil, M., Pizzalo, P.A. and Kubic, T.A., *Statistical Discrimination of Liquid Gasoline Samples from Casework.* Journal of Forensic Science, 2008. **53**(5): p. 1092 - 1101.

16. Petraco, N.D.K., Gambino, C., Kubic, T.A., Olivio, D. and Petraco, N., *Statistical Discrimination of Footwear: A Method for the Comparison of Accidentals on Shoe Outsoles Inspired by Facial Recognition Techniques.* Journal of Forensic Sciences, 2010. **55**(1): p. 34 - 41.

17. Ciagnis, C., Tsantili-Kakoulidou, A. and Theocharis, S., *Quantitative Structure Activity Relationship (QSAR) Methodology in Forensic Toxicology: Modelling Postmortem Distribution of Structurally Diverse Drugs Using Multivariate Statistics.* Forensic Science International, 2009. **190**: p. 9 - 15.

18. Hupp, A.M., Marshall, L.J., Campbell, D.I. and Smith, R.W., *Chemometrics Analysis of Diesel Fuel for Forensic and Environmental Applications.* Analytica Chimica Acta, 2008. **606**: p. 159 - 171.

19. Cserhati, T., *Multivariate Methods in Chromatography*. 2008, Chippenham, Wiltshire: John Wiley and Sons Inc.

CHAPTER 2: INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS (ANN) TECHNIQUES

2.1 INTRODUCTION

This chapter describes the theoretical background of the Artificial Neural Networks (ANNs) techniques used for data analysis in this study which are the unsupervised Self Organising Feature Maps (SOFM) and the supervised Multi-Layer Perceptron (MLP) ANNs.

2.2 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) is one of the many branches of the artificial intelligence (AI) discipline. It is a powerful technique that mimics the very basic function of the human brain [1, 2]. In their early stages, ANNs were developed using electrical circuits but with the emergence of computer technology, the interests in this field were channelled into developing suitable functions and codes using computer programming software such as Visual Basic, C++ and Pascal.

Unlike conventional computer programmes which execute specific functions, ANNs are often considered as a generalised learning machine, in other words a machine that can learn almost anything [3]. Although ANNs can be considered relatively new in their application to datasets of forensic relevance, the technique was developed around thirty years ago. Many scientific disciplines including engineering, business, ecology, water research and medicine have employed ANNs to resolve complex problems in these fields [4].

ANNs were first reported in the literature in 1943 when McCulloch and Pitts wrote their first paper on the systematic study of artificial neural networks [2]. In a later work, they explored network paradigms for pattern recognition using a simple neuron model [2]. Following this, in 1949, Hebb proposed a learning law that became the starting point for the ANNs training algorithms [2]. Early success of ANNs inspired other researchers

such as Minsky, Widrow and Rosenblatt [2] to develop networks consisting of a single layer of artificial neurons known as a perceptron and ANNs were applied in such diverse areas for weather prediction, electrocardiogram analysis and artificial vision [5]. However, Minsky and Papert [2] later on demonstrated some limitations of the perceptron, one of them was the inability of the perceptron to solve the simple exclusive-OR (X-OR) function. As a result of this, many researchers abandoned research in the field. In 1986 when Rumelhart, Hinton, Williams [2] developed a clear description of the back-propagation (BP) algorithm for the systematic training of ANNs which resolved the limitations described by Minsky and Papert.

2.2.1 The Biological Neurons

Human biological neural networks, such as the central nervous system, are constantly actively receiving signals or messages from muscles or other senses [3]. In the process of learning new skills, certain connections of neurons in the nervous system of the brain will receive frequent signals or messages as the result of the learning process. Over time these connections are strengthened and reinforced. Any type of signal or message can be sent either repeatedly or only once to the brain and it will respond appropriately to the signals.

An average brain contains approximately 100 billion neurons, each of which has 1000 to 10000 connections with other neurons [6]. These neurons are organised in a fully connected network. The neurons are like other cells in the body however they have special capabilities of receiving, processing and transmitting electrochemical signals over the pathways of the brain's communication system. A typical neuron consists of dendrites, a cell body and synapses as shown in Figure 2.1.

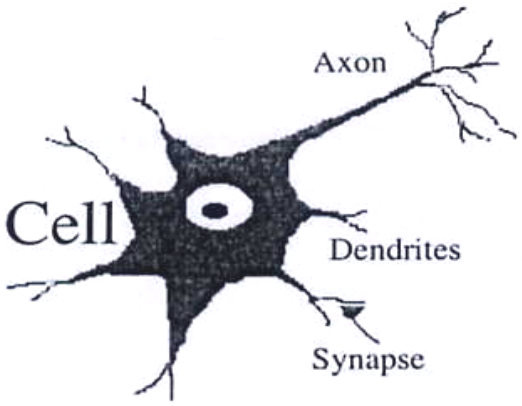


Figure 2.1 The Biological neuron. Reproduced from Kustrin [6].

The function of dendrites, which extend from the cell body to other neurons, is to receive signals or messages at a synapse. The synapse is a connection point between the neurons. The signals received at this point are channelled to the cell body where they are combined. Some signals tend to excite the cell body while others tend to inhibit excitation. Firing of the cell body happens when cumulative excitation exceeds a certain threshold, sending a signal or output through the axon connecting to other neurons. This is the most basic functional outline of a neuron, although in reality, the actual function may involve many complexities. Firing a signal is the only brain function model by the ANNs.

2.2.2 Perceptron

The basic element of ANN is known as a perceptron. As its function is to process information, it is also known as a processing element (PE). In general, a perceptron consists of a series of inputs which are equivalent to the signals (outputs) from other neurons channelled to the synapses in biological neurons. The inputs are given corresponding weights or coefficients, which are equivalent to the synaptic strength of the biological synaptic connection.

The weighted inputs are summed up in the summation block, which is analogous to the cell body to determine the activation level of the perceptron. The summed weighted inputs are then transformed by a transfer function to produce the output of the perceptron. Figure 2.2 show a typical architectural arrangement of a perceptron where X_i represents the inputs signals or patterns while W_i represents the weighting factors given to the inputs or patterns and ***y*** represents the perceptron output.

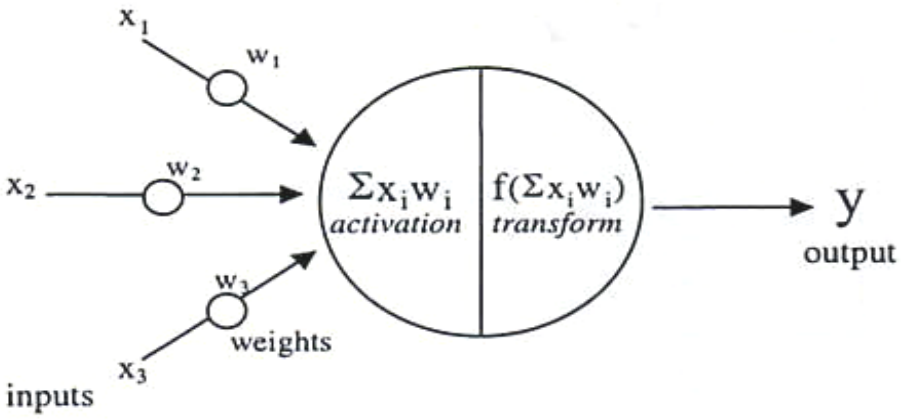


Figure 2.2 Architectural arrangement of a perceptron. Reproduced from Kustrin [6].

2.3. ANNs Learning Schemes

There are two types of learning schemes available for ANNs, namely supervised and unsupervised learning. In supervised learning, both input and output signals are required in the learning process where the output signals act as the target outcomes. With as unsupervised learning scheme only input signals are required in the learning process. An example of as ANN based on a supervised learning scheme is the Multi Layer Perceptron (MLP) or back-propagation network, and an example of ANN based on an unsupervised learning scheme is the Self Organising Feature Maps (SOFM) or Kohonen network.

2.3.1 Self Organising Feature Maps (SOFM)

The Self Organising Feature Maps (SOFM) neural network was first introduced by Kohonen in 1982 [7-9]. It represented the prototype for the unsupervised pattern recognition ANN and perhaps the most popular unsupervised pattern recognition ANNs by far [8]. As in other unsupervised learning frameworks, SOFM does not require a prior knowledge of the output patterns and uses only the input patterns or input vectors and draws from these, for example class membership of the input patterns. To achieve this, two basic assumptions are made; firstly the membership of each class or variable is defined by the users and is an input pattern that shares common features and secondly, the network will be able to identify these common features across the range of the input vectors [4].

The SOFM neural network is normally arranged in a two-layer system which consists of an input and an output layer. The output layer is normally arranged in a two-dimensional grid consisting of a number of units or neurons where the input patterns describing the objects are mapped. A typical arrangement of an SOFM is shown in Figure 2.3. The input neurons in the input layer are all fully connected to each neuron in the output layer.

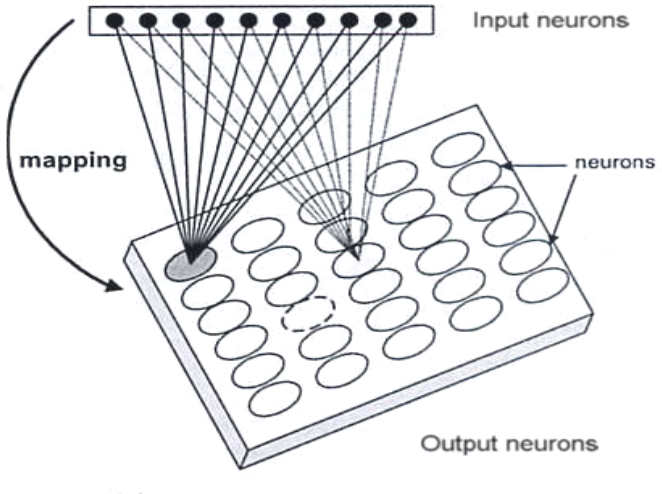


Figure 2.3 Typical arrangement of SOFM. The input neurons act as the input layer and the output neurons act as the output layer. The arrow shows mapping of the input patterns/vectors onto the output layer. Reproduced from Hong [10].

2.3.2 Learning in SOFM

Learning in SOFM is a competitive process in which every time the input patterns/vectors are introduced to the network, all neurons will compete to be stimulated by the input patterns/vectors. Winner neurons are 'picked up' in each learning process and these neurons which are also known as the best matching unit (BMU) have weight vectors similar to those of the input patterns. In SOFM, the similarity between the two vectors i.e. the input and the weight vectors are defined in term of Euclidean distance which is as given in Equation 1.2.

The weights of the BMU's are adjusted according to Equation 2.1. Such an adjustment is necessary so that the weights will become closer to the input vectors in the next iteration or epoch of the algorithm.

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)(x_i(t) - w_{ij}(t)) \quad \text{.....equation 2.1}$$

where $w_{ij}(t+1)$ is the new connection weight from input i to neuron j, $w_{ij}(t)$ is the old or previous connection weight from input i to neuron j and η (t) is the gain term that decreases in time.

The weights of neurons near to the BMU are also adjusted in each epoch although this adjustment will be made in proportion to their distance such that the further they are from the BMU the less adjustment occurs.

As the training process continues, the SOFM is organised into a state whereby similar input vectors are mapped onto similar neurons on the output layer. By the end of the training process, the output neuron is labelled according to the input or object that has stimulated or mapped onto it to reveal if clustering has emerged from the training.

2.3.3 Component Maps

Associated with the output map are component maps whose number corresponds to the number of variables used in the training of the SOFM. The component map is commonly employed to study the relationship of each variable to the position of samples on the output map such that if groupings exist after training of the SOFM, the responsible variables can be identified [11].

In most literature concerning SOFM, the architecture depicted in Figure 2.3 is most commonly, however the architecture described by Zupan [9] as shown in Figure 2.4 is perhaps easier to understand.

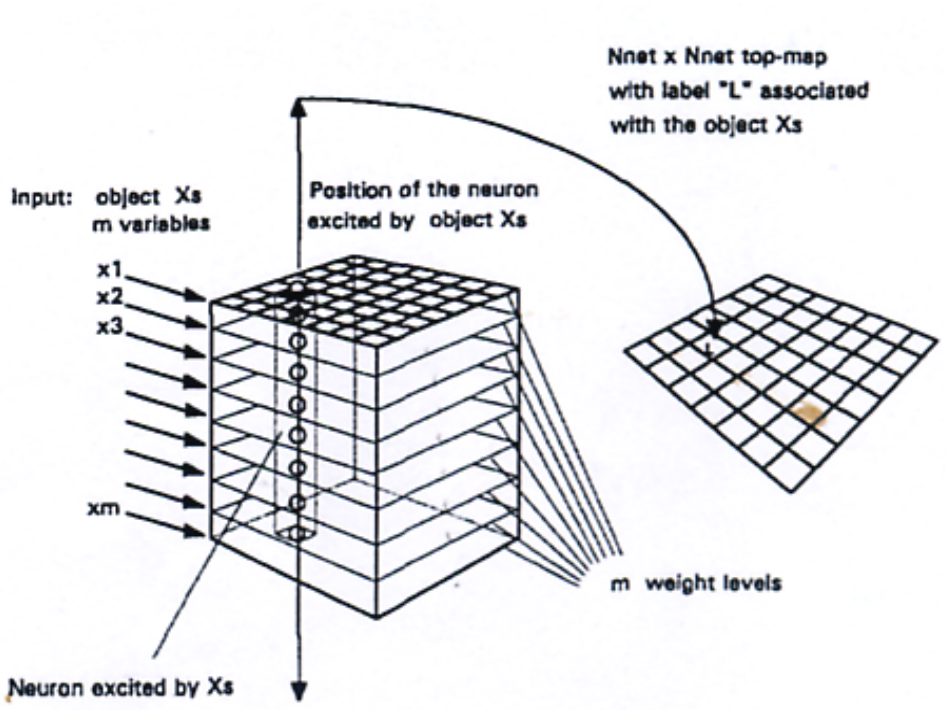


Figure 2.4 Arrangement of SOFM or Kohonen network according to Zupan [9] . Reproduced from Zupan [9]

The number of component maps in a given dataset corresponds with the number of attributes or variables. For example, if there are four attributes describing a given dataset then four component maps are expected. The first component map corresponds with the first attribute; the second component map corresponds with the second attribute and so forth. The arrangement of the component maps in SOFM can be visualised as multi-tier planes as illustrated in Figure 2.5

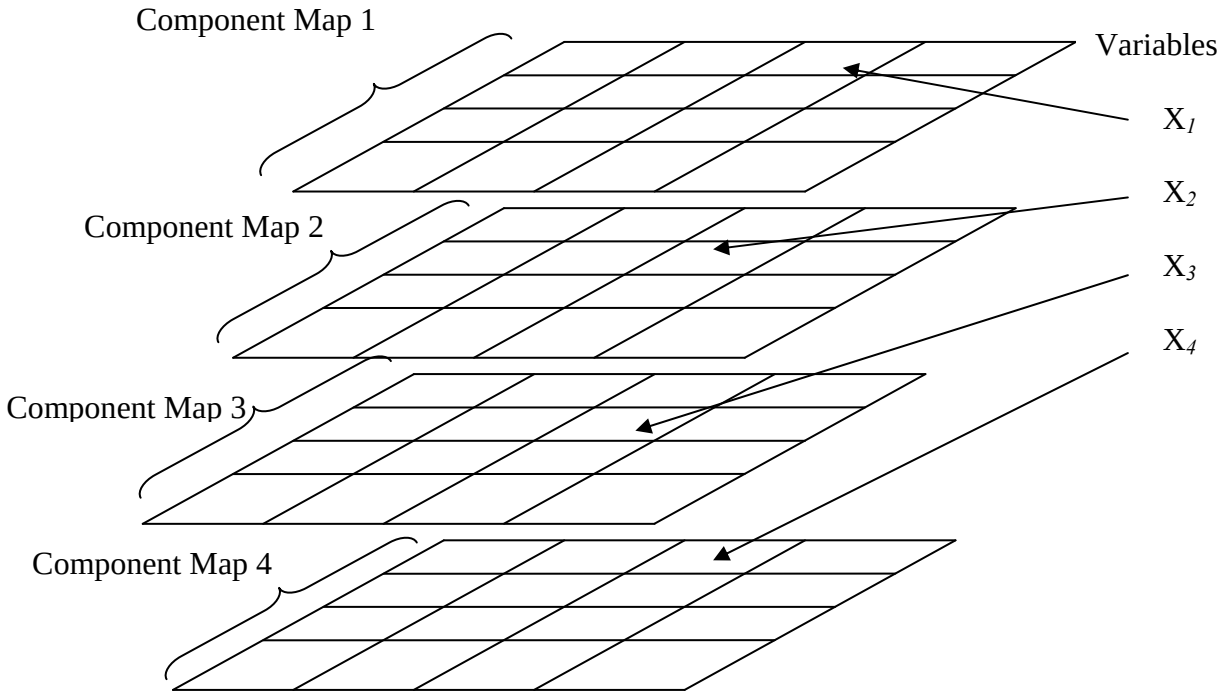


Figure 2.5 Arrangement of component maps in SOFM

To further understand the concept of component maps in SOFM, a trivial dataset can be considered. Consider a collection of sixteen animals based on the presence (= 1) or absence (= 0) of thirteen attributes (variables); small, medium, big, two legs, four legs, hair, hooves, mane, feathers, hunt, run, fly and swim . The dataset was initially used by Kohonen to describe the concept of SOFM [7].

The output layer of the dataset is given in Figure 2.6 which shows the distribution of the animals across the given attributes. After training, groupings of the animals according to similarity are apparent in the output layer.

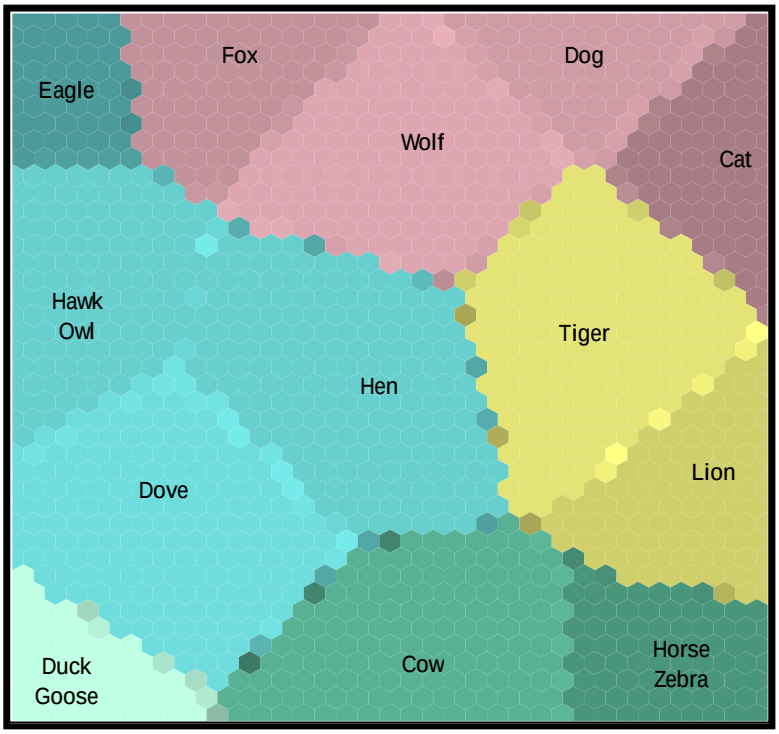


Figure 2.6 Output map of the animal dataset

Animals that belong to a bird family occupy the left part of the output map and animals that belong to a cat or dog family occupy the right middle and upper part of the output map whilst animals such as horse, zebra and cow occupy the lower right part of the output map. The thirteen component maps (one for each attribute) are presented in Figure 2.7. The colours (red, green and yellow, and blue) are used to represent the highest, intermediate and lowest magnitude across the component map. To a certain extent, the colours also reflect the magnitude of an attribute within the dataset and can be used to study the relationship between the animals for example duck and goose can both swim as illustrated by the swim component map in Figure 2.7.

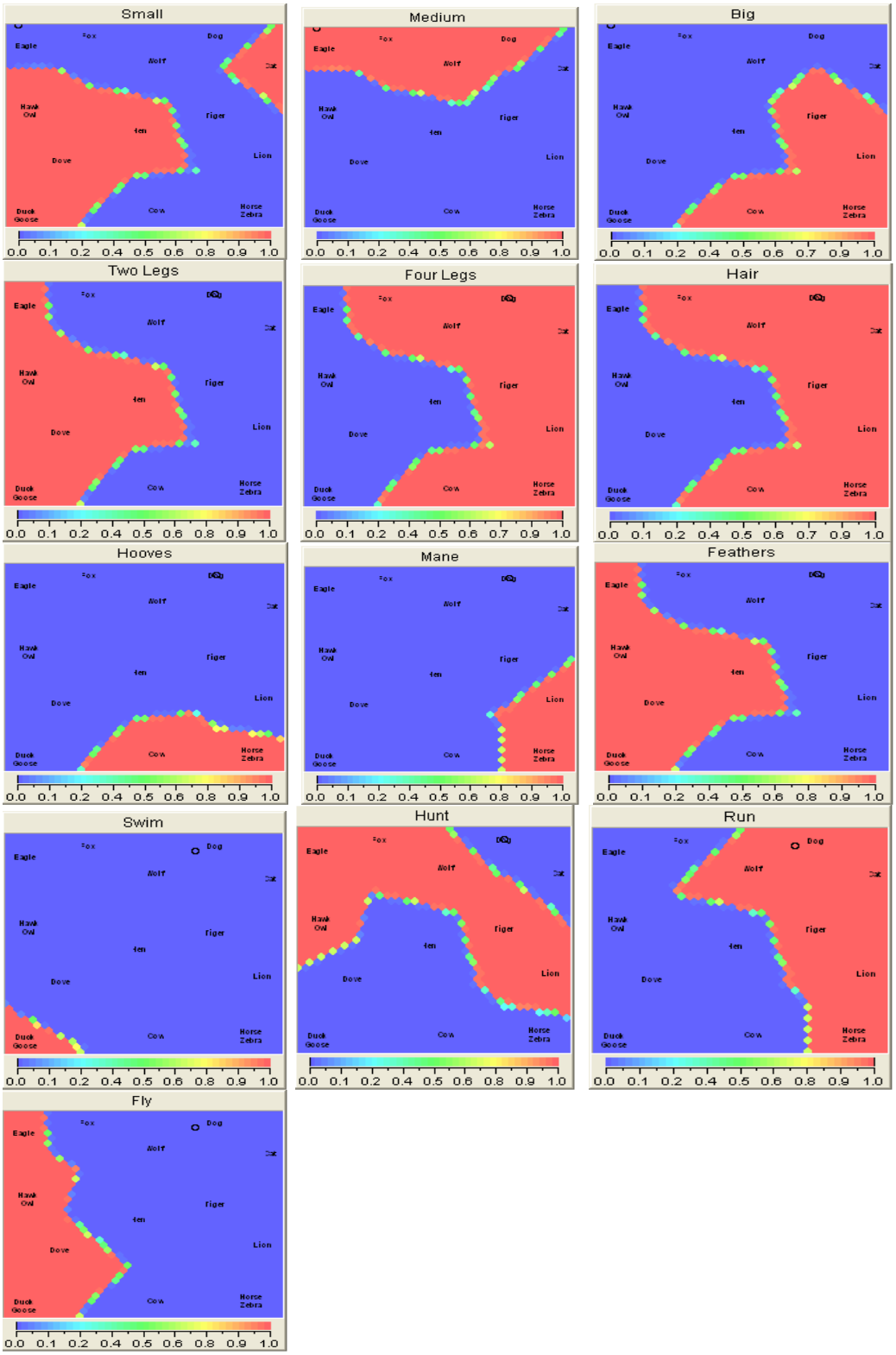


Figure 2.7 Component maps associated with the output map of the animal dataset

2.4 Theoretical Background to Multi-Layer Perceptron Neural Networks

Multi-Layer Perceptron (MLP) represents a supervised pattern recognition neural network model [12] and is the most popular neural network applied in areas involving classification and prediction [8, 13]. The most common learning algorithm used is known as a back propagation algorithm creating a back-propagation neural network (BP-NN) [2, 12]. Architecturally, MLP is organised in a three-layer form as illustrated in Figure 2.8.

The first layer is known as the input layer, and consists of neurons whose sole purpose is to distribute input vectors (patterns) to the neurons in the second layer. The second layer is the hidden layer and contains hidden neurons whose purpose is to 'process' the input vectors. The 'processed' input vectors are then transferred to neurons in the output layer, which is the final layer of the MLP.

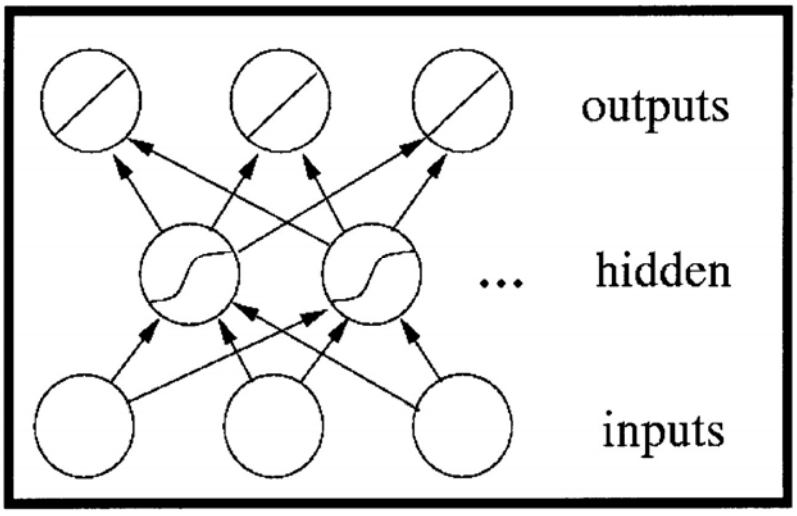


Figure 2.8 Typical arrangement of MLP. Reproduced from

Every neuron in MLP is connected to all neurons in adjoining layers rather than neurons within their own layer [2]. MLP is a fully connected feed-forward network, which means that information only flows from layer to layer rather than between neurons within the

individual layers. Associated with each neuron to neuron connection is a weight which is an adjustable parameter and plays an important role in the learning process of the MLP.

2.4.1 Learning in MLP

Learning in MLP involves a process of adjusting the connecting weights between the neurons using a systematic step-by-step approach or algorithm. For supervised pattern recognition frameworks, such as MLP, a specific desired output, such as a specific sample classification, is required.

In its simplest sense, the weighted input vectors from the input neurons/layer are propagated to the hidden neurons in the hidden layer where they are processed, through a transfer function, to produce the actual output of the network. The most common transfer function used for MLP is the sigmoid transfer function which is shown in Equation 2.2. The purpose of this transfer function is to ensure that the network output lies between one and zero [5] in order to give a non-linearity property to the MLP so that more complex boundaries can be achieved.

$$f(OUT_j) = \frac{1}{1 + e^{-NET_j}} \quad \text{..... equation 2.2}$$

where OUT_j represents the output from neuron j and NET_j is the summation of all weighted input signals of neuron $_j$ and is determined using Equation 2.3.

$$NET_j = \sum_{i=1}^{n} X_i W_{ij} \quad \text{..... equation 2.3}$$

where X_i is the input for neuron i and W_{ij} is the connection weight from input i to neuron j.

Another important transfer function employed in MLP is the hyperbolic tangent function. This function performs in a similar way to the sigmoid transfer function except that its output lies in the range of -1 to 1. The hyperbolic tangent function is given in Equation 2.4.

$$f(OUTj) = \frac{e^{NETj} - e^{-NETj}}{e^{NETj} + e^{-NETj}} \quad \text{..... equation 2.4}$$

Where *NETj* is given in Equation 2.4 above

2.4.2 Weight Adjustment – Back Propagation

Weight adjustment is perhaps one of the most interesting aspects in a neural network and can be regarded as the core part of the so called 'knowledge acquisition' process. In MLP, once the actual output is obtained after passing the input vectors to the network, it is then compared to the desired output.

The difference between the actual and desired output is the network error which is back-propagated layer by layer. The error is used to adjust the connection weights in every layer (if it is deemed necessary) before the input vectors are once again passed through the network. These processes are repeated until the difference in error between the two outputs reaches a low level and each input vector produces an output close to or the same as the desired output [2]. This kind of weight adjustment follows a gradient descent approach where adjustment is made by following the error surface toward a minimum [2]. The connection weights are adjusted according to a back-propagation rule (also known as the generalised delta rule), which is given in Equation 2.5;

$$\Delta w_{ji}(n) = \eta \delta_{pj} O_{pi} + \alpha \Delta w_{ij}(n-1) \quad \text{..... equation 2.5}$$

where Δw_{ij} is the adjustment in connection weight between neuron i in the hidden layer and neuron j in the input layer, η and α represent the learning rate and momentum respectively and n and $(n\text{-}1)$ represent present and previous iteration cycles or epochs respectively. In ANN sense, an epoch refers to one pass of all objects through the network [8]. Figure 2.9 which was reproduced from Reibnegger [12] summarises the learning process of MLP.

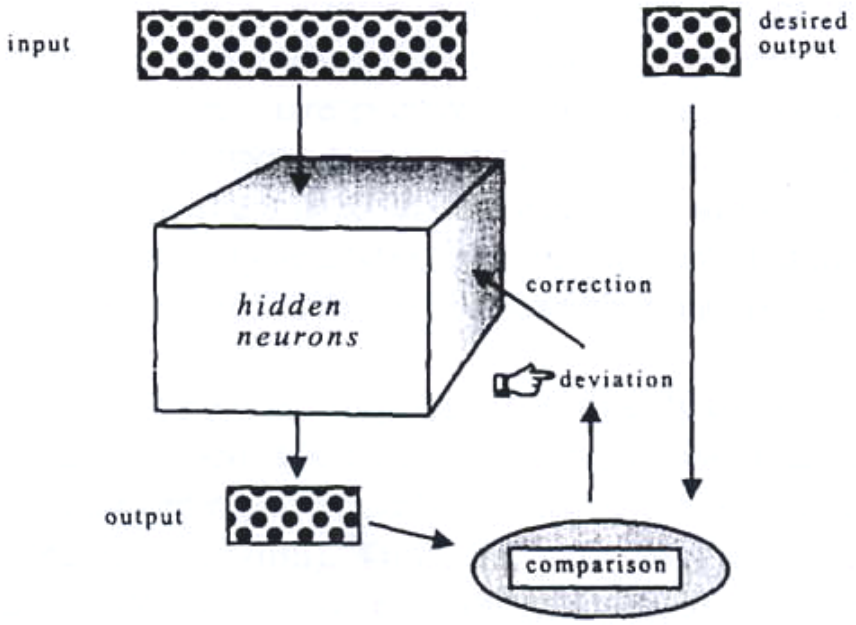


Figure 2.9 Learning process of MLP. Reproduced from Reibnegger [12]

2.4.3 Other Learning Paradigms

The standard back propagation algorithm operates in a batch mode, which means that the connection weights are adjusted or updated only after all input vectors are introduced to the network. Such an approach can be very slow, especially when there are a large number of input vectors. An alternative to this is to utilise an algorithm which operates in an incremental manner which adjusts or updates the connection weights after each input vector is introduced to the network. There are a number of algorithms which

operate in such an incremental fashion but for the purpose of this study only conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithms were investigated.

2.4.4 General Implementation Guidelines for MLP

There are a number of aspects that need to be considered when implementing an MLP neural network. These are the size of the network, the number of input and output neurons and the number of hidden neurons in the hidden layer.

2.4.4.1 Size of Network

Although networks with two or even three hidden layers are used, networks with a single hidden layer are normally sufficient for most applications [2]. These are faster to train because the adjustment of their connection weights is simpler thus requiring a minimal amount of training time.

2.4.4.2 Number of Input and Output Neurons

The number of input neurons in MLP can be easily determined and normally corresponds to the number of features or variables describing the objects or samples used in the training process. However, if the number of features describing the objects or samples is large, a large amount of training time is required in order to determine the 'correct' weights for all connections [14]. If this is the case, then it is possible to transform the data before application to the network.

2.4.4.3 Number of Hidden Neurons in the Hidden Layer

The number of hidden neurons in the hidden layer will greatly influence the performance of a given network. Unfortunately there is no standard rule regarding the number of hidden neurons and normally this is dependent on experimentation. The most commonly employed technique is to train a number of networks by varying the number of hidden neurons (starting with few and increasing the number) that are trained [13] and evaluated the results sequentially. A network with either too few or too many neurons in the hidden layer does not necessarily produce a good performance. Too few hidden

neurons will hinder the learning process and too many will depress prediction ability through overtraining [6].

2.4.4.4 Generalisation

Once trained correctly, networks can give reasonable answers when presented with inputs that they have not previously encountered. A new input vector can give an output similar to the correct output for input vectors used in the training process if these are similar to the new input vector introduced. This is known as ‘generalisation’ and makes it possible for networks to be trained only with a representative set of input and output vectors and the ability to generalise for new input vectors not previously encountered make MLP a suitable candidate for use as a predictive tool.

2.4.4.5 Over-Learning

Over-learning (or over fitting) is a situation which can occur when a network becomes ‘specialised’. This can happen if the network has ‘memorised’ all points within the training dataset sample space. Over fitting is not a preferable state in a neural network because an over fit network will not generalise well. Factors that can contribute to over fitting are using a limited training dataset and having too many hidden neurons in the hidden layer. Having said that, this does not mean that large training datasets and networks with small numbers of hidden neurons should be used in preference because the former can overly complicate a network and the latter may not learn the input vectors well enough to be able to produce the desired output.

The concept of over fitting is illustrated in Figure 2.10. In this case fitting a low order polynomial curve will not explain the data. A higher power polynomial curve, with a greater curvature such as illustrated in (a) fits the data to a much greater extent however fitting a polynomial with even greater curvature will result in a curve that is too flexible as in (b) which fails to describe the underlying function of the data.

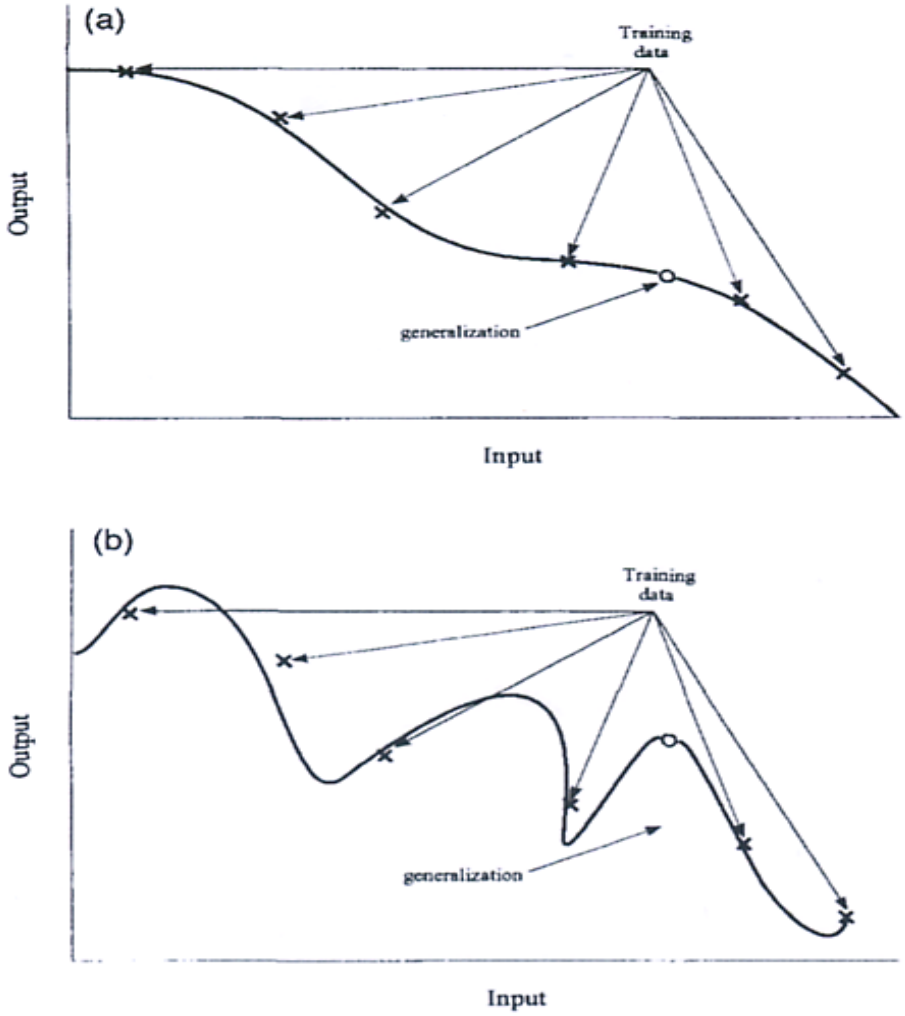


Figure 2.10 Principle of over learning, a) the curve properly fitted the data (good generalisation) b) the curve over-fitted the data (poor generalisation). Reproduced from Svozil [13].

2.4.4.6 Overcoming Network Over-Learning

Two techniques have been devised to overcome network over-learning (over fitting) known as early stopping and regularisation [4-6]. In the early stopping technique, the available data for network training is divided into three subsets normally in the proportion of 2: 1: 1 [8-10]. The first set is known as the training set, which is used to calculate the error and update the network weights. The second set is a validation set whose error is used to monitor the network training progress. The final set is the test set which is used to check on the generalisation ability of the network in other words

whether or not the network is capable of returning the correct or expected outcome for the input vectors not used in the network training. At the beginning of the training stage, both training and validation set errors normally decrease. An indication of over fitting is provided when the validation set error begins to rise. When this happens, training is halted and the weights at the minimum validation error are returned. This is the process which was used in this study.

2.5 Applications of Artificial Neural Networks (ANNs) in Forensic Science

Although the idea of ANNs has been around for almost 30 years since it first introduction in 1986 by Rumelhart, Hinton and Williams [2], its applications in forensic science are surprisingly limited compared to the 'conventional' pattern recognition techniques eventhough the concepts have been reported for decades [1]. Of the two ANNs techniques i.e. SOFM and MLP, the latter has found much wider applications. Amongst the areas where MLP has been applied in forensic science are ballistic studies [15], accelerant analysis [16], drugs of abuse [17-19], forensic hair analysis [20] and facial image recognition [21]. SOFM on the other hand has had much less impact in forensic science compared to MLP and to date, the only reported application has been to a dataset related to forensic digital investigation [22]

2.5.1 Application in Forensic Ballistic

MLP neural network consisting of 96 neurons in the input layer, 15 neurons in the hidden layer and one neuron in the output layer was used by Banno [15] to investigate the presence of striated toolmarks on bullets. The input patterns or vectors for the MLP were binary signals of the striation images derived using character-extraction software. The neurons in the input layer (of the MLP) were divided into two blocks i.e. Block A and Block B where each block contained 48 neurons (a total of 96 neurons). Figure 2.11 shows the arrangement of the MLP used in the study. Two patterns to be compared for similarity were inputted separately into the two blocks during the learning process and the MLP was tuned to give an output of 1 for similar input patterns and an output of near 0 for non-similar patterns.

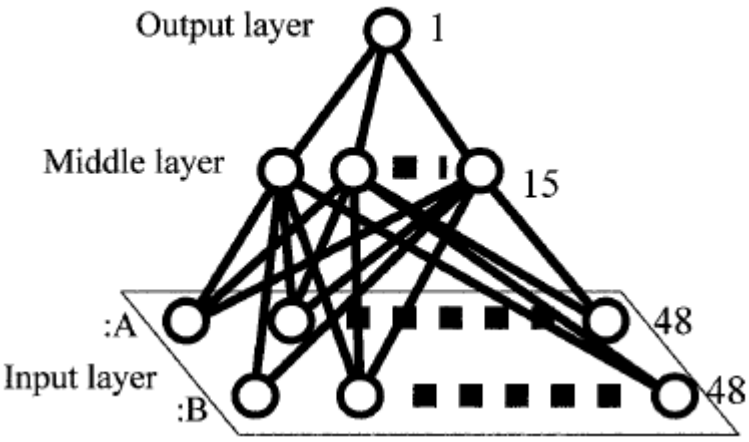


Figure 2.11 Structure of the MLP neural network used in the study. Reproduced from Banno [15]

Only 10 bullets were used in the study facilitating only100 comparisons to be made. An example of an output score for a similar pattern is illustrated in Figure 2.12. Although the MLP had demonstrated good outcomes by giving higher scores for two bullets fired from the same firearm, a much higher number of bullets were needed in order to validate the MLP algorithm.

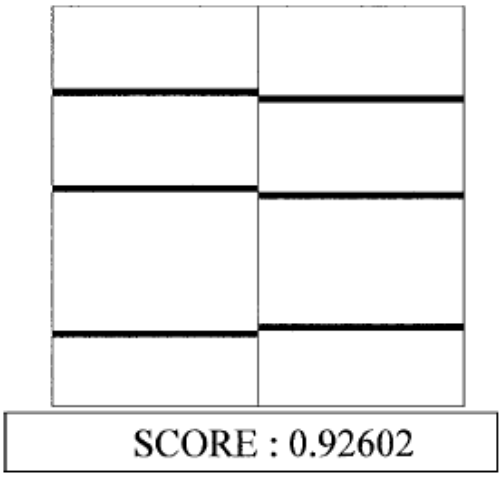


Figure 2.12 An example of output score generated by the MLP neural network

2.5.2 Application in Drugs of Abuse

Prior to the study conducted by Casale *et al* [17] in 1993, chromatographic pattern analyses was normally performed by manual pattern matching which are laborious, tedious, time consuming and potentially inaccurate [17]. Casale *et al* [17] suggested the use of an MLP neural network for cocaine pattern comparisons due to its ability to solve problems related to pattern-recognition. The input patterns used in the study were the chromatographic profiles of cocaine containing 16 quantificated cocaine related impurities. All together, 3426 different cocaine chromatographic patters (compiled into one large database) were used as the training set for the MLP neural network. An MLP neural network made up of 16 neurons in the hidden layer and 16 neurons in the output layer was found to be more superior to distance functions classifications techniques i.e. nearest neighbour and clustering. In the study, all chromatographic patterns were correctly recognised and matched [17]. Figure 2.13 illustrates the cocaine pattern matching procedure. The findings demonstrated that the MLP neural network is a reliable means to rapidly establishing common source identity of cocaine samples. The MLP neural network using a PC based software, is now being used on a daily basis at the North Carolina State Bureau of Investigation (NCSBI) to aid forensic experts identifying cocaine samples that potentially originate from the same batch.

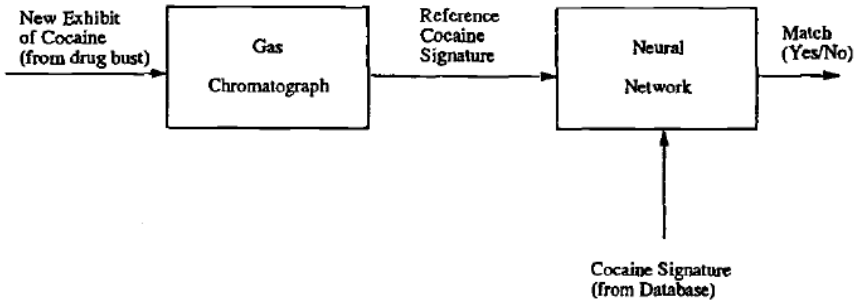


Figure 2.13 The cocaine signature matching procedure utilised at the NCSBI. Reproduced from Casale *et al* [17]

The Institute de Police Scientifique (IPS) at the University of Lausanne also utilised artificial neural network (ANN) to validate illicit heroin classification methods developed at the institute [19]. The ANN was employed along with PCA where the latter was used as a preliminary screening method to identify samples of similar chemical profiles. Correlation values (for each sample identified by the PCA) were then calculated in order to determine links between the heroin samples which were then recorded in a database known as Ibase® [19] and heroin samples with similar chemical profiles were grouped together. In total 20 chemical classes were identified. The input patterns used in the training (of the ANN) were the normalised peak areas of six target compounds (meconine, acetylcodeine, acetylthebaol, 6-monoacetylmorphine, papaverine and noscapine). In the training of the ANN, 468 samples were used as the training set, 60 samples as blind samples and 370 samples as non-linked samples. The best performing ANN correctly classified 96.6% of the test sample was a neural network with 6 input neurons in the input layer, 121 neurons in the hidden layer and 20 output neurons in the output layer which corresponded to the 20 chemical classes under study. The arrangement of the best performing ANN is as illustrated in Figure 2.14. The findings of the study indicated that chemical class determination could be successfully achieved using ANN [19].

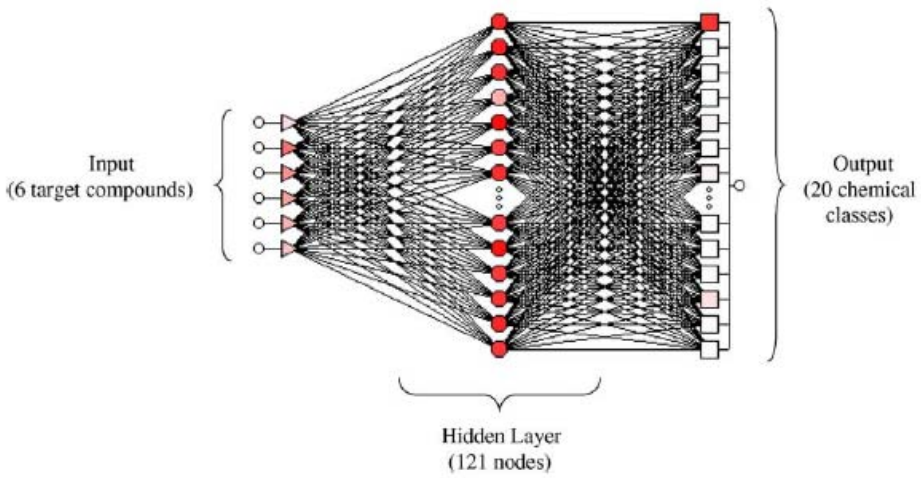


Figure 2.14 Arrangement of the best performing network. Reproduced from Esseiva *et al* [19]

2.5.3 Application in Forensic Hair Analysis

An automated forensic hair analysis and comparison system using a MLP neural network was studied by Verma *et al* [20] in 2002. The system inputs were microscopic images of two hairs which produced an output as to whether or not the hairs came from the same person. The hair images were captured using a NEXTDimension^{TM} video board in a NEXTDimension^{TM} colour turbo computer connected to a video camera [20]. Two image processing methods were undertaken, the wavelet analysis (for the medulla) and Halarick texture algorithm (for the cortex). Since the medulla and cortex are two different uncorrelated features of the hair, two separate MLP neural networks were trained for each feature. Figure 2.15 and 2.16 shows the process of creating the medulla and cortex network respectively.

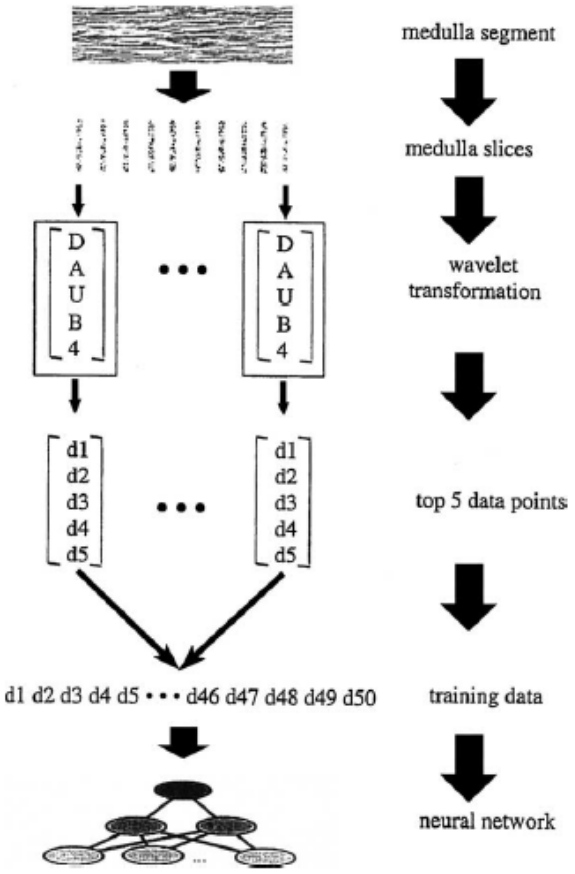


Figure 2.15 The process of creating the medulla network. Reproduced from Verma *et al* [20]

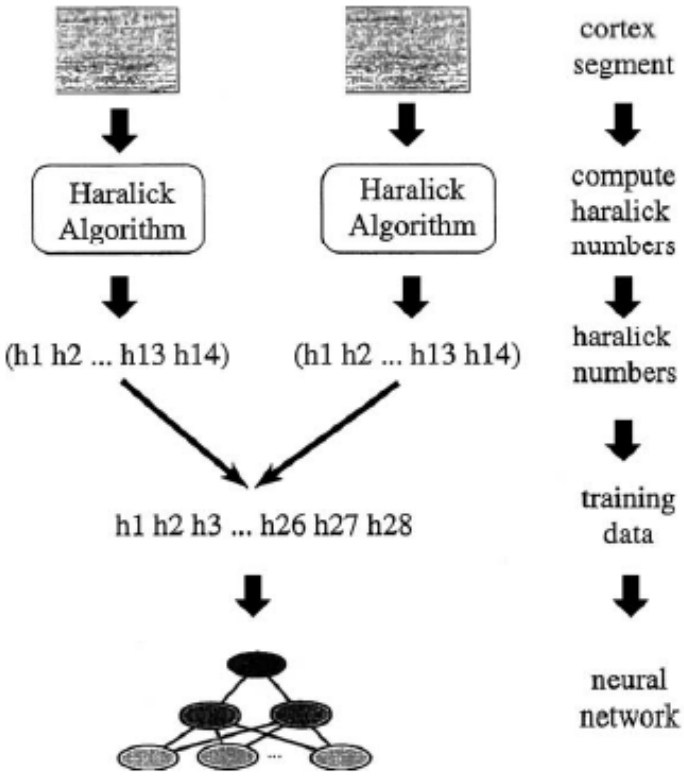


Figure 2.16 The process of creating the cortex network. Reproduced from Verma et al [20]

The use of the wavelet analysis and the Halarick texture algorithm to pre-process data facilitated the collection of a large amount of data acquired as images to be compressed into a much smaller but representative dataset [20]. As a result of this, the best performing ANN achieved an accuracy of 83% in matching hairs from the same person.

2.5.4 SOFM Application in Digital Forensics

Many employees who have access to a computer have, with the advent of the internet, been given the opportunity to explore new and interesting possibilities of the World Wide Web [22]. Having said that, excessive internet usage for non-work purposes for examples accessing websites that promote pornography and other illegal activities have become a risk [22]. To detect such activities, a study using a SOFM neural network coupled with a computer application 'Forensic Toolkit' was investigated by Fei *et al*

[22] at Pretoria University in 2006. The Forensic Toolkit itself is a computer application used to create a text file containing information about all files found in the temporary internet files folder. This data was then processed using SOFM. Four computer systems with four different users were the subjects of the study. For each computer system, three SOFM component maps were created where the first component map represented the file type; second component map represented the time when the temporary internet files were created and the last component map represented the day of the week on which the temporary files were created. Figure 2.17 shows the component maps generated from one of the computer systems used in the study.

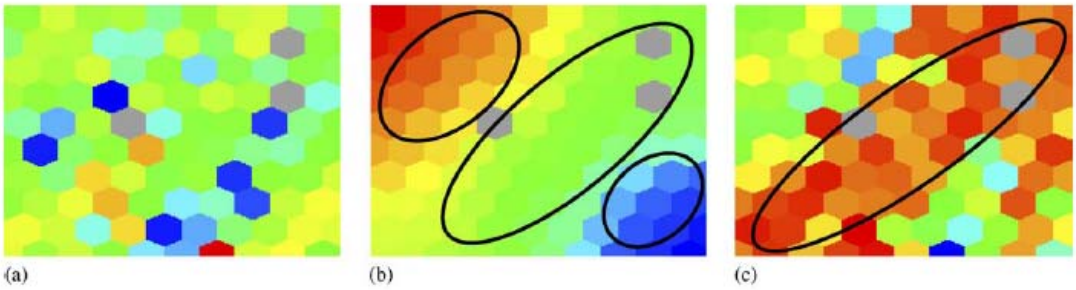


Figure 2.17 SOFM component maps generated from one of the computer systems used in the study. (a) - first component map, (b) - second component map and (c) – third component map. Reproduced from Fei *et al* [22]

Based on the component maps generated for each computer system, the study indicated that anomalous behaviour was detected in one of the computer system. This was due to the irregularities in the 'behaviours' of its component maps compared to the other computer systems. The findings of the study demonstrated that SOFM could be of use to digital forensic investigators conducting investigation on anomalous behaviour of internet users within an organisation.

2.6 REFERENCES

1. Kingston, C., *Neural Networks in Forensic Science.* Journal of Forensic Sciences, 1992. **37**(1): p. 252 - 264.

2. Wassermann, P.D., *Neural Computing Theory and Practice*. 1989, New York: Von Nostrand Reinhold.

3. Cartwright, H.M., *Applications of Artificial Intelligence in Chemistry*. 1993, Oxford: Oxford Science Publications.

4. Beale, R. and Jackson, T., *Neural Computing: An Introduction*. 1990, Bristol: Hilger.

5. Patterson, D.W., *Artificial Neural Networks - Theory and Applications*. 1996, London: Prentice Hall.

6. Kustrin, S.A. and Beresford, R., *Basic Concepts of Artificial Neural Network (ANN) Modeling and its Application in Pharmaceutical Research.* Journal of Pharmaceutical and Biomedical Analysis, 2000. **22**: p. 717 - 727.

7. Kohonen, T., *Self Organising Maps*. 1991, Berlin: Springer-Verlag.

8. Zupan, J. and Gasteiger, G., *Neural Networks in Chemistry and Drug Designs*. 1999, Weinheim: Wiley-VCH.

9. Zupan, J., Novic, M. and Ruisanchez, I., *Kohonen and Counterpropagation Artificial Neural Networks in Analytical Chemistry.* Chemometrics and Intelligence Laboratory Systems, 1997. **38**: p. 1 - 23.

10. Hong, Y.S.T., Rosen, M.R. and Bhamidimarri, R., *Analysis of Municipal Wastewater Treatment Plant using a Neural Network-Based Pattern Analysis.* Water Research, 2003. **37**: p. 1608 - 1618.

11. Marini, F., Bucci, R., Magri, A.L. and Magri, A.D., *Artifial Neural Networks in Chemometrics: Histrory, Examples and Persepctives.* Microchemical Journal, 2008. **88**: p. 178 - 185.

12. Reibnegger, G. and Wachter, H., *Self Organising Neural Networks - An Altenative Way of Cluster Analysis in Clinical Chemistry.* Clinica Chimica Acta, 1996. **248**: p. 91 - 98.

13. Svozil, D., Kvasnicka, V. and Pospichal, J., *Introduction to Multi-Layer Feed-Forward Neural Networks.* Chemometrics and Intelligence Laboratory Systems, 1997. **39**: p. 43 - 62.

14. Bishop, C.M., *Neural Networks for Pattern Recognition*. 1995, Oxford: Oxford University Press.

15. Banno, A., *Estimation of Bullet Striation Similarity using Neural Networks.* Journal of Forensic Sciences, 2004. **49**(3): p. 1 - 5.

16. Doble, P., Sandercock, M., Pasquier, E.D., Petocz, P., Roux, C. and Dawson, M., *Classification of Premium and Regular Gasoline by Gas Chromatography/Mass Spectrometry, Principal Component Analysis and Artificial Neural Networks.* Forensic Science International, 2003. **132**: p. 235 - 240.

17. Casale, J.F. and Watterson, J.W., *A Computerised Neural Network Method for Pattern Recognition of Cocaine Signature.* Journal of Forensic Science, 1993. **38**(2): p. 292 - 301.

18. Waddell, R.J.H., NicDaeid, N. and Littlejohn, D., *Classification of Ecstasy Tablets using Trace Metal Analysis with The Application of Chemometric Procedures and Artificial Neural Network Algorithms.* Analyst, 2004. **129**: p. 235 - 240.

19. Esseiva, P., Anglada, F., Dujourdy, L., Taroni, F., Margot, P., Pasquier, E.D., Dawson, M., Roux, C. and Doble, P., *Chemical Profiling and Classification of Illicit Heroin by Principal Component Analysis, Calculation of Inter Sample Correlation and Artificial Neural Networks.* Talanta, 2005. **67**: p. 360 - 367.

20. Verma, M.S., Pratt, L., Ganesh, C. and Medina, C., *Hair-MAP: A Prototype Automated System for Forensic Hair Comparison and Analysis.* Forensic Science International, 2002. **129**: p. 168 - 186.

21. Sinha, P., *A Symmetry Perceiving Adaptive Neural Network and Facial Image Recognition.* Forensic Science International, 1998. **98**: p. 67 - 89.

22. Fei, B.K.L., Eloff, J.H.P., Olivier, M.S. and Venter, H.S., *The Use of Self Organising Maps for Anamalous Behaviour Detection in a Digital Investigation.* Forensic Science International, 2006. **162**: p. 33 - 37.

CHAPTER 3: UNSUPERVISED PATTERN RECOGNITION TECHNIQUES APPLIED TO DATA DERIVED FROM THE ANALYSIS OF WAX BASED PRODUCTS

3.1 INTRODUCTION

This chapter describes the application of the unsupervised 'conventional' pattern recognition techniques of Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) and the unsupervised Artificial Neural Networks (ANNs) technique of Self Organising Feature Maps (SOFM) to datasets derived from the analysis of a variety of lipsticks, lip balms and shoe polishes. Datasets were generated using Thin Layer Chromatography (TLC), Microspectrophotometry (MSP), Ultra-Violet and Visible (UV/Vis) spectroscopy and Gas Chromatography with Flame Ionisation Detector (GC-FID).

3.2 Wax Based Products

Waxes are among the oldest worked materials used by humans [1]. They are used for many different purposes in industrial and consumer products. In dental surgery for instance, waxes are used for moulding and casting patients' teeth. They are also used to preserve the freshness of vegetables and fruits, in pharmaceuticals, as electrical insulations, for manufacture of explosives and pyrotechnics, in cosmetics, polishes, and as lubricants and adhesives [2].

There is no generally accepted definition for the term wax [1]. In some chemistry textbooks, they are defined as esters of long-chain carboxylic acids with long chain alcohols [1] however such a definition only covers beeswax and other natural waxes but not waxes derived from crude petroleum (defined as petroleum waxes).

According to the German Association for Fat Sciences [1], which has produced perhaps the most comprehensive definition of waxes, substances defined as waxes must fulfil the following criteria;

1. Melt at 40°C or greater without decomposition,
2. Polishable under slight pressure and have strongly temperature dependent consistency and solubility,
3. Kneadable at 20°C or hard to brittle, coarse to finely crystalline, transparent to opaque, but not glassy, or highly viscous or liquid,
4. Burn with sooty flame after ignition,
5. Able to form pastes or gels and poor conductors of heat and electricity,

3.2.1 Classification of Waxes

Although waxes can be classified according to their chemical, physical and engineering properties as well as application [1], perhaps the simplest classification is based on the origin of waxes and their synthesis [1].

Waxes can be classified into natural and synthetic waxes as illustrated in Figure 3.1. The former are metabolic products of insects, animals and plants or fossilised materials while the latter are derived from chemical synthesis.

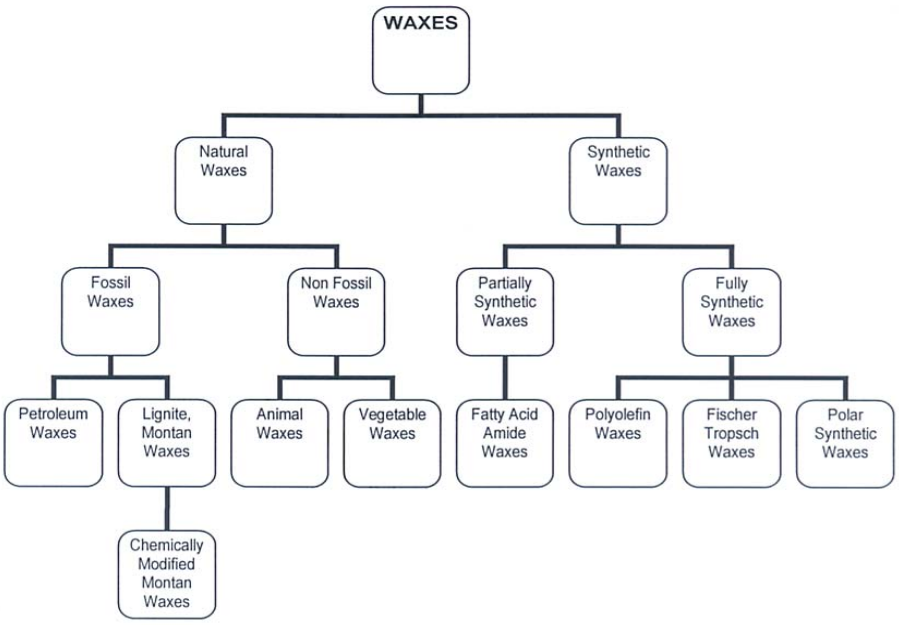


Figure 3.1 Classifications of Waxes. Reproduced from Elvers *et al* [1]

3.2.2 Natural Waxes

The most common natural waxes are derived from bees (hence the term beeswax) or other animals such as lanolin, derived from sheep, goats and sperm whales. Plant waxes, such as carnauba wax, are derived from the carnauba palm, which grows abundantly in parts of South America [1, 2] and fossilised wax, such as paraffin wax, is derived from fractionation of crude petroleum [1, 2]. Natural waxes are commonly used in consumer products such as cosmetics, polishes, cleaning materials and pharmaceuticals [1, 2].

3.2.3 Synthetic Waxes

There are two main groups of synthetic waxes known as Fischer Tropsch and polyolefin waxes. The former is synthesised from methane obtained either from the reaction of steam with carbon or from natural gas [1] whilst the latter is synthesised from ethane and propane [1]. These waxes are used mostly in industry as lubricants in plastic processing, as mould releasing agents, in corrosion protection and in photocopying [1].

3.2.4 Wax Based Products and Their Technology

In consumer products, the waxes are mostly of natural origin (paraffin wax, carnauba wax, candelilla wax and beeswax). To manufacture the end products which meet the needs of consumers, other materials such as dyes, pigments, fragrance and solvents are added in varying proportions to the waxes.

Beeswaxes contain a high proportion of esters of C_{30}, as well as alkanes in the range of $C_{25} - C_{31}$, acids and a small proportion of alcohols. Candelilla waxes contain mostly saturated alkanes in the range of $C_{29} - C_{33}$, esters, alcohols and acids whilst carnauba waxes contain mostly esters and acids. Most commonly, these waxes are blended to give certain desired properties to the final product such as solvent binding power, polishing capability and flexibility [2].

Paraffin waxes are mainly a mixture of straight chain alkanes with minor amounts of branched chain and cyclic alkanes in the range of $C_{20} - C_{35}$ and microcrystalline waxes which is of fossilised origin consist mainly of branched chain alkanes of higher molecular weight than paraffin waxes [3].

3.2.5 The Importance of Wax Based Products in Forensic Science

Wax based products such as lipsticks and shoe polish can sometimes be found at crime scenes as stains or smears on clothing, cigarette butts, glasses and cups. The presence of such smears may be indicative of direct physical contact resulting in the transfer of the material. Smears of lipstick and shoe polish have been reported as being encountered in cases such as assault [4] burglary [5], sexual assaults [6] and homicides [7]. Like any other transfer evidence, analysis of such stains may suggest potential sources and eventually may help to provide a link to a suspect and as such can be of considerable value depending on the circumstances of the case [8].

3.2.6 Lipsticks

Lipsticks, as their name suggests, are moulded in the form of crayons and are dispensed normally from a swivel-up case [9]. They constitute a mixture of waxes, oils and pigments which contain various proportions of these constituents to give different final characteristics [9]. According to Barker [4], a lipstick typically consists of 65% castor oil, 15% beeswax, 10% carnauba wax, 5% lanolin and a number of soluble and insoluble dyes, pigments and perfumes. Long-lasting lipsticks for example may contain higher amounts of waxes and pigments with less amounts of oil. Lipsticks with higher amounts of oil with less waxes, by contrast, do not last as long but can be easily and smoothly applied [9].

3.2.6.1 Colour in Lipsticks

Colour is a major selling point of any lipstick [10] although other beneficial characteristics for example, ultra violet (UV) and microbial protection, are also incorporated into modern lipsticks. In general, lipstick colours contain some measure of

red and this allows shades ranging between orange-yellow and purple-blue to be prepared [10].

The colour in a lipstick is imparted by both dyes and pigments present in the matrix [10] The most common dyes used in lipsticks are water soluble eosin and other halogenated derivatives of fluorescein (‘bromo acids’) [10]. Figure 3.2 and 3.3 shows the chemical structures of eosin and fluorescein respectively.

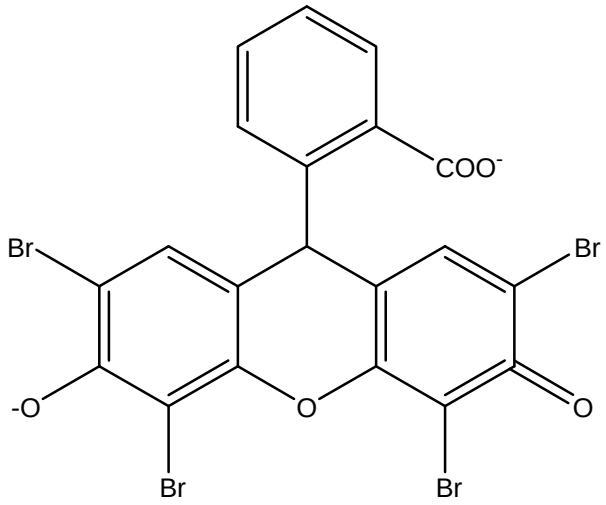


Figure 3.2 Chemical structure of Eosin.

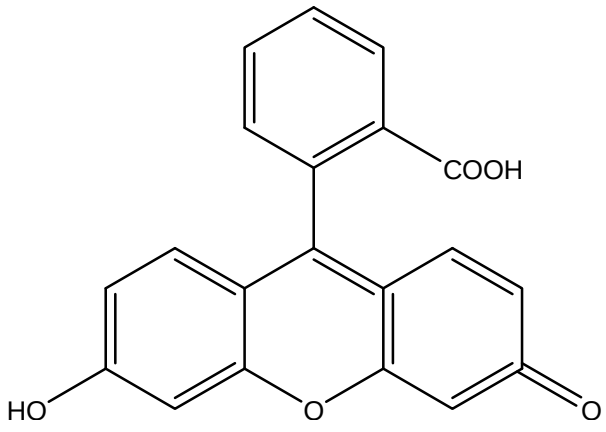


Figure 3.3 Chemical structure of Fluorescein.

The most common pigment used in lipsticks is titanium dioxide (TiO_2) which has been found to be the most effective white pigment for brightening the colour and obtaining a pink shade in lipsticks [10]. Titanium dioxide is also effective as a sunscreen. Pigments containing calcium, barium and strontium may also be used in some lipsticks [10]. Table 3.1 shows typical proportion of colours in lipsticks.

Table 3.1 Typical proportions of colours in lipsticks [14]

COLOUR	PERCENTAGE (%)
Staining dyes	2-3
Oil Soluble Pigment	2
Insoluble Pigment	8-10
Titanium Dioxide	1

Solvents also form some part of the lipstick base and contribute to the physical properties of lipsticks. Fatty materials, for example castor oil, are commonly used as solvents [5, 9, 10]. Other solvents include fatty alcohols and esters [10] which impart a creamy and moisturising feel on the lips [13]. An excess of the solvent can result in a smelly greasy film [9].

3.2.7 Shoe Polish

Shoe care has been practised ever since the introduction of leather sandals by the Romans as a way of softening, protecting and prolonging the life of footwear [11]. At that time, animal fats were used to replenish any loss of natural oils on the leather resulting from exposure to the environment. Today the substances used for shoe care have changed considerably. Modern shoe polishes no longer contain animal fats but are formed from a complex mixture of waxes, silicone and polymers dissolved in appropriate solvents. This combination produces a protective coating which can be easily applied, adheres strongly and can be easily buffed to the desired gloss [11]. The most common shoe polishes are either pastes (solvent based) or creams (emulsion based).

3.2.7.1 Solvent Based Shoe Polish

Solvent based shoe polish, as its name suggests, consists primarily of a solvent, which can be turpentine, mineral spirit or naphtha, on its own or in combination with other components [11]. Other major components are waxes and colours. The roles of the solvents are to solubilise the wax components and to provide a uniform layer of thin film coating on the shoe's surfaces. Upon evaporation of the solvent, the film can be buffed to produce the desired shine. The first stage in the production of solvent based shoe polish is the preparation of the wax blend. In order to produce a paste with different hardness, a high melting wax, for example carnauba wax, is combined with low melting wax, for example paraffin wax, in varying proportions [11]. A typical solvent based shoe polish contains approximately 20 – 40% of the wax component.

The colours used in this type of shoe polish are mostly Nigrosine (in black shoe polish), Rhodamine and Chrysoiodine (in brown shoe polish). The chemical structures for these materials are presented in Figures 3.4, 3.5 and 3.6 respectively.

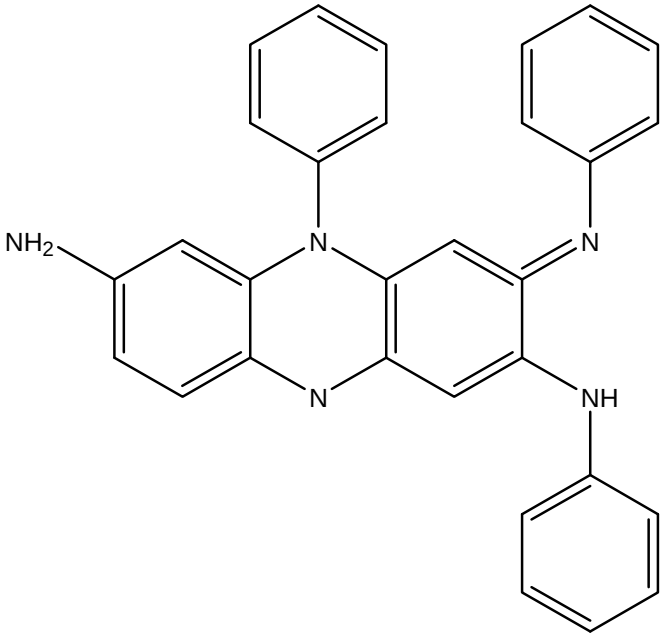


Figure 3.4 Chemical structure of Nigrosine

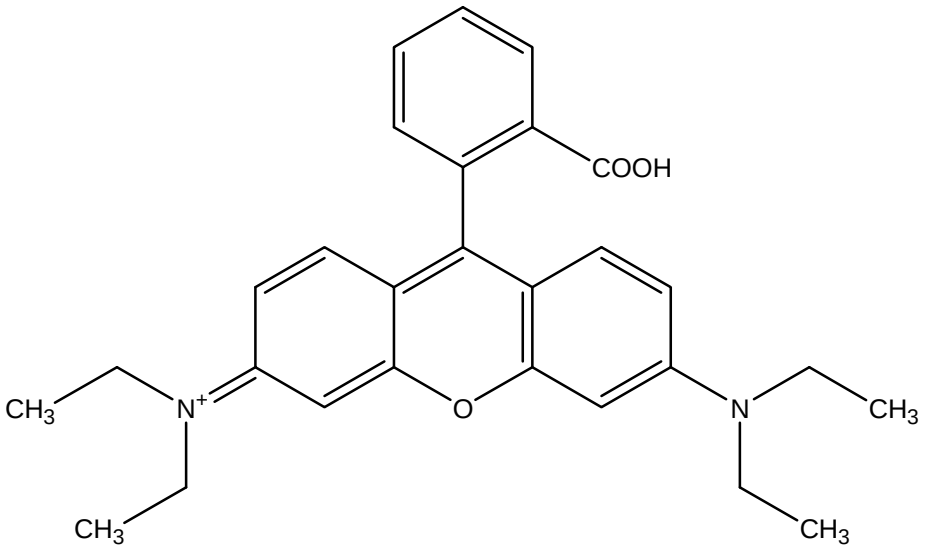


Figure 3.5 Chemical structure of Rhodamine

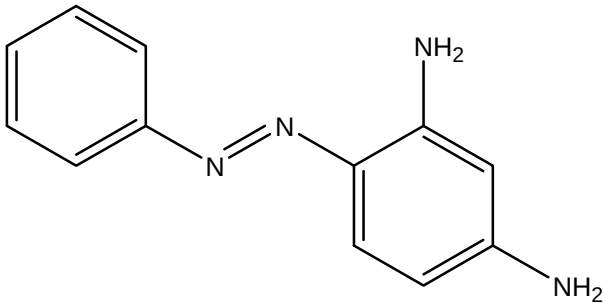


Figure 3.6 Chemical Structure of Chrysoiodine

3.2.7.2 Emulsion Shoe Polish

In general, emulsion shoe polishes are much softer than solvent based shoe polish. They are normally used for fashion shoes and are therefore available in many colours [11]. Water seems to be the primary component and since waxes are immiscible with water; a solvent such as turpentine is commonly added to the formulation. In the preparation of this type of shoe polish, turpentine and waxes are pre-mixed before combining with water to form an emulsion. In order to produce a stable emulsion, an emulsifier such as

potassium or sodium soap is added [11]. For coloured shoe cream, oil soluble dye is used, while for white coloured shoe cream, titanium dioxide is commonly used as the whitening agent.

3.2.8 Lip Balms

Lip balms are also moulded in the form of crayons and are dispensed from a swivel-up case similar to lipstick. It can also be found in small glass or plastic jars or tubs. Lip balms are normally used for the protection of lips against the drying effect of wind, cold or harsh weather. Upon application to the lips, lip balm produces a thin layer which is water resistant that minimises, if not prevents the loss of moisture from the lips [10]. In contrast to lipstick, most lip balm is colourless which means that no dyes are added in their formulation, however sometimes colour is added to meet user preferences. This is achieved through the incorporation of small amounts of oil-soluble dye [10].

Most of the lip balms available in the market today contain sunscreen [12]. This is a compound which helps to absorb or reflect the ultra-violet A (UVA - 315 – 400nm) and ultra-violet B (UVB - 290 – 315nm) light from the sun. Examples of sunscreens are octyl salicylate (Figure 3.7) and titanium dioxide. The former absorbs nearly the full range of the UVB and the latter rather blocks the UVB light from reaching the skin.

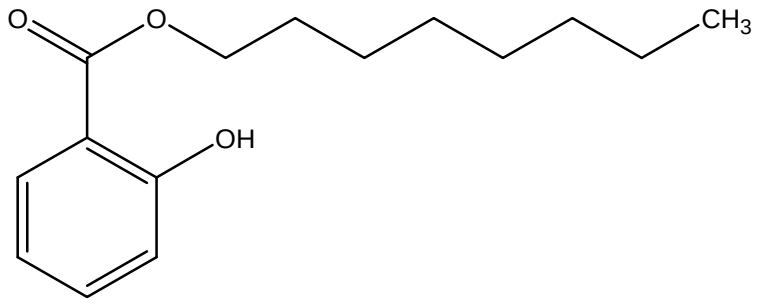


Figure 3.7 Chemical structure of Octyl Salicylate

3.3 Theoretical Backgrounds of the Analytical Techniques Employed

The analytical techniques employed reside within two areas of analytical chemistry namely chromatographic and spectroscopic analysis. The analytical techniques used included Thin Layer Chromatography (TLC), Microspectrophotometry (MSP), Ultra-Violet and Visible (UV/Vis) Spectroscopy and Gas Chromatography with Flame Ionisation Detector (GC-FID). A comprehensive introduction to each of the analytical technique is presented in Appendix A.

3.4 EXPERIMENTAL

3.4.1 Sample Collection and Preparation

In total sixty three (63) samples consisting of 21 lip balms, 17 shoe polishes and 25 lipsticks were analysed. These were obtained from various retail outlets in Glasgow, Scotland. All of the samples investigated as part of this study were given an individual code and are listed in Table 3.2a to 3.2c.

Table 3.2a List of lip balm (LB) samples analysed in the study.

Sample Name	Type	Colour	Code
Body Shop Born Lippy Strawberry	LB	Red	LB1
Body Shop Born Lippy Watermelon	LB	Light Red/Pink	LB2
Body Shop Born Lippy Passionberry	LB	Purple	LB3
Body Shop Born Lippy Mango Peach	LB	Yellow	LB4
Body Shop Born Lippy Raspberry	LB	Pink	LB5
Chap Stick Flava Craze Fruit Craze	LB	Light Red/Pink	LB6
Chap Stick Flava Craze Blue Crazeberry	LB	Light Blue	LB7
Chap Stick Flava Craze Grape Craze	LB	Purple	LB8
Superdrug Strawberry Lip Balm	LB	Red	LB9
Superdrug Vanilla Lip Balm	LB	Colourless	LB10
Lipmate Original Lip Balm	LB	Colourless	LB11
Lypsyl Original Lip Balm	LB	Colourless	LB12
ASDA Original Lip Balm	LB	Colourless	LB13
TESCO Medicated Lip Balm	LB	Colourless	LB14
NIVEA Lip Care Essential Lip Balm	LB	Colourless	LB15
Metholatum Soft Lips French Vanilla	LB	Colourless	LB16
Lypsyl Strawberry	LB	Light Pink	LB17
Vaseline Original	LB	Colourless	LB18
Vaseline Rosy Lips	LB	Red	LB19
Vaseline Aloe Vera	LB	Colourless	LB20
Vaseline Sun Protection	LB	Colourless	LB21

Table 3.2b List of shoe polish samples (SP) analysed in the study.

Sample Name	Type	Colour	Code
Kiwi Shoe Polish Dark Tan	SP	Tan	SP1
Kiwi Shoe Polish Black	SP	Black	SP2
Kiwi Shoe Polish Blue	SP	Blue	SP3
Kiwi Shoe Polish Red	SP	Red	SP4
Granger's Shoe Polish Brown	SP	Brown	SP5
Granger's Shoe Polish Black	SP	Black	SP6
Cherry Blossom Shoe Polish Tan	SP	Tan	SP7
Safeway Shoe Polish Dark Tan	SP	Tan	SP8
ASDA Shoe Polish Brown	SP	Brown	SP9
ASDA Shoe Polish Neutral	SP	Colourless	SP10
Kiwi Shoe Polish Neutral	SP	Colourless	SP11
Granger's Shoe Polish Neutral	SP	Colourless	SP12
Punch Taupe Shoe Cream	SP	Light Black	SP13
Meltonian Shoe Cream Black	SP	Black	SP14
Cherry Blossom Iris Blue	SP	Blue	SP15
Meltonian Shoe Cream Navy Blue	SP	Blue	SP16
Tuxan Shoe Care	SP	Brown	SP17

Table 3.2c List of lipstick (L) samples analysed in the study

Sample Name	Type	Colour	Code
Star Gazer Lipstick 133	L	Purple	L1
Star Gazer Lipstick 105	L	Blue	L2
Star Gazer Lipstick 110	L	Black	L3
Clinique Super Spice	L	Brown	L4
Jane Seymour Pineapple Pink	L	Orange	L5
Body Collection Crushed Rose	L	Brown	L6
RIMMEL London Rich Raisin	L	Brown	L7
Clinique Raspberry Glace	L	Brown	L8
Collection 2000 Advance Colour Cream	L	Brown	L9
Boots Mulberry	L	Light Brown	L10
Estee Lauder French Fig	L	Brown	L11
Amway Debut Peach	L	Orange	L12
Clinique Ginger Flower	L	Red	L13
Estee Lauder All Day Carol Melon	L	Red	L14
Dare Dolly	L	Orange	L15
Apricot Crush	L	Orange	L16
Nutmeg	L	Brown	L17
Red Carnation	L	Red	L18
Marie France Touch of Spice	L	Brown	L19
Revlon Moondrops	L	Orange	L20
TESCO Shade 12	L	Red	L21
TESCO Shade 3	L	Brown	L22
TESCO Shade 13	L	Dark Brown	L23
Max Factor Sunset Rose	L	Red	L24
Body Collection Mango Mood	L	Orange	L25

In each case, smears of the sample were prepared by rubbing the sample onto a clean white cotton cloth swatch (1 cm x 1 cm). The colour of each sample smear was first observed by eye. This was followed by extraction of the sample smear using an organic solvent. Analysis was performed by TLC initially and each developed spot was then analysed by MSP using reflectance mode. The sample extracts were also analysed by UV/Vis and GC-FID.

The robustness of the extraction and analysis in terms of repeatability and reproducibility were established and verified in each case. Comprehensive details of the extraction and analysis (including their validation) are provided in Appendix B.

3.4.2 Dataset Collection

Once the extraction and analytical procedures were optimised all of the remaining samples were analysed accordingly using the optimised procedure. Four different datasets, one from each of the analytical technique, TLC, MSP, UV/Vis and GC-FID were generated by the end of this process. Normalised and power transformation (square and fourth root) datasets were also generated. These datasets were used individually and in combination to identify groupings or relationships amongst the wax based products under study.

In total 54 datasets were analysed using the pattern recognition techniques. In each case the data was analysed using Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA) and Self Organising Feature Maps (SOFM) neural network to assess the level of discrimination between the product types, if any, provided by the unsupervised pattern recognition techniques.

The full list of the datasets used is presented in Appendix C.

3.4.3 Effect of Weathering

The effect of weathering of the sample smears was studied by ageing selected samples according to the conditions employed by Cole [13] by naturally exposing the smears at room temperature under normal lighting condition and by ageing artificially at 56°C in an oven for the period of 14 and 28 days. After each period, the smears were subjected to the same extraction and analytical and regime as previous samples.

3.4.5 Software Packages

All PCA and HCA were performed using MATLAB® R2008a (MathWorks Inc.) Statistics Toolbox while the SOFM was performed using a Viscovery SOMine® 4.0.2 (Eudaptics) software.

3.5 RESULTS AND DISCUSSION

Of the 54 datasets analysed, only the TLC, MSP, and TLC/MSP combined (for 33 coloured samples) and UV/Vis and GC square root pre-processed data combined (for all 63 samples) gave meaningful discrimination of the samples. As such only these datasets are discussed in detail in the following sections.

3.5.1 Limitations of the TLC and MSP Dataset

Of the 63 samples available in this study, 33 samples (22 lipsticks and 11 shoe polishes) as presented in Table 3.3 produced coloured spots when analysed by TLC. The rest of the samples, which included the lip balms, some lipsticks and shoe polishes, did not produce any coloured spots which meant that they could also not be analysed by these methods.

Table 3.3 Samples that produced coloured spot when analysed by the TLC

NO.	SAMPLE NAME	CODE	TYPE
1	Clinique Super Spice	L4	Lipstick
2	Jane Seymour Pineapple Pink	L5	Lipstick
3	Body Collection Crushed Rose	L6	Lipstick
4	RIMMEL London Rich Raisin	L7	Lipstick
5	Clinique Raspberry Glace	L8	Lipstick
6	Collection 2000 Advance Colour Cream	L9	Lipstick
7	Boots Mulberry	L10	Lipstick
8	Estee Lauder French Fig	L11	Lipstick
9	Amway Debut Peach	L12	Lipstick
10	Clinique Ginger Flower	L13	Lipstick
11	Estee Lauder All Day Carol Melon	L14	Lipstick
12	Dare Dolly	L15	Lipstick
13	Apricot Crush	L16	Lipstick
14	Nutmeg	L17	Lipstick
15	Red Carnation	L18	Lipstick
16	Marie France Touch of Spice	L19	Lipstick
17	Revlon Moondrops	L20	Lipstick
18	TESCO Shade 12	L21	Lipstick
19	TESCO Shade 3	L22	Lipstick
20	TESCO Shade 13	L23	Lipstick
21	Max Factor Sunset Rose	L24	Lipstick
22	Body Collection Mango Mood	L25	Lipstick
23	Kiwi Shoe Polish Dark Tan	SP1	Shoe Polish
24	Kiwi Shoe Polish Black	SP2	Shoe Polish
25	Kiwi Shoe Polish Blue	SP3	Shoe Polish
26	Kiwi Shoe Polish Red	SP4	Shoe Polish
27	Granger's Shoe Polish Brown	SP5	Shoe Polish
28	Granger's Shoe Polish Black	SP6	Shoe Polish
29	Safeway Shoe Polish Dark Tan	SP8	Shoe Polish
30	ASDA Shoe Polish Brown	SP9	Shoe Polish
31	Punch Taupe Shoe Cream	SP13	Shoe Polish
32	Meltonian Shoe Cream Black	SP14	Shoe Polish
33	Meltonian Shoe Cream Navy Blue	SP16	Shoe Polish

3.5.2 Principal Component Analysis (PCA)

3.5.2.1 PCA for the TLC Dataset

The TLC dataset consisted of the retardation factor of the spot recorded after the first and second development respectively. As a consequence only two principal components could be derived from the dataset. After the principal component transformations, it was found that the first principal component (PC1) accounted for 98% of the variance in the dataset while the second principal component (PC2) accounted for only 2% of the variance. Figure 3.8 illustrates the score plot for the principal components generated

from the TLC dataset. Arbitrary boundaries have been constructed to illustrate the different groupings achieved in the dataset.

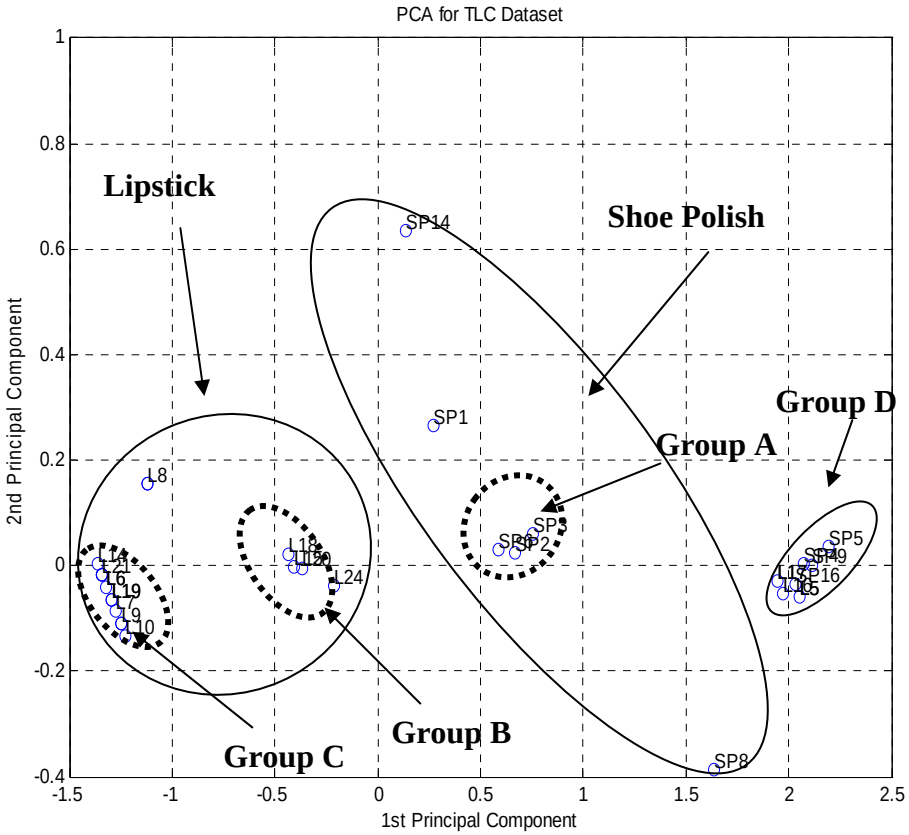


Figure 3.8 Score plot for the principal components for the TLC dataset

It is evident that full discrimination of the samples based on its type was not possible using the TLC dataset alone. Although most of the lipstick and shoe polishes can be grouped into two separate clusters, a cluster located at the bottom far right of the score plot (Group D) is convoluted and contains both lipstick and shoe polish samples.

Smaller clusters of samples were also evident within each of the larger groups. One separate group (Group A) could be further established within the shoe polish cluster. Two groups (Group B and C) could be discriminated within the lipstick cluster. Group

A consisted of samples SP2, SP3 and SP6. Group B consisted of L15, L18, L20 and L24 whilst Group C consisted of L6, L9, L7, L10, L14, L19 and L21 samples. In order to gain further clarity of the samples within each of the smaller congested groups, A and B were expanded and are presented in Figure 3.9.

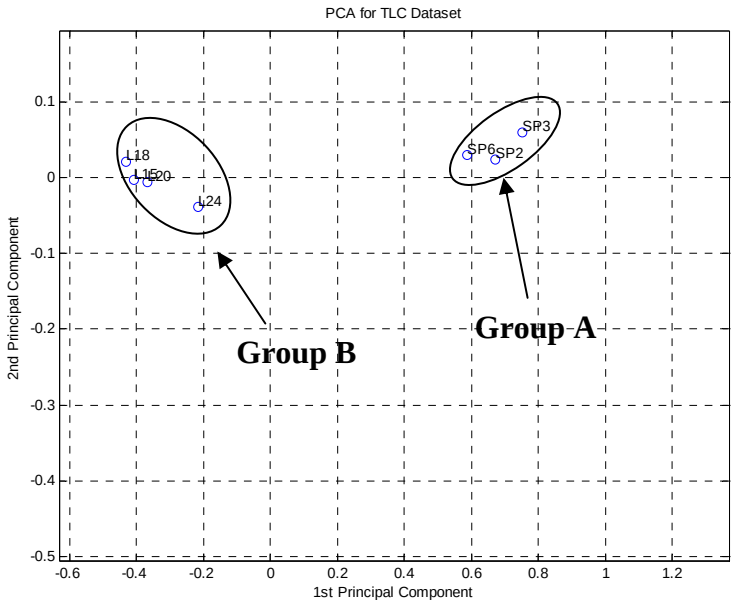


Figure 3.9 Enlarged view of the Group A and Group B cluster

Each of the samples within these various groups have spots with similar retardation factor values as presented in Table 3.4 to 3.6.

Table 3.4 TLC outcomes of the samples within the Group A cluster

SAMPLE	R_F1	R_F2	Spot Colour
SP2	0.52	0.55	Black
SP3	0.55	0.56	Blue
SP6	0.55	0.53	Black

Table 3.5 TLC outcomes of the samples within the Group B cluster

SAMPLE	R_F1	R_F2	Spot Colour
L15	0.24	0.32	Orange
L18	0.24	0.31	Orange
L20	0.25	0.33	Orange
L24	0.28	0.37	Orange

Table 3.6 TLC outcomes of the samples within the Group C cluster

SAMPLE	R_F1	R_F2	Spot Colour
L6	0.00	0.13	Red
L9	0.00	0.16	Pink
L7	0.00	0.15	Red
L10	0.00	0.17	Pink
L14	0.00	0.11	Red
L19	0.00	0.14	Red
L21	0.00	0.12	Pink

The congested cluster (Group D) located at the bottom far right of the score plot was also expanded and presented in Figure 3.10. This reveals that Group D consisted of samples SP4, SP5, SP9, SP16, L5, L16 and L17. Examination of these samples revealed that their TLC spots have quite similar R_f values (in the range of 0.83 to 0.91) although their spot colours were different as presented in Table 3.7.

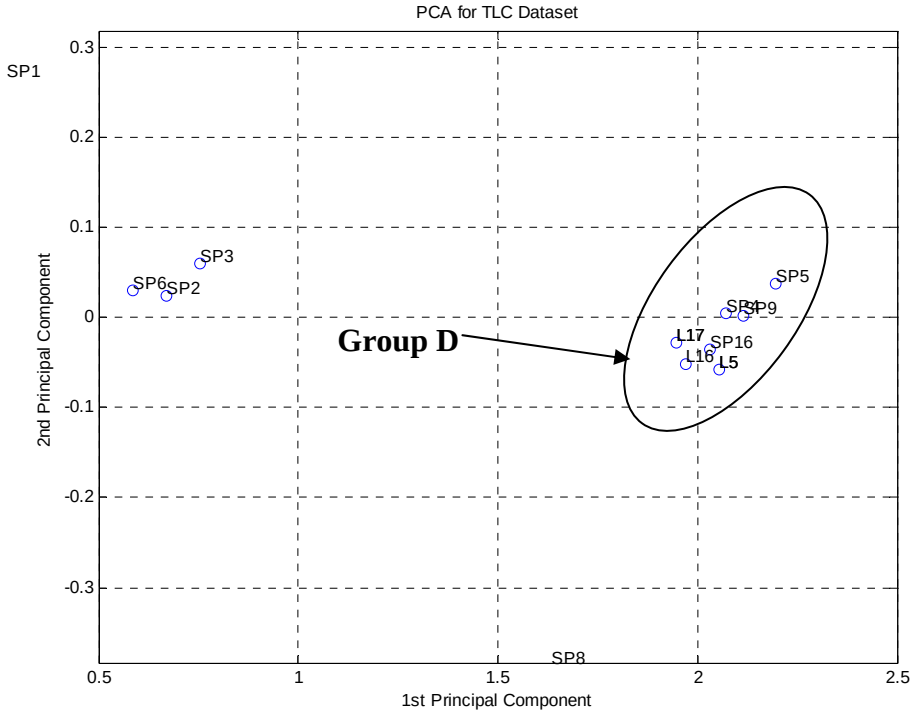


Figure 3.10 Enlarged view of the Group D cluster

Table 3.7 TLC outcomes of the samples within the Group D cluster

SAMPLE	R_F1	R_F2	Spot Colour
SP4	0.87	0.86	Red
SP5	0.91	0.88	Black
SP9	0.88	0.87	Orange
SP16	0.85	0.86	Pink
L5	0.85	0.87	Orange
L16	0.83	0.85	Orange
L17	0.83	0.84	Orange

3.5.2.2 PCA for the MSP Dataset

The MSP dataset produced a UV spectrum for each spot present on the TLC plates. The spectral data was entered into PCA and generated two principal components. The score plot constructed from the two principal components derived from the dataset is illustrated in Figure 3.11. The first principal component (PC1) accounted for 99% of the variance in the dataset and the second principal component (PC2) accounted for less than 1% of the variance in the dataset.

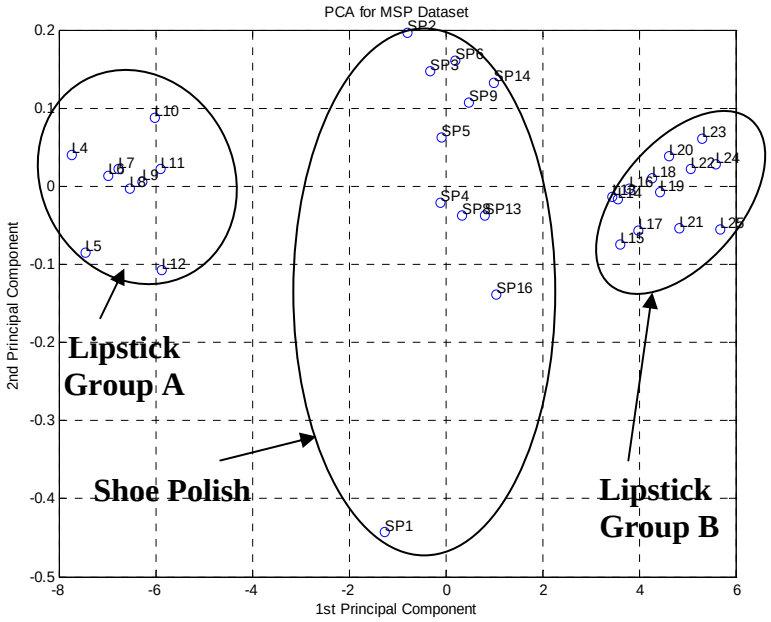


Figure 3.11 Score plot for the first two principal components for the MSP dataset

Three clusters of samples were evident as illustrated. All of the shoe polish samples could be grouped into a large spread out group in the centre of the score plot whilst two clusters were presented for the lipstick samples i.e. Group A and B. Lipstick Group A consisted of samples L4 to L10 inclusive while Lipstick Group B consisted of samples L13 to L25 inclusive. Examinations of the reflectance spectra of the samples in each group revealed that the samples in the Lipstick Group A cluster have reflectance spectra with percentage reflectance lower than the samples in the Group B as illustrated in Figure 3.12 and 3.13 which could perhaps explain the groupings.

Further examinations of the reflectance spectra of the shoe polish cluster located at the centre of the score plot is illustrated in Figure 3.14 reveal that their percentage reflectance is in between that of Group A and Group B samples, further suggesting that this variable rather than wavelength is the grouping variable.

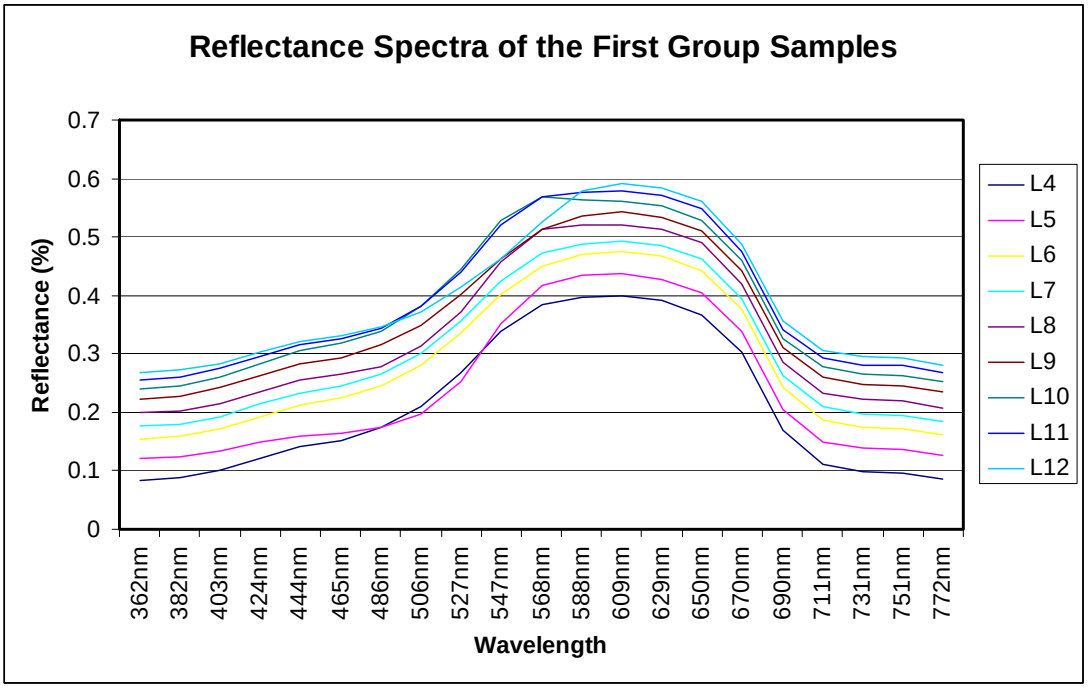


Figure 3.12 Reflectance spectra of the Lipstick Group A samples

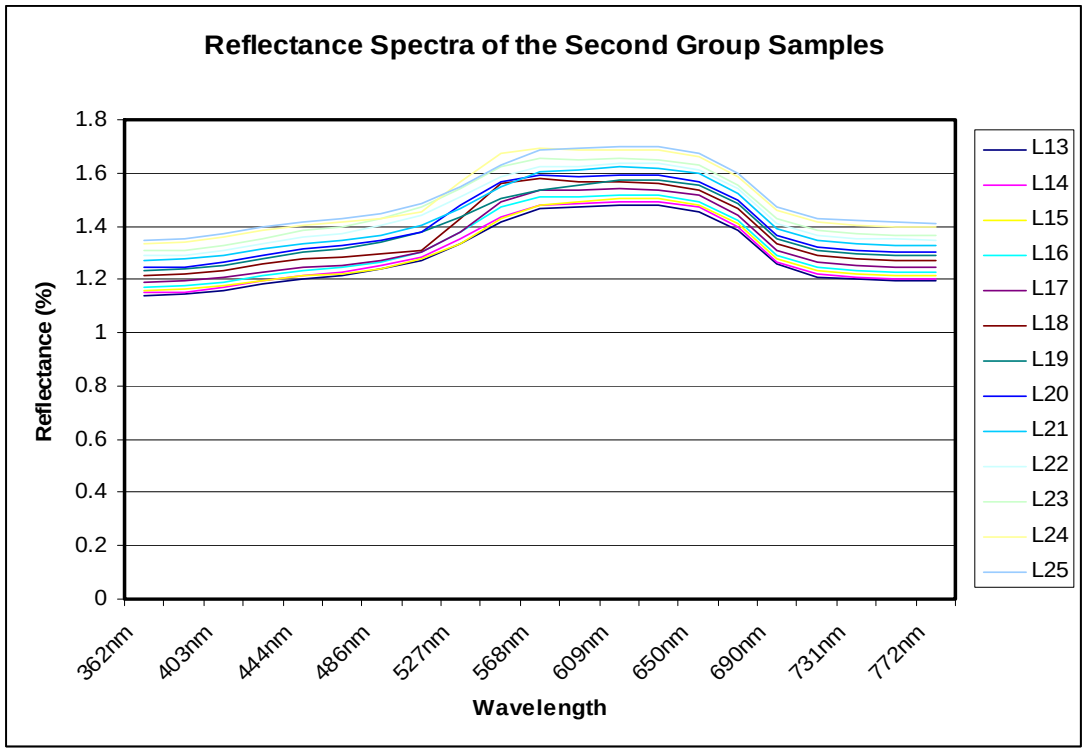


Figure 3.13 Reflectance spectra of the Lipstick Group B samples

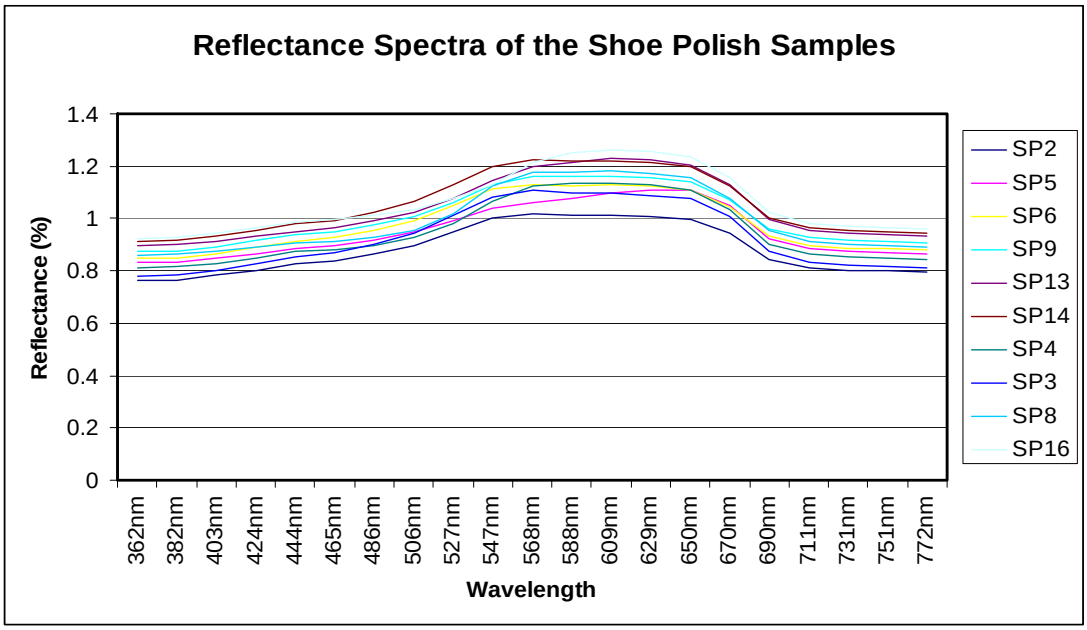


Figure 3.14 Reflectance spectra of the shoe polish samples

3.5.2.3 PCA for the TLC and MSP Combined Dataset

The scree plot generated from the principal component transformation of the TLC and MSP combined dataset is presented in Figure 3.15

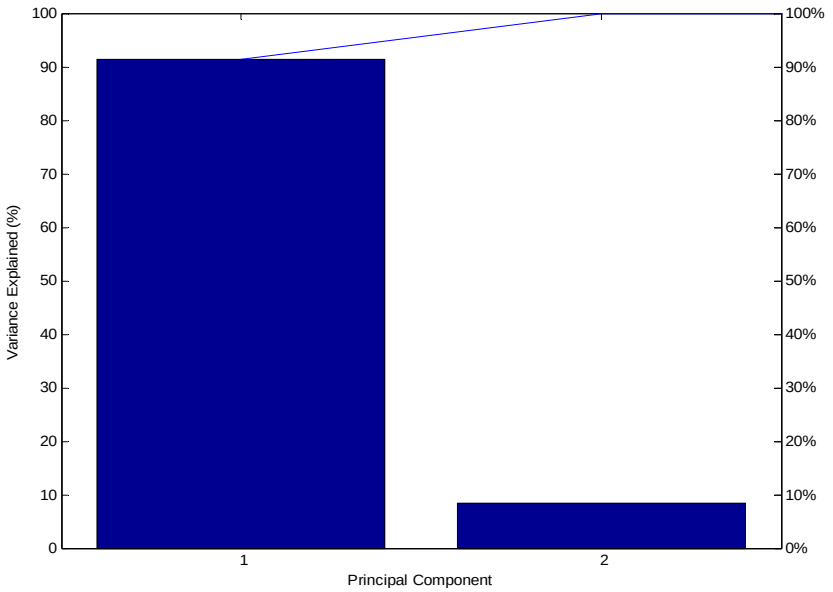


Figure 3.15 Scree plot for the first two principal components of the TLC and MSP combined dataset

The first principal component (PC1) accounts for approximately 90% of the variance in the dataset while the second principal component (PC2) accounts for approximately 8% of the variance in the dataset. The combination of these principal components therefore accounts for approximately 98% of the total variance in the dataset.

The score plot constructed using the first two principal components derived from the TLC and MSP combined dataset is presented in Figure 3.16. Three groupings similar to that of the MSP dataset are evident in the score plot. Two lipstick clusters were evident with some smaller clusters within these groups. A better defined cluster was achieved for the shoe polish samples than for TLC or MSP alone and again smaller clusters within that grouping were evident.

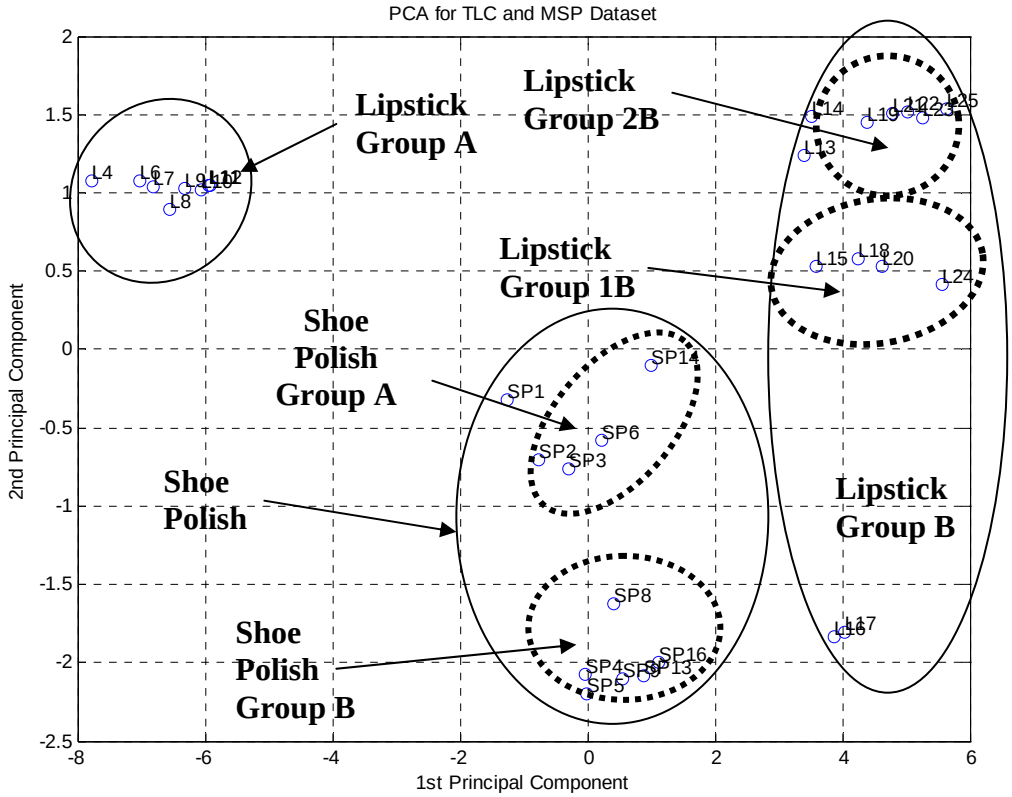


Figure 3.16 Score plot for the first two principal components for the TLC and MSP combined dataset

Enlargement of Lipstick Group A cluster illustrated in Figure 3.17 reveals that it consists of samples L4, L6, L11, L8, L9, L10, L11 and L12. Further examinations of the samples within this cluster reveal that they presented similar retardation factors and reflectance spectra as presented in Table 3.8 and Figure 3.18.

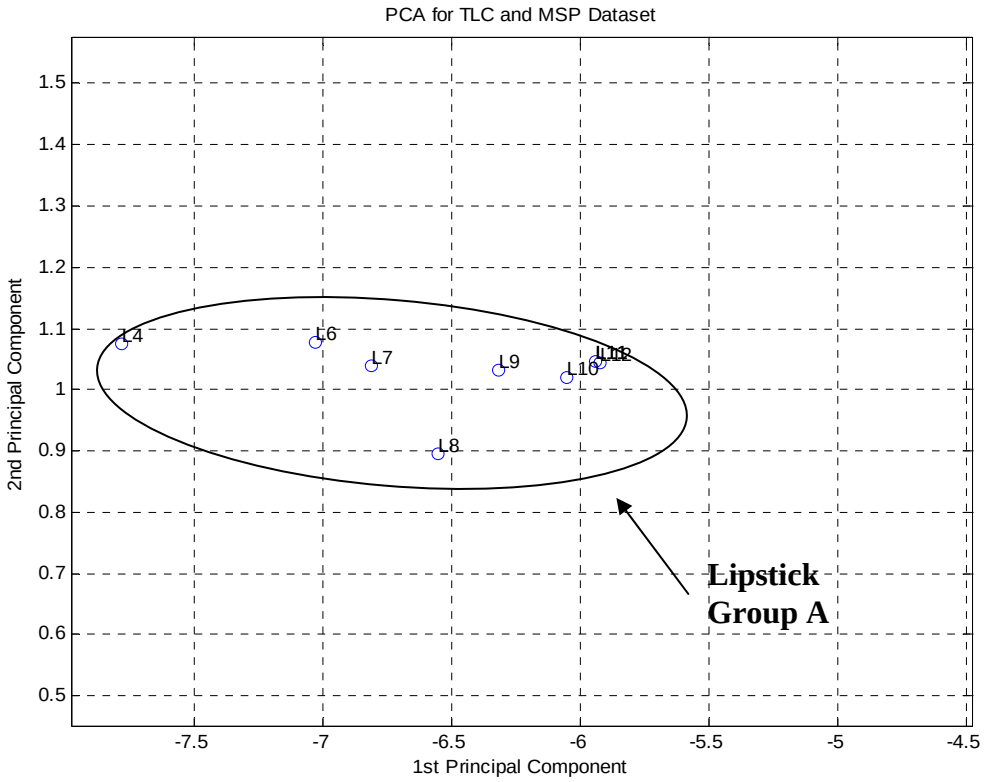


Figure 3.17 Enlarged view of the Lipstick Group A cluster

Table 3.8 TLC outcomes for samples within the Lipstick Group A cluster

SAMPLE	R_F1	R_F2	Spot Colour
L4	0.00	0.12	Red
L6	0.00	0.13	Red
L11	0.00	0.16	Red
L8	0.00	0.13	Red
L9	0.00	0.16	Pink
L10	0.00	0.17	Pink
L11	0.00	0.16	Red
L12	0.00	0.16	Red

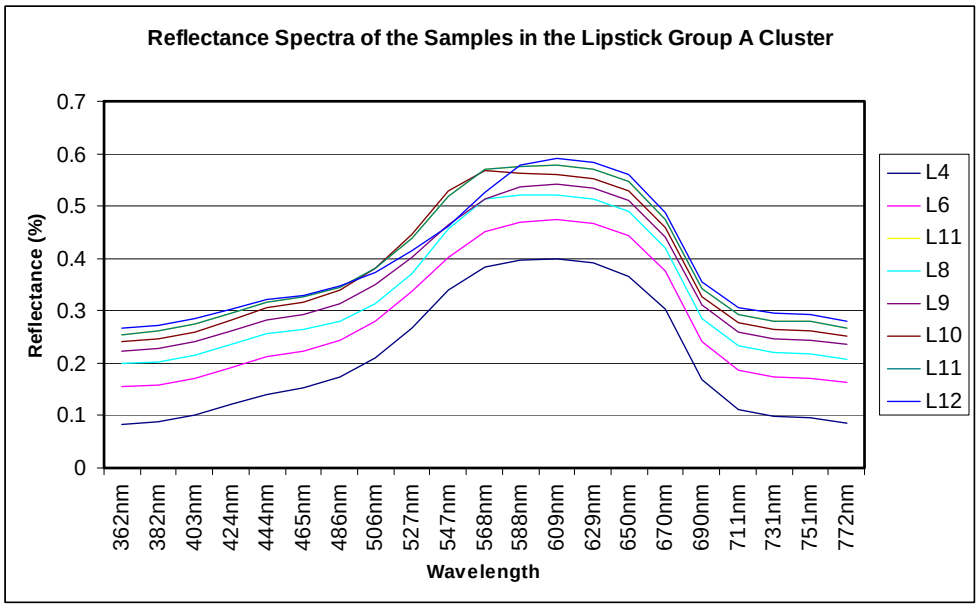


Figure 3.18 Reflectance spectra of the samples in the Lipstick Group A cluster

Two clusters were further established within the Lipstick Group B cluster. Enlargement of these revealed that Group 2B cluster consisted of samples L13, L14, L19, L21, L22, L23 and L25 and Lipstick Group 1B consisted of samples L15, L18, L20 and L24. These groupings are illustrated in Figure 3.19.

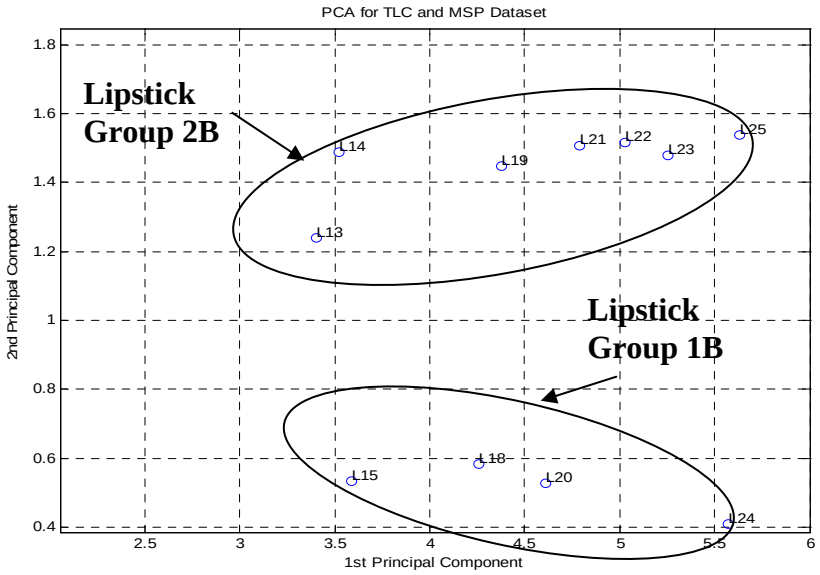


Figure 3.19 Enlarged view of the Lipstick Group 1B and 2B cluster

Further examinations of the Lipstick Group 1B and 2B cluster reveal that the samples recorded similar retardation factors and reflectance spectra respectively and are presented in Tables 3.9 and 3.10 and Figures 3.20 and 3.21 respectively.

Table 3.9 TLC outcomes of the samples within the Lipstick Group 1B cluster

SAMPLE	R_F1	R_F2	Spot Colour
L15	0.24	0.32	Orange
L18	0.24	0.31	Orange
L20	0.25	0.33	Orange
L24	0.28	0.37	Orange

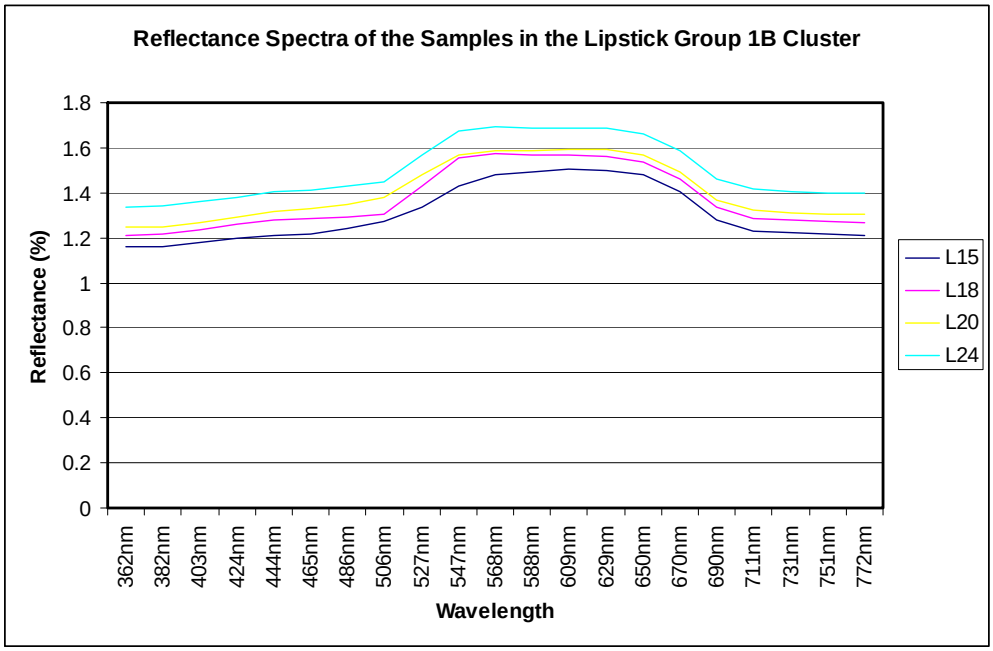


Figure 3.20 Reflectance spectra of the samples in the Lipstick Group 1B cluster

Table 3.10 TLC outcomes of the samples within the Lipstick Group 2B cluster

SAMPLE	R_F1	R_F2	Spot Colour
L13	0.10	0.13	Red
L14	0.00	0.11	Red
L19	0.00	0.14	Red
L21	0.00	0.14	Pink
L22	0.00	0.12	Pink
L23	0.00	0.14	Pink
L25	0.00	0.12	Red

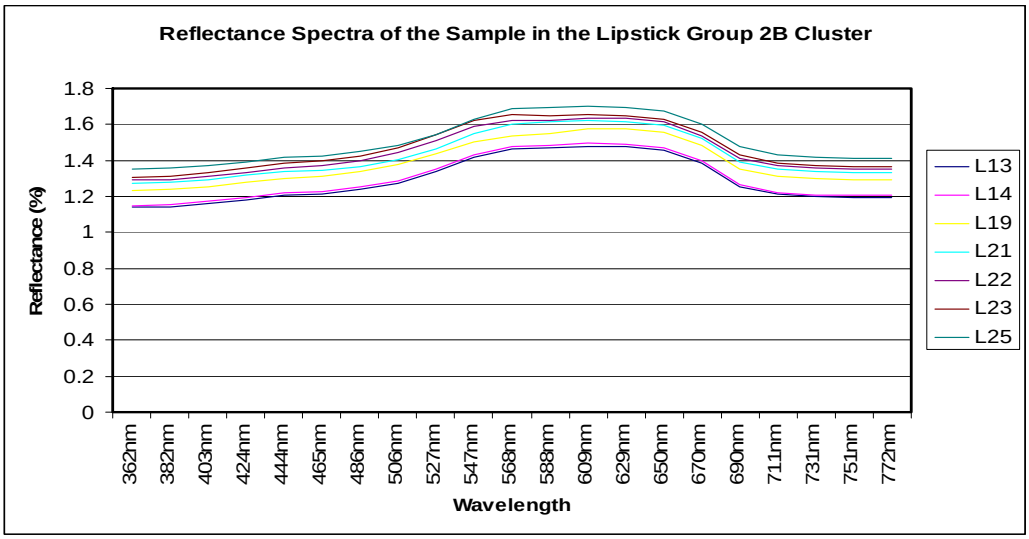


Figure 3.21 Reflectance spectra of the samples in the Lipstick Group 2B cluster

Within the shoe polish cluster, two groups can be further established. Group A consists of sample SP2, SP3, SP6 and SP14 and Group B consists of samples SP8, SP4, SP5, SP9, SP16 and SP13 samples. TLC examination of the samples in the Group A revealed R_f values lower than 0.60 while the samples in Group B have R_f values higher than 0.66. The relevant data is presented in Tables 3.11 and 3.12. The reflectance spectra of both shoe polish groups provided no discriminating information and are illustrated in Figures 3.22 and 3.23.

Table 3.11 TLC outcomes for the samples in the Shoe Polish Group A cluster

SAMPLE	R_F1	R_F2	Spot Colour
SP2	0.52	0.55	Black
SP3	0.55	0.56	Blue
SP6	0.50	0.53	Black
SP14	0.54	0.53	Black

Table 3.12 TLC outcomes for the samples in the Shoe Polish Group B cluster

SAMPLE	R_F1	R_F2	Spot Colour
SP4	0.87	0.86	Orange
SP5	0.91	0.88	Red
SP8	0.66	0.85	Black
SP9	0.88	0.87	Orange
SP13	0.88	0.87	Pink
SP16	0.85	0.86	Orange

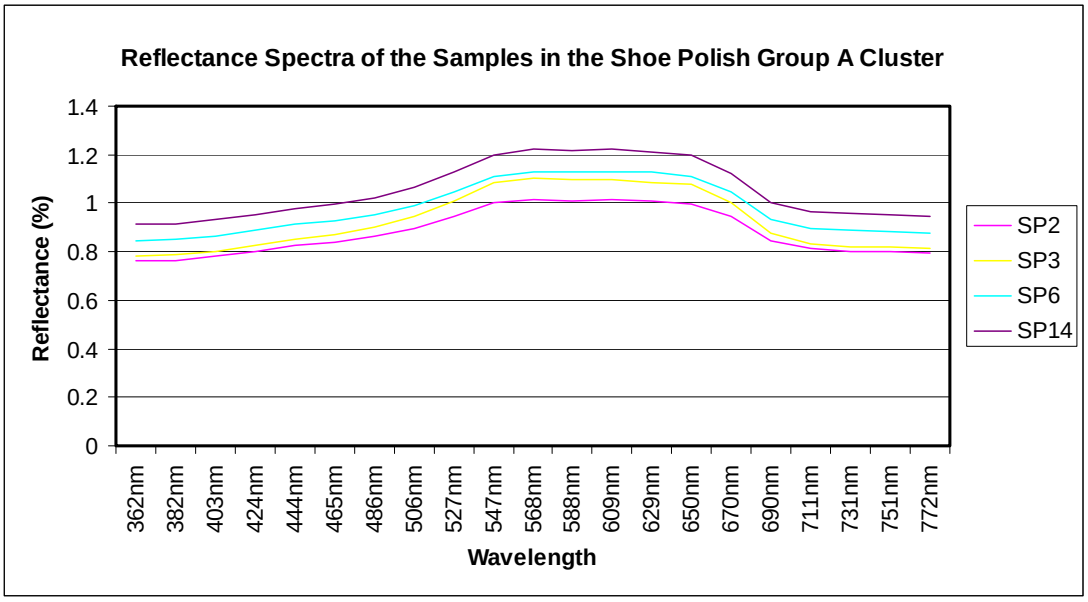


Figure 3.22 Reflectance spectra of the samples in the Shoe Polish Group A cluster

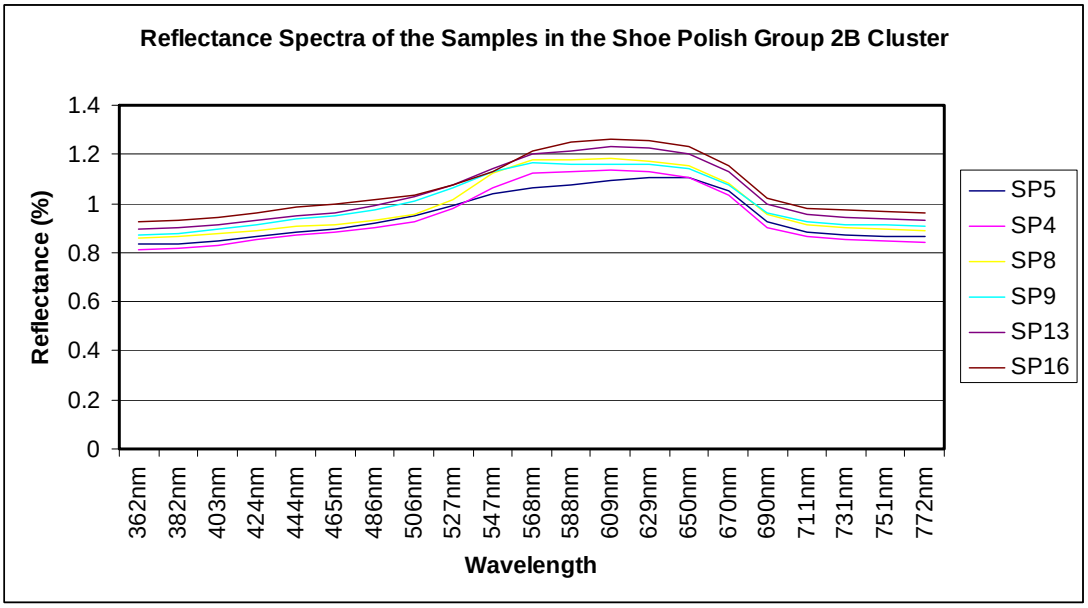


Figure 3.23 Reflectance spectra of the samples in the Shoe Polish Group B cluster

3.5.2.4 PCA for the UV/Vis and GC Square Root Combined Dataset

The Scree plot derived from the UV/Vis and GC square root combined dataset is presented in Figure 3.24. The first two principal components account for 56% of the variance in the combined dataset. The third and the fourth principal components account for approximately 7% and 6% variance respectively and cumulatively, these four principal components account for 70% of the total variance in the dataset.

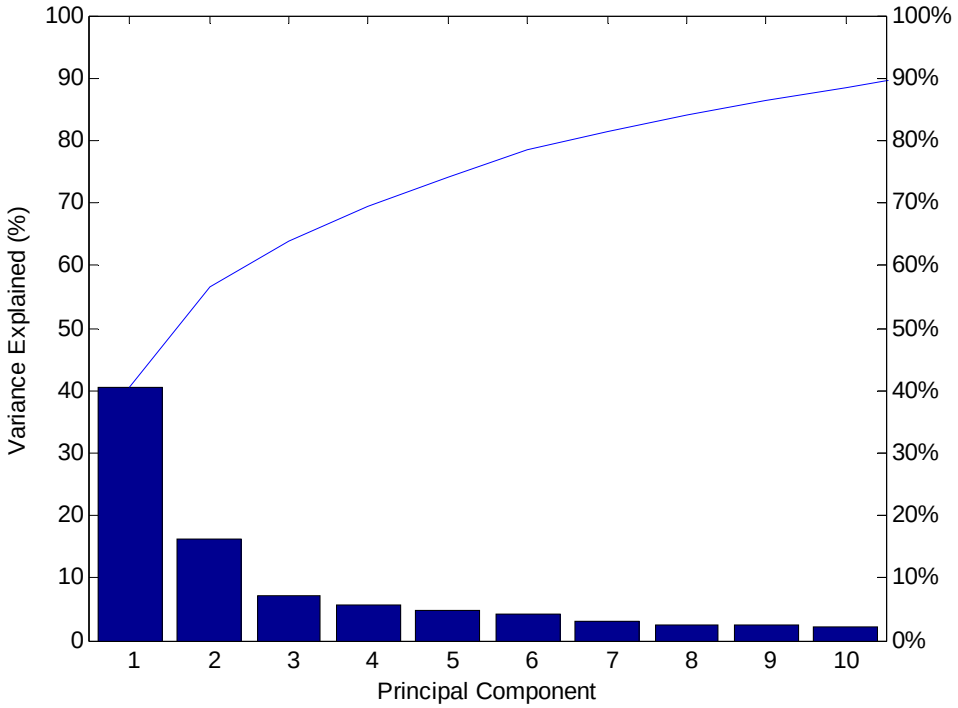


Figure 3.24 Scree plot for the first ten principal components for UV/Vis and GC square root combined dataset

Figure 3.25 shows the score plot for the first two principal components for this data set. The plot is convoluted and no clear groupings of the samples are evident therefore no meaningful discrimination of the samples can be made.

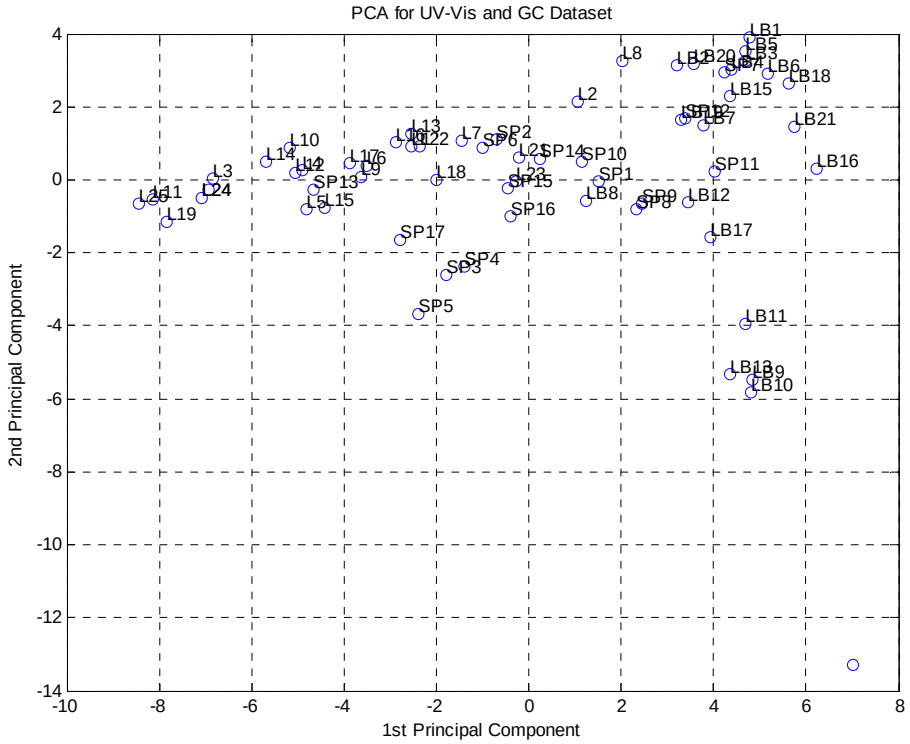


Figure 3.25 Score plot for the first two principal components for the UV/Vis and GC square root combined dataset

3.5.3 Hierarchical Cluster Analysis (HCA)

HCA for the selected datasets were performed using Euclidean distance as the proximity measure. Euclidean distance was chosen as it is the simplest and most commonly used proximity measure [42]. To establish the effect of the linkage mechanism chosen; single, average and complete linkage strategies were applied to the datasets.

The dendrogram produced from each dataset with each linkage strategy was examined to determine which strategy best described the dataset. The accuracy of the dendrogram in describing the dataset was evaluated against the known sample provenance.

3.5.3.1 HCA for the TLC Dataset

None of the linkage strategies applied in the HCA using the TLC dataset on its own was capable of successfully discriminating the samples under study. Of the three linkage strategies, the complete linkage strategy was considered to be the best and resulted in two clusters (Cluster 1 and 2) as illustrated in Figure 3.26. Cluster 1 consists mainly of lipstick samples while Cluster 2 consists of a mixture of lipstick and shoe polish samples. Table 3.13 describes in details the memberships of each cluster.

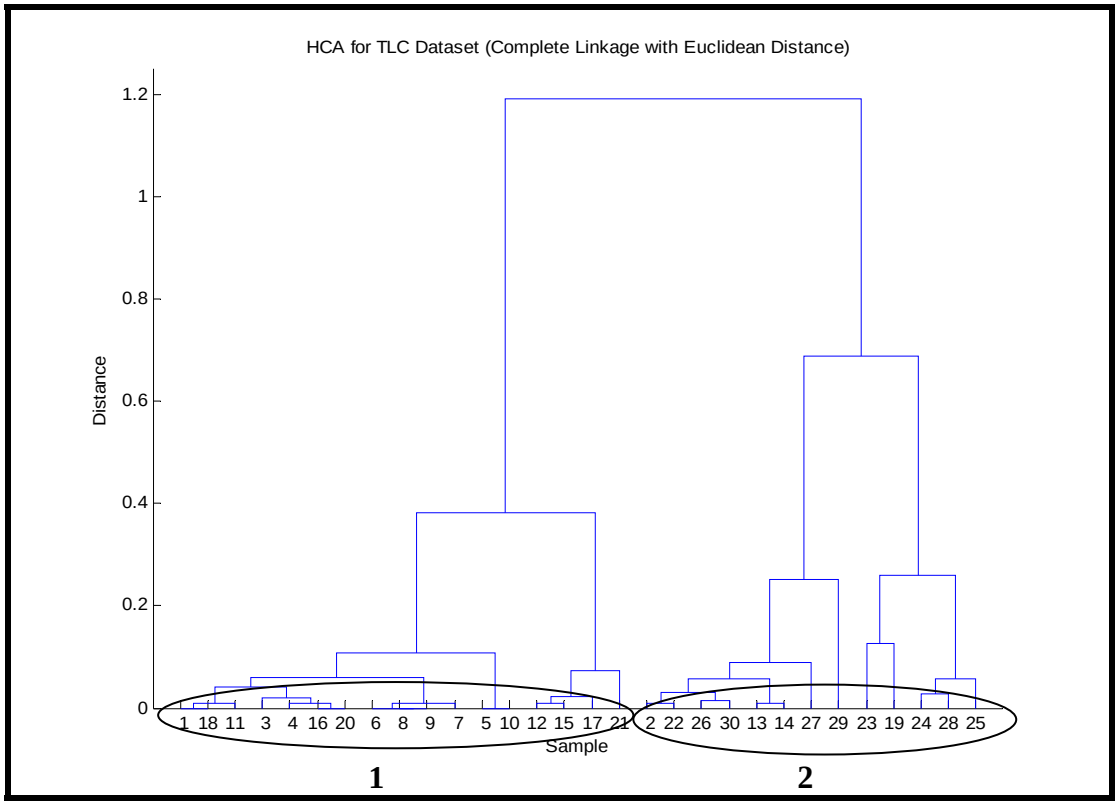


Figure 3.26 Dendrogram for the TLC dataset constructed using Euclidean Distance and Complete Linkage

Table 3.13 Identified cluster and members in each cluster for the TLC dataset.

CLUSTER	NODE NO.	MEMBER(S)
1	1	L4, L22, L25
	18	L21
	11	L14
	3	L6
	4	L7
	16	L19
	20	L7
	6	L9
	8	L11
	9	L12
	7	L10
	5	L8
	10	L13
	12	L15
	15	L18
	17	L20
	21	L24
2	2	L5
	22	SP16
	26	SP4
	30	SP9, SP13
	13	L16
	14	L17
	27	SP5
	29	SP8
	23	SP1
	19	SP14
	24	SP2
	28	SP6
	25	SP3

3.5.3.2 HCA for the MSP Dataset

No full discrimination of the samples was achieved when the MSP dataset was subjected to HCA. The complete linkage was found to best describe the discrimination of the samples in the MSP dataset. In this case three clusters (Cluster 1, 2 and 3) resulted from the MSP dataset as illustrated in Figure 3.27. Cluster 1 and 3 consists of lipstick samples while Cluster 2 consists of the shoe polish samples. Table 3.14 describes in detail the memberships of each cluster.

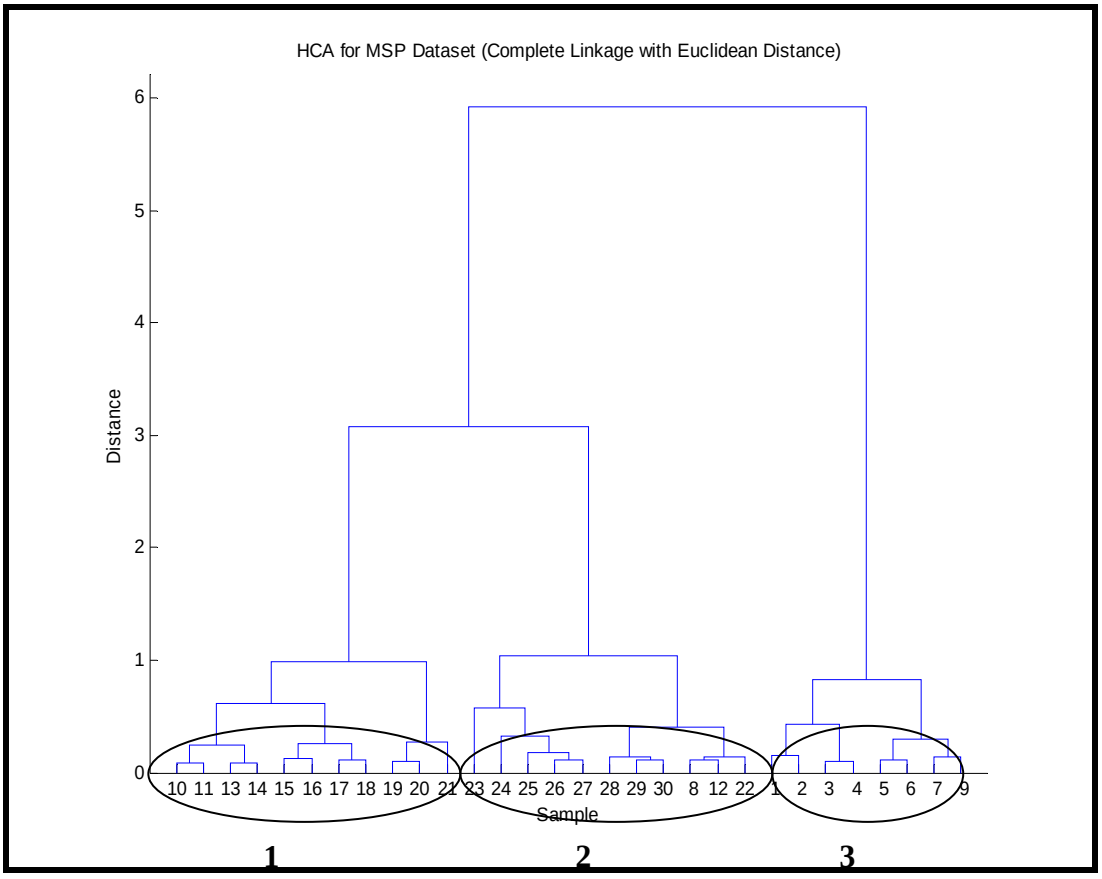


Figure 3.27 Dendrogram for the MSP dataset constructed using Euclidean Distance and Complete Linkage

Table 3.14 Identified cluster and members in each cluster for the MSP dataset.

CLUSTER	NODE NO.	MEMBER(S)
1	10	L13
	11	L14, L15
	13	L16
	14	L17
	15	L18
	16	L19
	17	L20
	18	L21
	19	L22
	20	L23
	21	L24, L25
2	23	SP1
	24	SP2
	25	SP3
	26	SP4
	27	SP5
	28	SP6
	29	SP8
	30	SP9
	8	SP13
	12	SP14
	22	SP16
3	1	L4
	2	L5
	3	L6
	4	L7
	5	L8
	6	L9
	7	L10, L11
	9	L12

3.5.3.3 HCA for the TLC and MSP Combined Dataset

Similar to the TLC and MSP datasets, no full discrimination of the samples under study was achieved when the TLC and MSP combined dataset was subjected to HCA. The best clustering strategy was by complete linkage revealing three clusters (Cluster 1, 2 and 3) as illustrated in Figure 3.28. Cluster 1 and 3 consisted of lipstick samples while cluster 2 consisted of shoe polishes. Table 3.15 describes in details the memberships of each identified cluster.

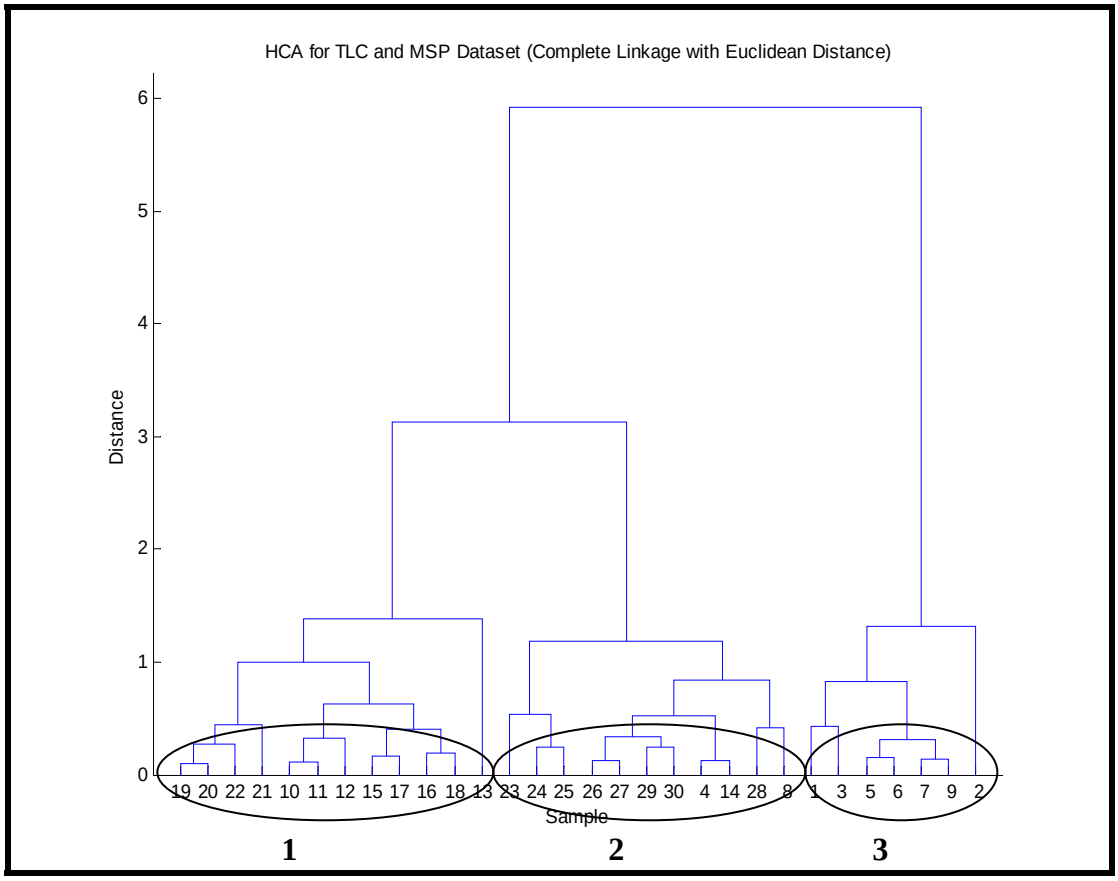


Figure 3.28 Dendrogram for TLC and MSP combined dataset constructed using Euclidean Distance and Complete Linkage

Table 3.15 Identified cluster and members in each cluster for the TLC and MSP combined dataset.

CLUSTER	NODE NO.	MEMBER(S)
1	19	L22
	20	L23
	21	L24
	10	L13
	11	L14
	12	L15
	15	L18
	17	L20
	16	L19
	18	L21
	13	L16, L17
2	23	SP1
	24	SP2
	25	SP3
	26	SP4
	27	SP5
	29	SP8
	30	SP9
	4	SP13
	14	SP16
	28	SP6
	8	SP14
3	1	L4
	3	L6, L7
	5	L8
	6	L9
	7	L10, L11
	9	L12
	2	L5

3.5.3.4 HCA for the UV/Vis and GC Square Root Combined Dataset

Poor discrimination was achieved when the UV/Vis and GC square root combined dataset was subjected to HCA. The dendrogram produced using complete linkage as illustrated in Figure 3.29 provides the best clustering strategy however no meaningful clusters were evident.

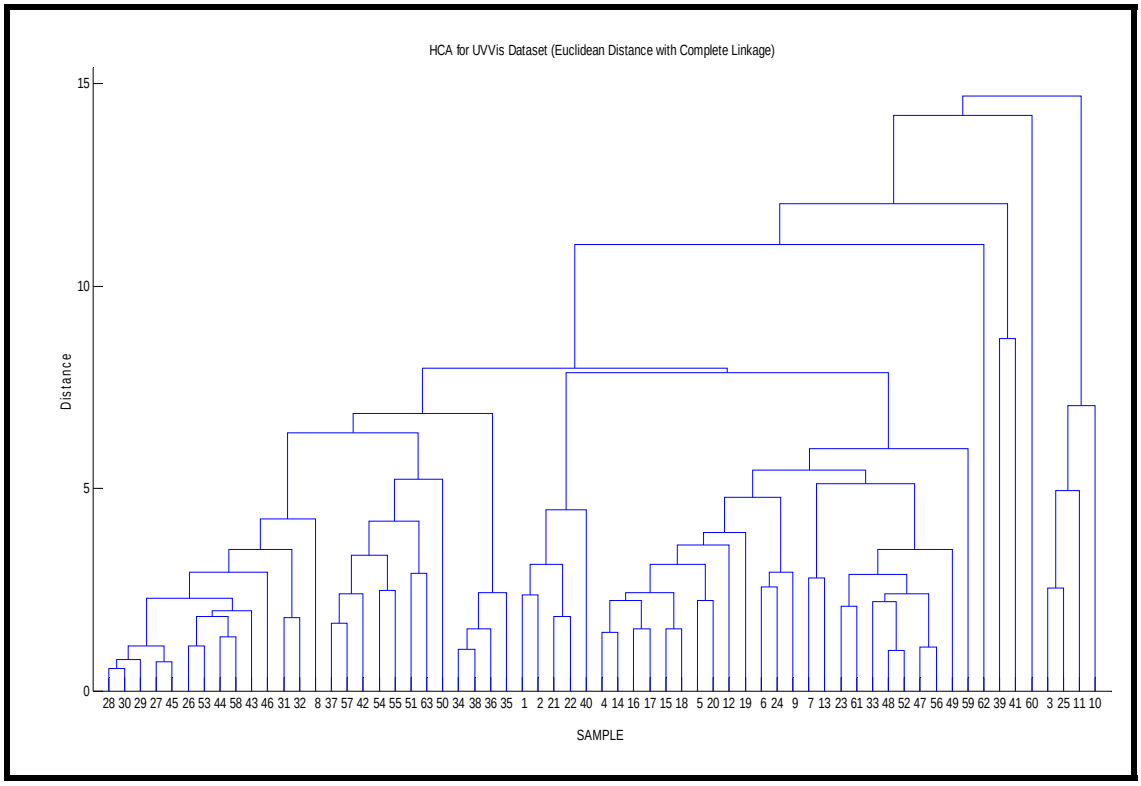


Figure 3.29 Dendrogram for the UV/Vis and GC square root combined dataset constructed using Euclidean Distance and Complete Linkage

3.5.4 Self Organising Feature Maps (SOFM)

Each dataset previously described were subjected to analysis using self organising feature map (SOFM) using Viscovery SOMine® 4.0.2 (Eudaptics) software. The SOFM output maps generated from the TLC, MSP, TLC and MSP combined dataset and UV/Vis and GC square root combined data sets are presented in the following sections.

3.5.4.1 SOFM of the TLC Dataset

The SOFM output map for the TLC dataset revealed four groupings and is illustrated in Figure 3.30. The first group consists mainly of the lipstick samples which occupy the neurons located mostly at the right hand side of the output map. The second group occupies the neurons located primarily at the centre and stretching to the top of the output layer and consists of a mixture of shoe polish and lipstick samples. The third group consists of shoe polish samples and occupies the neurons located at the upper left hand side of the output maps and finally the fourth group consists of a mixture of lipstick (L5, 16 and L17) and shoe polish samples (SP4, SP13, SP9, SP5 and SP8).

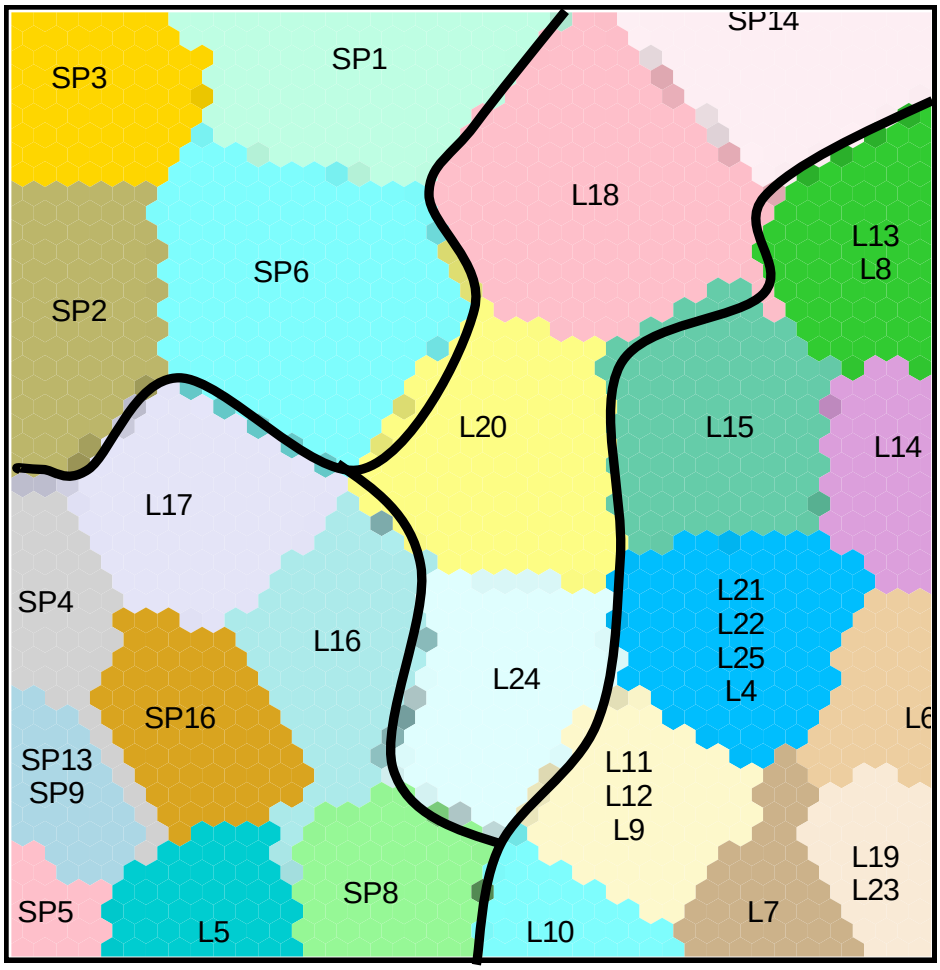


Figure 3.30 SOFM output map for the TLC dataset

An examination of the component map for the R_f value of the spot after the second development is illustrated in Figure 3.31 and reveals these sample groupings and illustrates that this variable is the predominant variable involved in the assignation of samples into the various groups.

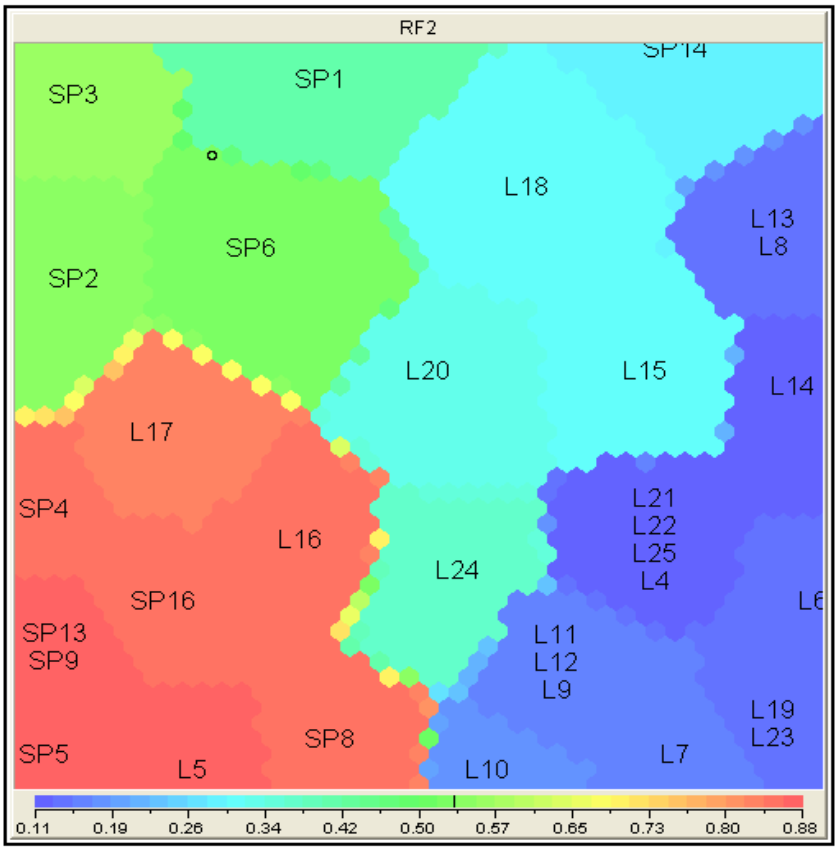


Figure 3.31 Component map representing the variable R_f2 of the TLC dataset

3.5.4.2 SOFM of the MSP Dataset

The SOFM output map for the MSP dataset is illustrated in Figure 3.32 and three groupings can be identified. The first two groups consists of the lipstick samples which are split and occupy the neurons mostly located at the left and right hand side of the output map. The third group contains the shoe polish samples and occupies the neurons located mostly at the centre and top of the output map with one sample split and separated from the main group. The component map of the percentage reflectance at

362nm illustrates the influence that variable has on grouping the samples and is illustrated in Figure 3.33.

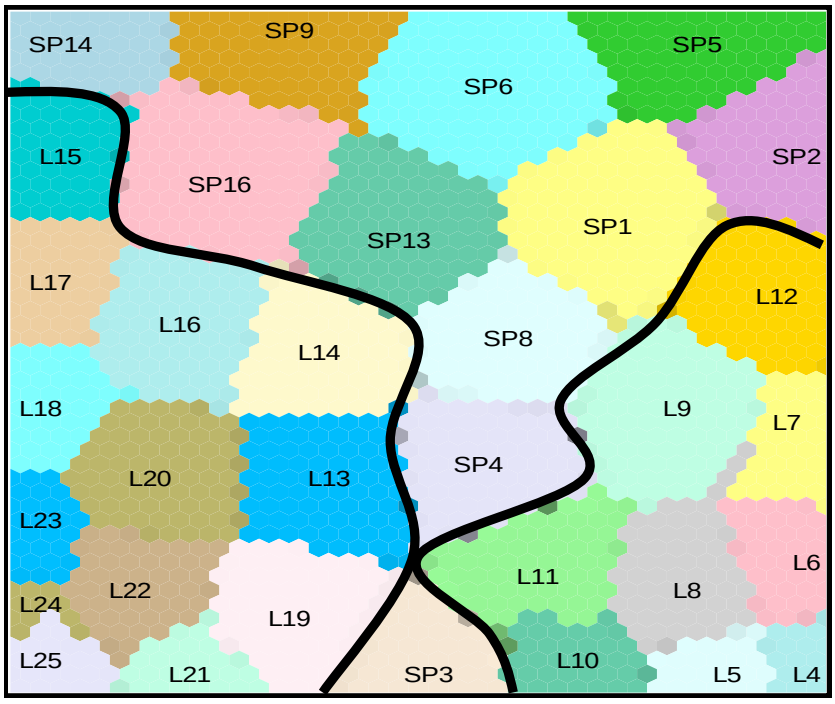


Figure 3.32 SOFM output map for the MSP dataset

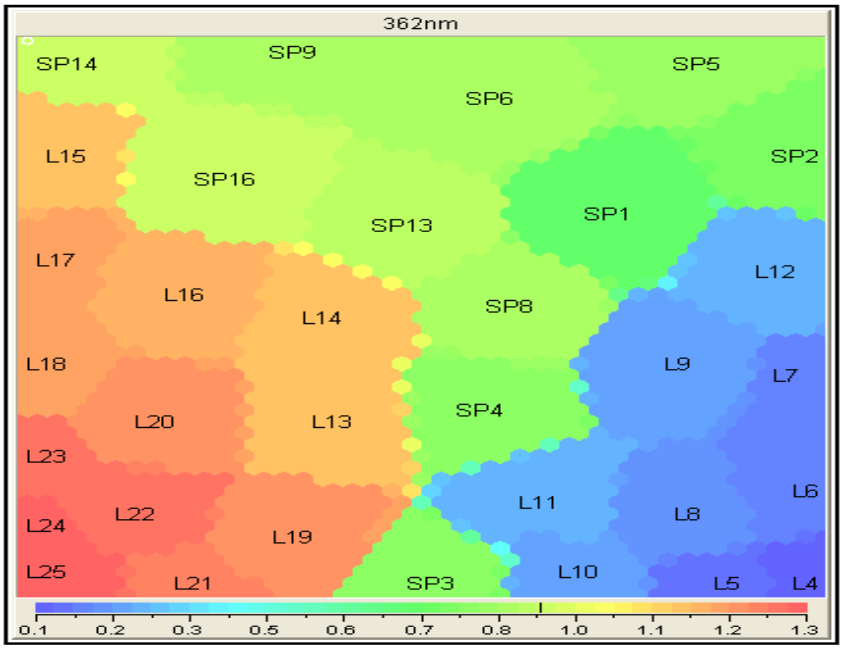


Figure 3.33 Component map representing the variable 362 nm of the MSP dataset

3.5.4.3 SOFM of the TLC and MSP Combined Dataset

The SOFM output map for the combined TLC and MSP dataset is illustrated in Figure 3.34 and clearly separates the samples into two distinct groupings based upon their product types. The first group occupying the bottom neurons of the output map group all of the lipstick samples whereas the entire shoe polish samples are grouped in the top portion of the output map.

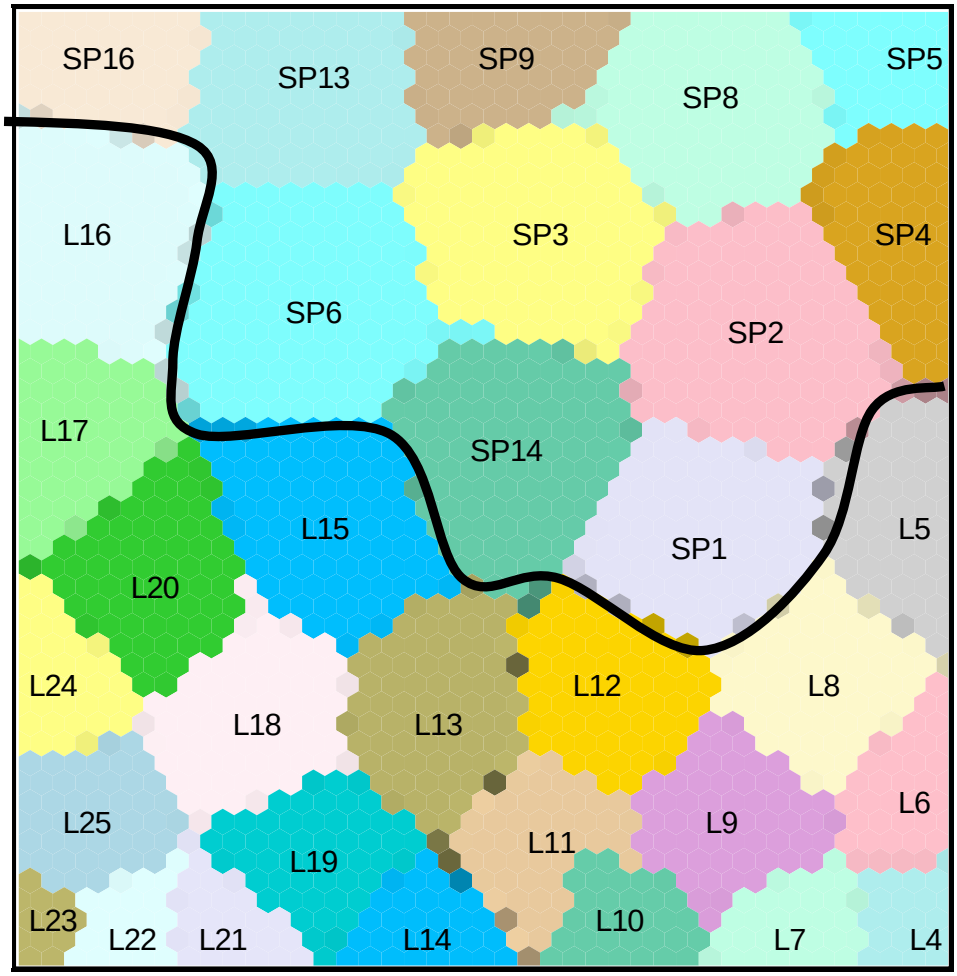


Figure 3.34 SOFM output map for the TLC and MSP combined dataset

This clearly illustrates that the SOFM technique was superior to both HCA and PCA in its abilities to correctly discriminate the samples presented based even on simple measurements of retardation factor and UV spectra.

3.5.4.4 SOFM of the UV/Vis and GC Square Root Combined Dataset

The SOFM output map for the UV/Vis and GC square root combined dataset is given in Figure 3.35. This sample set includes all 63 samples whereas the previous data sets excluded samples which were not coloured. All of the lipstick samples are mapped onto neurons located in the left hand region and all of the lip balm samples were mapped onto neurons on the right hand side of the output map. The shoe polish samples were all mapped onto neurons in the centre of the output map.

Again this is a very clear indication of the power of the SOFM technique in sample discrimination. All of the samples were successfully classified into their appropriate product types using in this case more complex data sets (UV/Vis and chromatographic data). No successful classifications were achieved using either HCA or PCA.

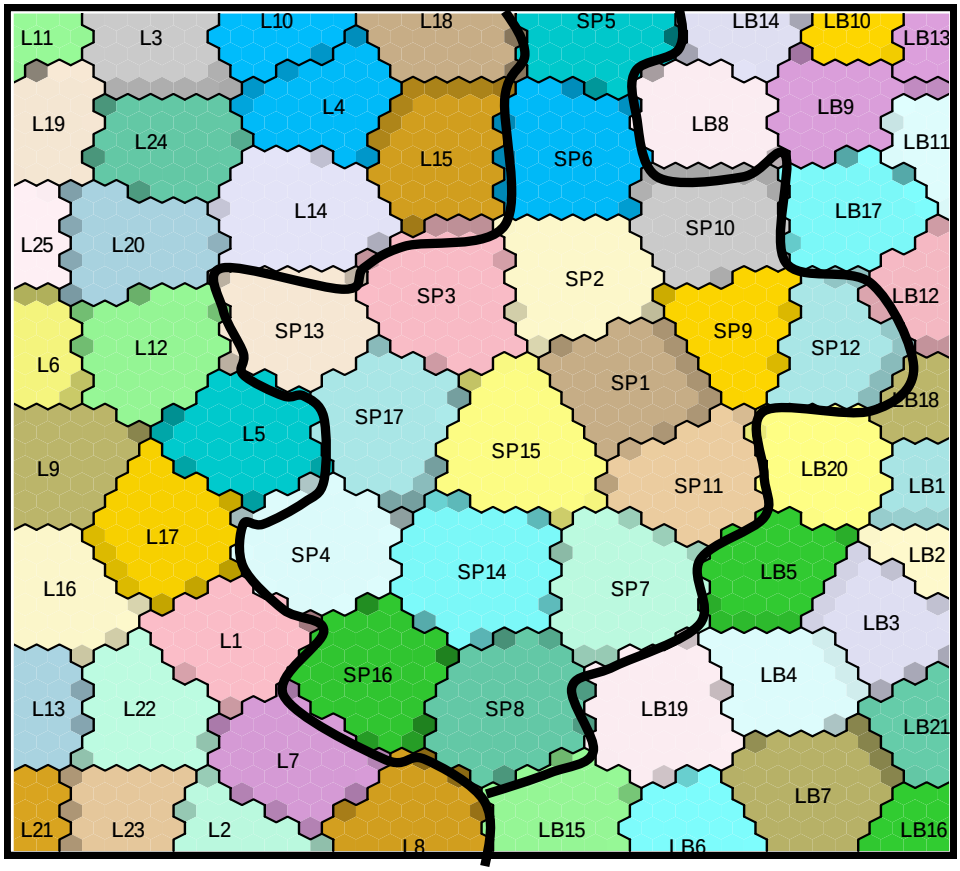


Figure 3.35 SOFM output map for the UV/Vis and GC square root combined dataset

An added advantage of SOFM is that the component maps can be used to determine the variable(s) which influence sample discrimination.

In total, fourty five (45) component maps were used in the production of the final successful output map. Analysis of each individual component map demonstrated that no one variable examined could account for the discrimination between products, rather a combination of the variables was required. This was hardly surprising given the nature and similarities of the products in question. Combinations of variables were examined and investigated to determine if any trends were evident across the data. At the visible part of the UV/Vis spectra, the lip balm samples demonstrated low or almost no absorbance, the shoe polish samples demonstrated an intermediate to low absorbance and the lipstick samples produced an intermediate to high absorbance. This is illustrated in Figure 3.36 to 3.37 where an increase in absorbance intensity is demonstrated by a red/yellow colour within the three different component maps. Since all samples were of same concentration, the increase in absorbance is not due to differences in concentration of the samples.

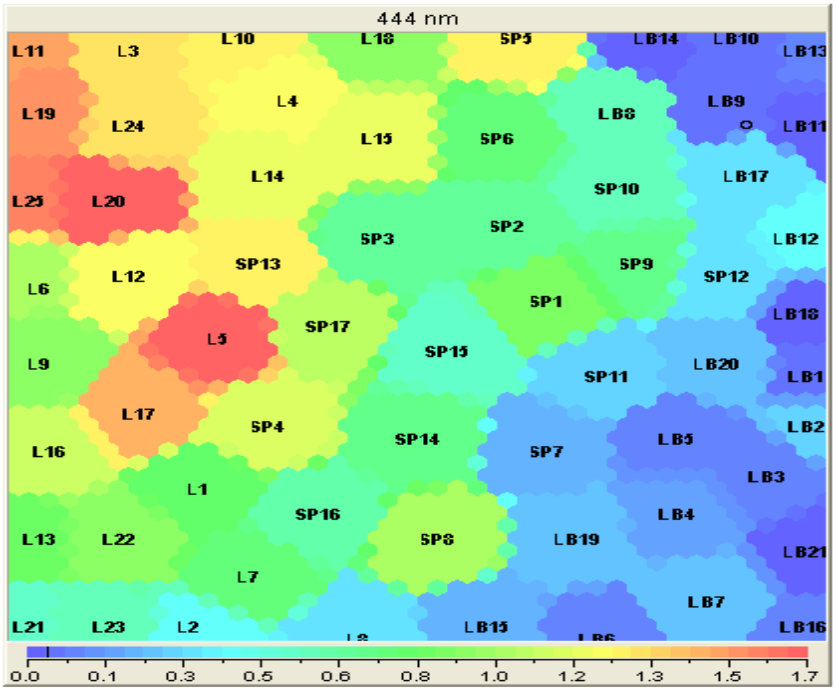


Figure 3.36 Component map for the variable representing 444 nm in UV/Vis data

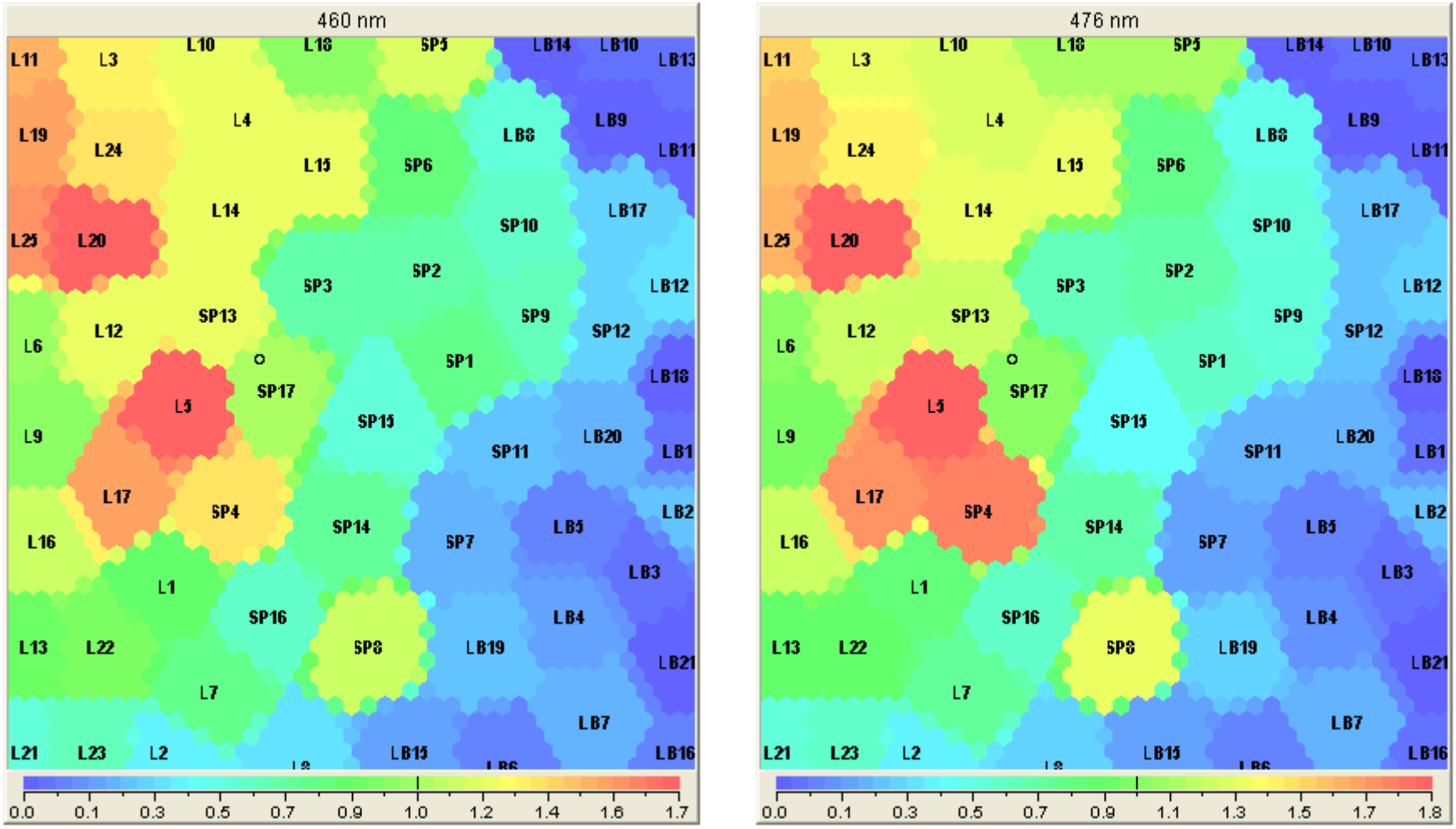


Figure 3.37 Component maps for the variables representing 460 and 476 nm respectively in the UV/Vis data.

Analysing the component maps for the variables derived from the GC profiles of the wax-based products allows for some generalisations to be made. Lipstick samples in general are constituted from a combination of low and high molecular weight hydrocarbons which are present in the five component maps corresponding to retention times of 23, 25, 28, 38 and 39 minutes and represented in Figures 3.38 to 3.40 respectively. Conversely, Lip balms and shoe polishes tend to have only high molecular weight hydrocarbons corresponding to retention times 33, 35, 38 and 39 minutes.

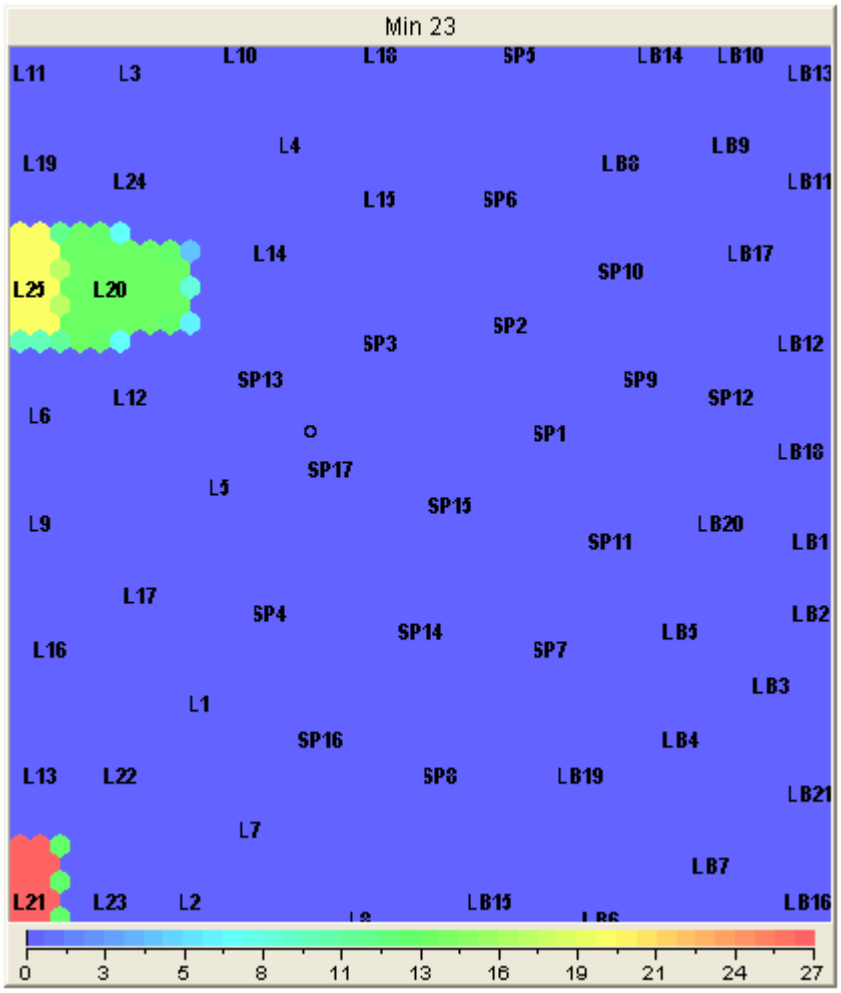


Figure 3.38 Component map representing the variable for 23 minutes for lower boiling point hydrocarbons which demonstrate higher scores for lipstick samples

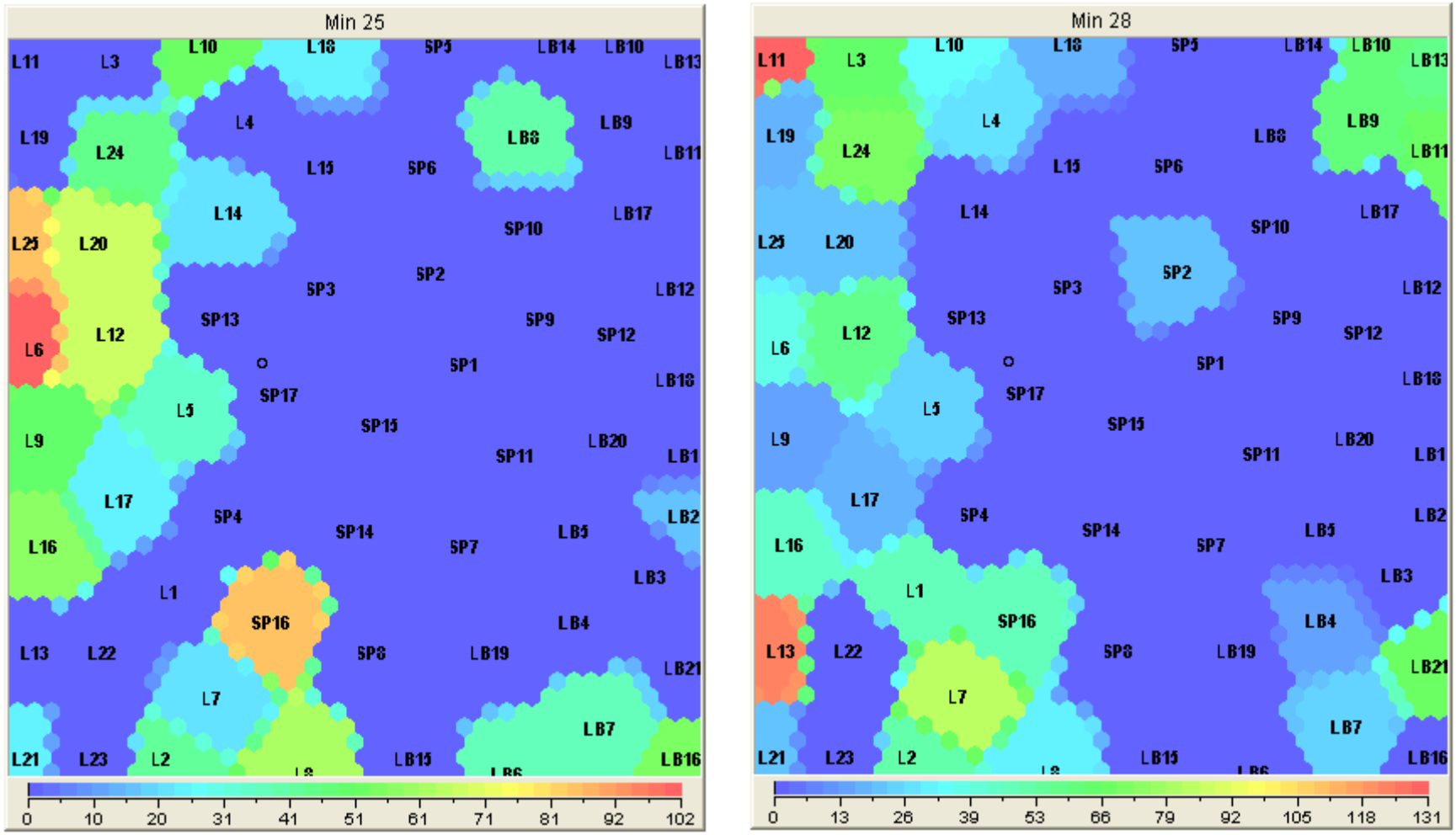


Figure 3.39 Component maps for the variables representing 25 and 28 minutes respectively for lower boiling point hydrocarbons which demonstrate higher scores for lipstick samples

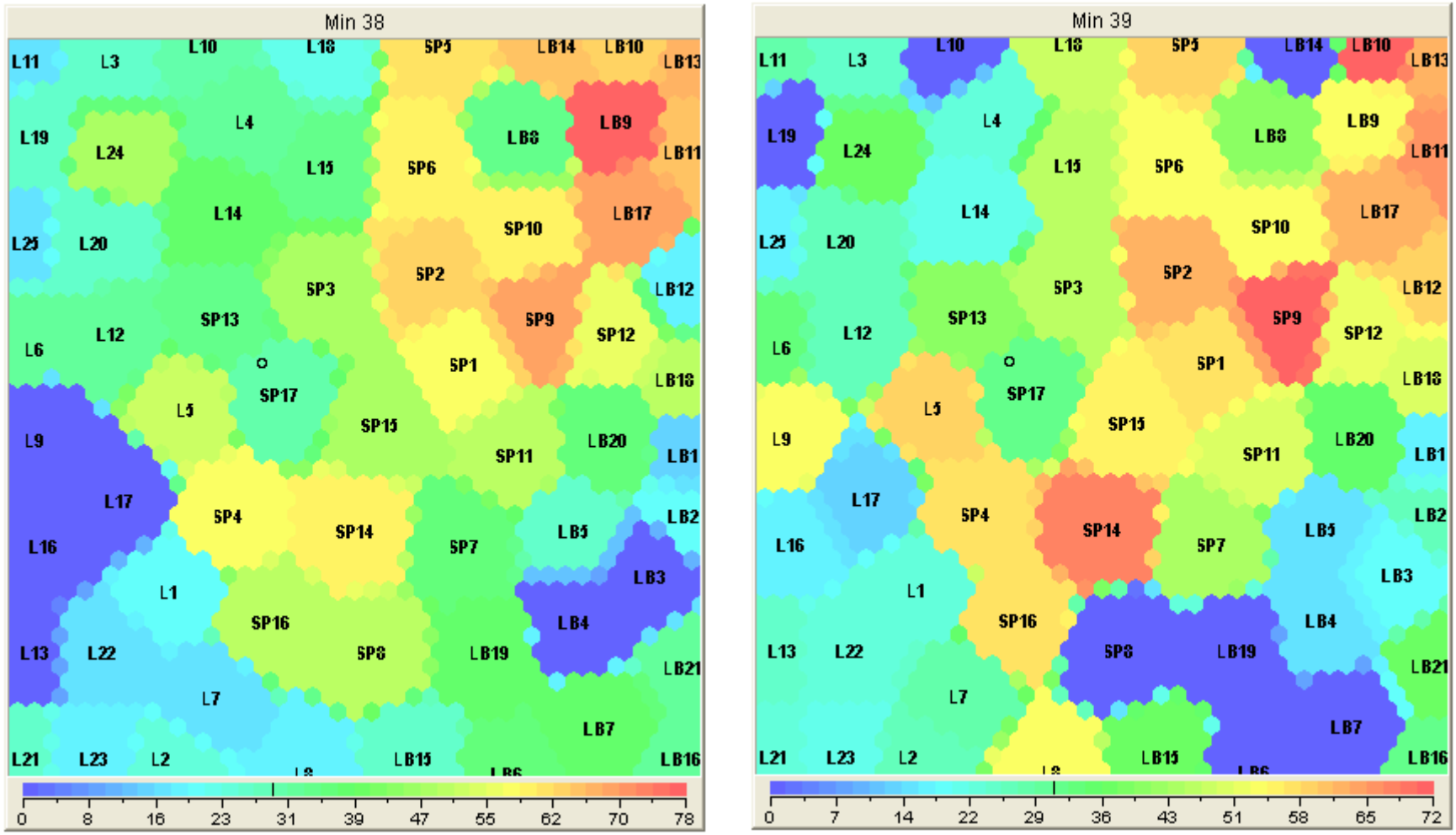


Figure 3.40 Component maps for the variables representing 38 and 39 minutes respectively for higher boiling point hydrocarbons which demonstrate higher scores for shoe polish and lip balm samples

Another factor of the sample type which is exposed within the component maps are the analytical responses of the non coloured wax based products which can be found in both lip balm and shoe polish (samples SP10, SP11, SP12, LB10, LB11, LB13, LB14, LB15, LB16, LB20 and LB21). For example, four of the lip balm samples (LB10, LB11, LB13 and LB14) can be readily distinguished from other non-coloured shoe polish and lip balm samples by the very high absorbance at the ultra-violet (UV) part of the spectra which is presented in the component map representing the variable 316 nm and is illustrated in Figure 3.41.

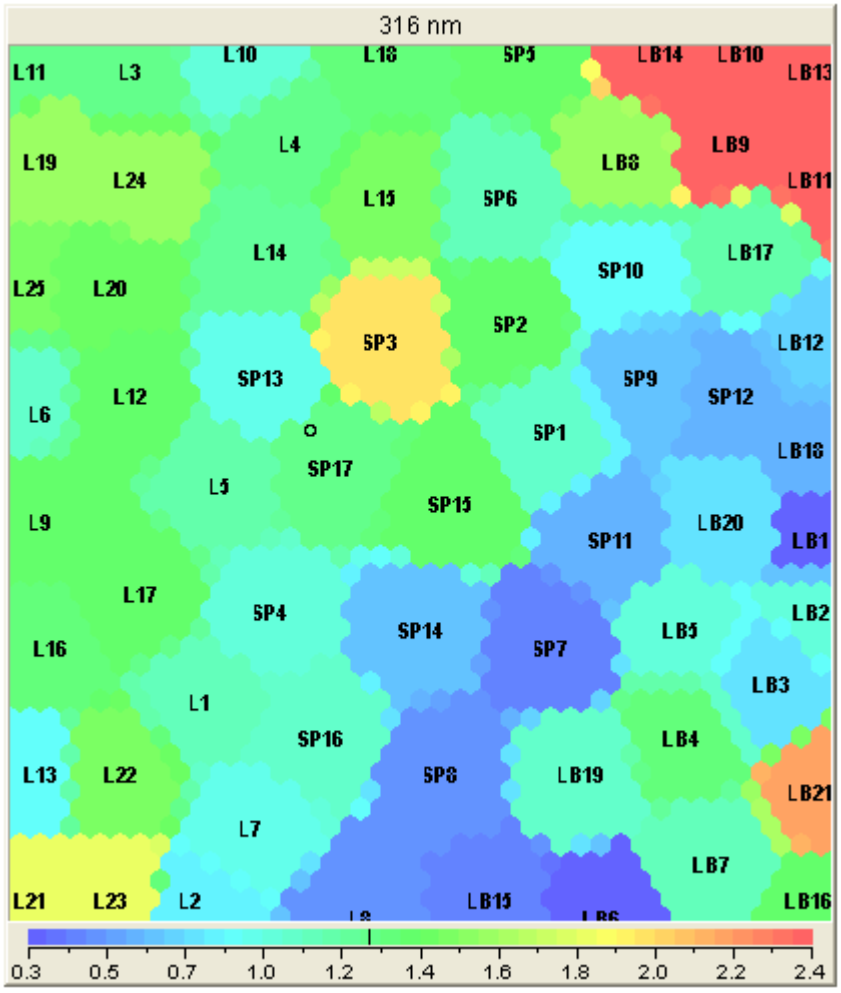


Figure 3.41 Component map for the variable representing 316 nm.

3.6 Determining the Effect of Ageing and Weathering of Wax Based Products Using SOFM.

3.6.1 Sample Selection and Preparation

Six different brands of samples consisting of two lipsticks, two lip balms and two shoe polishes as presented in Table 3.16 were used in the ageing and weathering study. All samples were prepared according to the sample preparation protocol previously outlined.

Table 3.16 List of samples used in the study. The code is given to identify the samples

SAMPLE NAME	TYPE	CODE	COLOUR
Rimmel London Rich Raisin	Lipstick	L7	Brown
TESCO Shade 12	Lipstick	L21	Red
Kiwi Black Shoe Polish	Shoe Polish	SP2	Black
Tuxan Shoe Polish	Shoe Polish	SP17	Brown
Mentholatum Soft Lips	Lip Balm	LB16	Colourless
Vaseline Rosy Lips	Lip Balm	LB19	Red

3.6.2 Degradation of Samples

The samples were degraded by exposing the sample smears to the open air at room temperature (approximately 20^{o}C) and in an oven at 56^{o}C for a period of 14 (two weeks) and 28 (four weeks) days. After each ageing and weathering period, a portion of the smears was extracted using the method described previously. The sample extracts were then analysed using TLC, MSP, UV/Vis spectrophotometer and Gas Chromatography. The GC profile of each sample extract was pre-processed using the square root power transformation. Undegraded sample smears were also prepared, extracted and analysed. In each case the following degraded samples were produced:

1. Room temperature 14 days @ 2 weeks (RT2W)
2. Room temperature 28 days @ 4 weeks (RT4W)
3. Oven 14 days @ 2 weeks (AC2W)
4. Oven 28 days @ 4 weeks (AC4W)

3.6.3 Dataset Preparation for SOFM

The TLC and MSP profile of each sample extract were combined to form the degraded TLC and MSP combined dataset while the UV/Vis spectra and the gas chromatographic profile of each sample extract were combined to form a degraded UV/Vis and GC square root combined dataset. Each combined dataset was introduced to SOFM analysis.

3.6.4 SOFM for the Degraded TLC and MSP Combined Dataset

The SOFM output map for the degraded TLC and MSP combined dataset is presented in Figure 3.42. Three groups were revealed on the output map and all of the samples were successfully resolved into their appropriate groups according to product type. The degraded samples were correctly associated with their undegraded sample but only by product type. No discrimination was possible by individual sample.

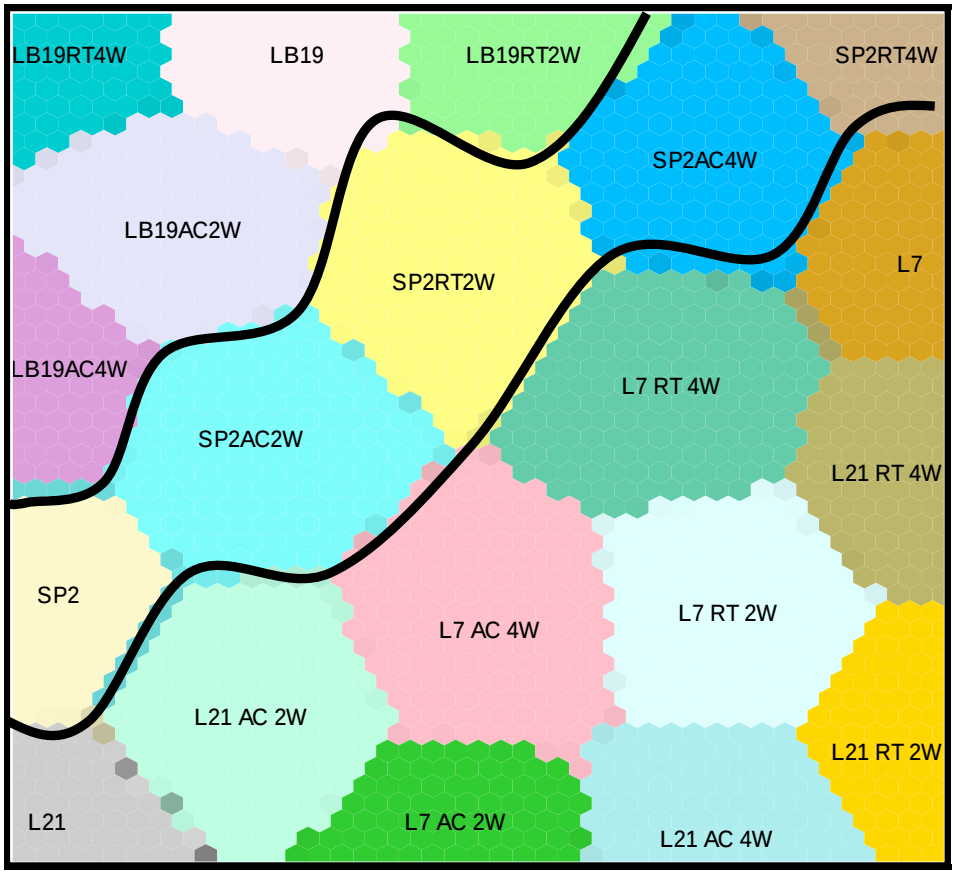


Figure 3.42 SOFM output map for the degraded TLC and MSP combined dataset

3.6.5 SOFM for the Degraded UV/Vis and GC Square Root Combined Dataset

The SOFM output map for the degraded samples dataset is presented in Figure 3.43. It is evident from the output map that six groups can be established which corresponds to the number of samples used in the study (See Table 3.16).

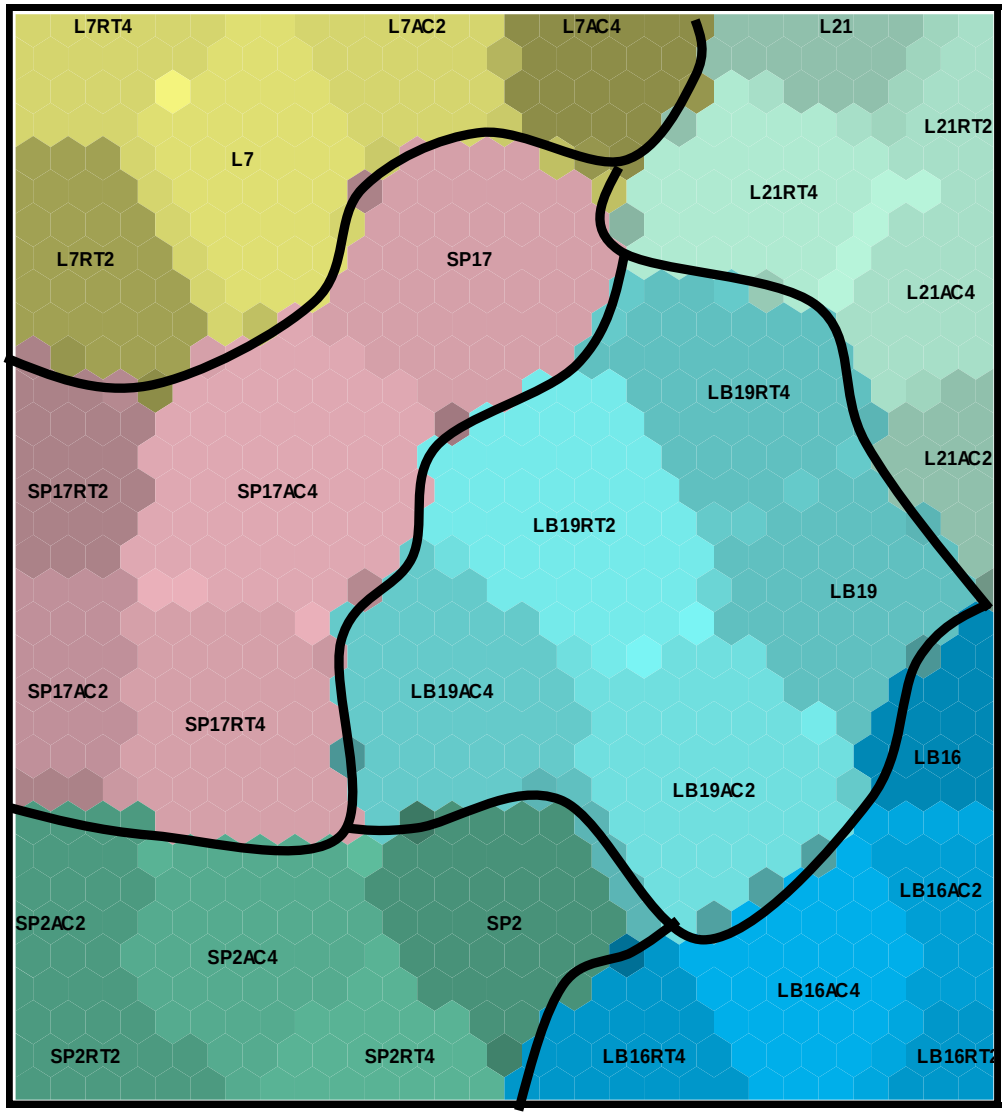


Figure 3.43 SOFM output map for the degraded samples dataset

All of the degraded samples were successfully grouped to the corresponding fresh samples with no overlap of the groups. This finding demonstrates that the SOFM neural network can successfully group related samples together and distinguish these related

samples successfully from other samples even of the same product class. This result also demonstrates, for the first time, that even in cases where considerable degradation may have occurred (for example at prolonged exposure to elevated temperature) the analytical regime in combination with SOFM can correctly associate related wax based samples together.

3.7 CONCLUSIONS

PCA and HCA analysis of the data sets presented did not facilitate complete grouping of the samples in any case. SOFM, on the other hand, grouped the wax based products according to type even with very basic data derived from a combination of TLC and MSP analysis. This provided a robust mathematical method for verifying the application of two independent, cost effective and rapid techniques for the discrimination of coloured wax based products by type.

SOFM analysis of the UV/Vis and GC square root combined data set was considered. All samples irrespective of their colour were discriminated into their separate product groups. Furthermore, the component maps elucidated with SOFM provided an insight into the strength of association provided by individual variables such as a specific retention time or wavelength. Again this provides valuable mathematical evidence of the validity of these techniques in the discrimination of such samples.

When degraded samples were examined, SOFM again facilitated the grouping by product of such samples using only TLC coupled with MSP analysis. When UV/VIS and GC-FID (pre treated using the square root method) data were considered, the discrimination obtained was at an individual sample level, whereby degraded samples were unambiguously associated to their specific parent sample. This demonstrates the power of the mathematical technique and potentially provides a mathematical means for the comparison and discrimination of a wide variety of samples of relevance in forensic science once a numerical value can be attached to a measurable attribute.

3.8 REFERENCES

1. Elvers, B. and Hawkins, S., *Ullman's Encyclopedia of Industrial Chemistry*. 5th Ed. (A28) , 1996, Weinheim: VCH.

2. Warth, A.H., *The Chemistry and Technology of Waxes*. 1956, New York: Reinhold Publishing Corporation.

3. Griffin, G.M.E., Doolan, K., Campbell, M., Hamill, J. and Kee, T.G., *Analysis of Wax Based Products by Capillary Gas Chromatogrpahy-Mass Spectrometry*. Science and Justice, 1996. **36**.

4. Barker, A.M.L. and Clarke, P.D.B., *Examination of Small Quantities of Lipsticks.* Journal of Forensic Science Society, 1972.

5. Keagy, R.L., *Examinations of Cosmetic Smudges Including Transesterification and Gas Chromatographic/Mass Spectrometric Analysis.* Journal of Forensic Sciences, 1983. **28**(3): p. 623 - 631.

6. Giles, J.J., *The Analysis of Waxes and Greases using High Resolution Gas Chromatography.* Journal of Forensic Science Society, 1987. **27**: p. 231 - 239.

7. Russell, L.W., *Analysis of Lipsticks.* Forensic Science International, 1984. **25**: p. 105 - 116.

8. Chaudry, M.Y., *Comparison of Minute Smears of Lipstick by Microspectrophotometry and Scanning Electron Microscopy/Energy Dispersive Spectroscopy.* Journal of Forensic Sciences, 1991. **36**: p. 366 - 375.

9. Williams, D.F. and Schmitt, W.H., *Chemistry and Technology of The Cosmetics and Toiletries Industry*. Second Edition ed. 1992, London: Chapman and Hall.

10. Harry, R.G., *Harry's Cosmeticology*. 6th Ed., 1973, London: Leonard Hill Books.

11. Elvers, B., Hawkins, S., Russey, W., and Schulz, G., *Ullman's Enclyclopedia of Industrial Chemistry*. 5th Ed. (A23), 1993, Weinheim: VCH.

13. Cole, M.D., Milligan, F. and Thorpe, J.W., *The Examination of Black Wax Shoe Polish Stains After Ageing and Weathering.* Journal of Forensic Science Society, 1994. **34**: p. 23 - 27.

CHAPTER 4: UNSUPERVISED PATTERN RECOGNITION TECHNIQUES APPLIED TO COMPLEX DATA DERIVED FROM THE ANALYSIS OF LIGHTER FUELS

4.1 INTRODUCTION

This chapter describes the application of the unsupervised Artificial Neural Networks (ANNs) technique of Self Organising Feature Maps (SOFM) and its comparison with the unsupervised pattern recognition techniques of Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) to a more complex dataset. Datasets were derived from the analysis of both un-evaporated lighter fuels from multiple sources and evaporated lighter fluid samples from the same sources. The challenge in this case was two fold, initially to discriminate between different brands of complex hydrocarbon mixtures and secondly to link degraded samples unambiguously to their original brands. The lighter fuels (both un-evaporated and evaporated), were obtained from five different manufacturers and were analysed using gas chromatography-mass spectrometry (GC-MS). The results demonstrated that SOFM acted as a powerful means of evaluating and linking degraded lighter fuels to their parent un-evaporated liquids. The findings indicate the usefulness of SOFM for complex datasets of forensic science interest.

4.2 Ignitable Liquids

Ignitable liquids are commonly used as accelerants to intensify the rate of fire development in cases of deliberate fire setting. Identification and discrimination of ignitable liquid or ignitable liquid residues within recovered samples are, therefore, of interest and the characterisation, identification and linkage of such ignitable liquids is highly desirable. At present, gas chromatography-flame ionisation detector (GC-FID) and gas chromatography-mass spectrometry (GC-MS) are widely accepted as effective methods for the analysis and identification of ignitable liquids [1-3]. Interpretation of the instrumental data and sample classification processes are based mainly on visual comparison of the sample to an ignitable liquid database [4-6]. Other approaches involve selective ion monitoring of the resultant chromatographic data producing a target compound chromatogram (TCC) of the sample which can be compared to TCC of the

ignitable liquid from a database [7-10]. These methods however can be difficult, time consuming, highly subjective and rely heavily on the skill and experience of the analyst. Petrol for example, consists of a wide range of over 500 hydrocarbon compounds together with other additives [11]. Exposure to heat or aging can result in substantial changes in the liquid's composition, which in turn greatly affects their chromatographic profile. Another common complication encountered in fire debris analysis is the appearance of hydrocarbon by-products from the combustion and pyrolysis of background matrices [12-15].

The applications of multivariate pattern recognition to discriminate and classify ignitable liquid samples are suggested as a means of facilitating the pattern matching process and rendering it less subjective. Pattern recognition techniques using chemometric approaches have been used to establish underlying relationships amongst variables within complex datasets and can be used to differentiate groups within a given dataset. Principle component analysis (PCA) and hierarchical cluster analysis (HCA) are both regarded as unsupervised learning methods which do not require a training set of known categories to derive the classification model. Instead they use the given data to self-establish grouping structures [16, 17].

PCA and HCA have been employed for the classification of ignitable liquid analysis [18-22]. Typically, numerical datasets of selected compounds within the sample are processed by PCA, Canonical Variate Analysis (CVA) and Orthogonal Canonical Variate Analysis (OCVA) coupled with linear discriminate analysis (LDA) for discrimination. Petraco *et al.* demonstrated the use of such methods for the discrimination of 20 petrol samples, however some misclassification occurred [22]. The authors emphasised the importance of statistical methods of pattern recognition for fire debris casework rather than relying on visual comparison alone.

Doble *et al*. demonstrated the application of supervised ANNs using a multilayer perceptron (MLP) arrangement with a back propagation (BP) algorithm to discriminate

between regular and premium petrol and reported a 97% correct classification-prediction rate [20]. To date no other application of ANN in ignitable liquid analysis has been reported and there is no reported application of Self Organising Feature Map (SOFM) [23, 24].

In ignitable liquid analysis, the relevant chromatographic data are isolated by targeting compounds identified as discriminating for that particular sample [22, 25-27]. The selected data may also undergo further transformations or data pre-treatment (such as normalisation, logarithmic transformation and root transformation), which are commonly used to minimise the effect of large peaks or eliminate signal noise in order to make the data distribution more symmetrical and facilitate multivariate analysis [24, 28-31].

4.3 EXPERIMENTAL

4.3.1 Materials and Methods

Fifteen re-fill lighter fluid samples from 5 different brands (Zippo, Swan, Ronsonol, Perma and Dunhill) were purchased from commercial outlets. For each lighter fluid brand, a set of partially evaporated samples was generated. This was achieved by gently heating the un-evaporated lighter fluid (100mL) and removing a sub sample at specific intervals when the original liquid had been reduced by 10, 25, 50, 75, 90 and 95 mL in volume. This produced a set of partially evaporated samples for each lighter fluid liquid at approximately 10, 25, 50, 75, 90 and 95 percent evaporation. Each sample was diluted to 2% in pentane (HPLC grade, WVR International) with 0.5mg/mL Tetrachloroethylene (Sigma Aldrich >99% purity) as internal standard. All samples were stored in screw capped vials and darkness at room temperature until analysed. An aluminium foil shield was placed inside the screw cap to prevent evaporation of the sample. The prepared samples were at the approximate percentages of evaporation of interest and as such, any additional evaporation which may have occurred during storage was considered to be of little effect to the overall results obtained.

According to the ASTM E1618 guidelines, pentane is listed as one of the recommended non-polar solvents for fire debris analysis and is as efficient as other solvents but relatively less toxic, safer and easier to handle [1, 32]. The solvent delay on the mass-spectroscopy detector (MSD) was set in order to allow the pentane peak to elute undetected and this has minimal effect on the overall patterns of interest in the test samples, particularly in relation to the evaporated samples. The application and efficiency of using a volatile chlorinated compound as the internal standard for fire debris analysis, such as Tetrachloroethylene has been reported [33, 34].

4.3.2 Instrumentation

Gas chromatographic analysis was performed using a Hewlett Packard (HP) 6890/5973 gas chromatograph with a mass selective detector (GC-MSD). Data acquisitions were performed by MS ChemStation (version B.00.01, Hewlett Packard, Agilent Technologies). Chromatographic separation was achieved using a DB1-MS fused silica capillary column (25.0m x 0.20mm internal diameter x 0.33µm film thickness). The injection port temperature set at 250ºC. The oven temperature was set at 40ºC for 5 minutes and ramped at a rate of 15ºC/min to 280ºC, and maintained at this temperature for 2 minutes. Helium was used as the carrier gas and was maintained at a constant flow rate of 1.0mL/min. The temperatures of the ion source and the quadruple were set at 150ºC and 280ºC respectively. The MS analyses were performed at full scan mode (from 30 to 300 AMU) with two minutes solvent delay. Injections were carried out using a 7673A Hewlett Packard automatic liquid sampler. Each sample was analysed in triplicate and the injection volume for each sample was 1µL with a 20:1 split ratio.

4.3.3 Data Collection and Pre-processing

The chromatograms were visually compared and components of similar retention time and relative standards deviation of less than 5% on triplicate analysis were selected. Peak response data was acquired as comma separated values (CSV) files and converted into an Excel (Version 10) spread-sheet for ease of use. In total, 51 components were selected as variables from the pure and evaporated samples. Missing components in any

individual sample were given a zero value. The peak area response were normalised against the internal standard. Further data pre processing was undertaken by applying a square root transformation and a fourth root transformation to the normalised data set. The four resultant datasets (i.e. raw and processed) were inputted to MATLAB® 2008a (Version 7.6, Mathworks Inc.) where principal component analysis (PCA) and hierarchical cluster analysis (HCA) were performed. SOFM-artificial neural network analysis was performed using Viscovery®SOMine (Version 5.0.2, Viscovery Software GmbH).

4.4 RESULTS AND DISCUSSION

4.4.1 Chromatographic Analysis

Initial examination of the chromatographic pattern of each lighter fluid sample from each of the 5 brands revealed compositional differences for Swan, Dunhill and Zippo samples. By contrast Perma and Ronsonol revealed chromatographic patterns which were very similar to each other as shown in Figure 4.1. Identification of the components common in the various lighter fluid samples is presented in Table 4.1. Identification was made based on NIST mass spectral database using AMDIS software (Version 2.0) [35].

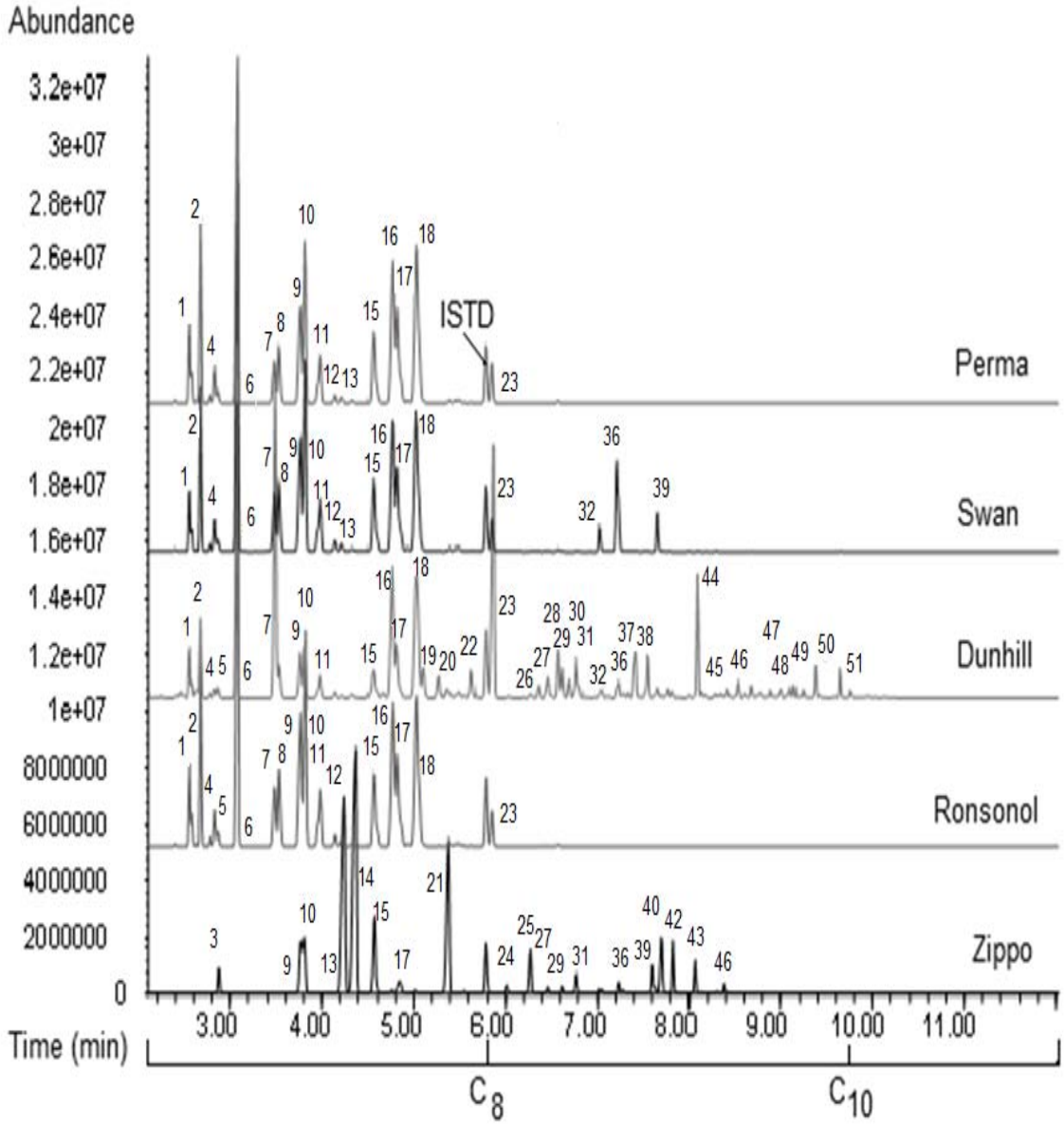


Figure 4.1 Chromatograms of pure lighter fluid samples (2% in pentane with 0.5mg/mL Tetrachloroethylene as ISTD).

Table 4.1 List of components identified in lighter fuel sample

No.	Retention Time	Peak Identification
1	2.55	2-Methylhexane
2	2.68	3-Methylhexane
3	2.84	3-Ethylpentane
4	2.89	2,2,4-Trimethylpentane
5	3.08	Heptane
6	3.47	1,2-Dimethylcyclopentane
7	3.48	Cyclomethylhexane
8	3.53	2,2-Dimethyl-3-hexene
9	3.77	2,5-Dimethylhexane
10	3.81	2,4-Dimethylhexane
11	3.99	3,3-Dimethylhexane
12	4.15	1,2,3-Trimethylcyclopentane
13	4.23	2,3,4-Trimethylpentane
14	4.37	2,3,3-Trimethylpentane
15	4.56	2,3-Dimethylhexane
16	4.77	2-Methylheptane
17	4.82	4-Methylheptane
18	5.03	3-Methylheptane
19	5.11	1,4-Dicyclohexane
20	5.28	1,1-Dicyclohexane
21	5.37	2,2,4-Trimethylhexane
22	5.64	1,2-Dicyclohexane
23	5.85	Octane
24	6.01	2,4,4-Trimethylhexane
25	6.28	2,3,5-Trimethylhexane
26	6.36	2,2-Dimethylheptane
27	6.47	2,4-Dimethylheptane
28	6.58	Ethylcyclohexane
29	6.63	2,6-Dimethylheptane
30	6.70	1,1,3-Trimethylcyclohexane
31	6.78	2,5-Dimethylheptane
32	7.02	2,3,4-Trimethylhexane
33	7.04	3-Methylheptane
34	7.06	1,3,5-Trimethylcyclohexane
35	7.21	m-Xylene

Table 4.1 Continued

No.	Retention Time	Peak Identification
36	7.23	2,3-Dimethylheptane
37	7.43	2-Methyl octane
38	7.56	3-Methyl octane
39	7.66	p-Xylene
40	7.71	2,3,6-Trimethylheptane
41	7.78	1-Ethyl-4-methylcyclohexane
42	7.83	2,2,4-Trimethylheptane
43	8.07	3,3-Dimethyloctane
44	8.09	Nonane
45	8.30	3,4-Dimethyloctane
46	8.41	2,4-Dimethyl-3-ethylpentane
47	8.54	cis-1,1,3,5-Tetramethylcyclohexane
48	8.69	2,6-Dimethyloctane
49	9.12	1-ethyl-1,3-dimethylcyclohexane
50	9.15	Toluene
51	9.65	Decane

The effect of evaporation of an ignitable liquid is twofold - lower boiling compounds diminish or are lost completely while higher boiling compounds increase in their abundance relative to neighbouring compounds, and as such the chromatographic profiles of evaporated samples are different to those of pure samples [5]. The exemplar of this is illustrated in Figure 4.2 which clearly demonstrates substantial changes in peak abundance.

4.4.2 PCA Classification

PCA was performed on each data set (raw data, normalised data only, normalised square root and normalised fourth root) to assess the effectiveness of the method to distinguish between the various lighter fluid brands and whether it was possible to establish a link between the pure and evaporated samples of a specific lighter fluid brand. The score plots obtained when all datasets were subjected to PCA analysis are given in Figure 4.3. Each score plot obtains a series of principal components (PC) where these values represent variances within the data sets.

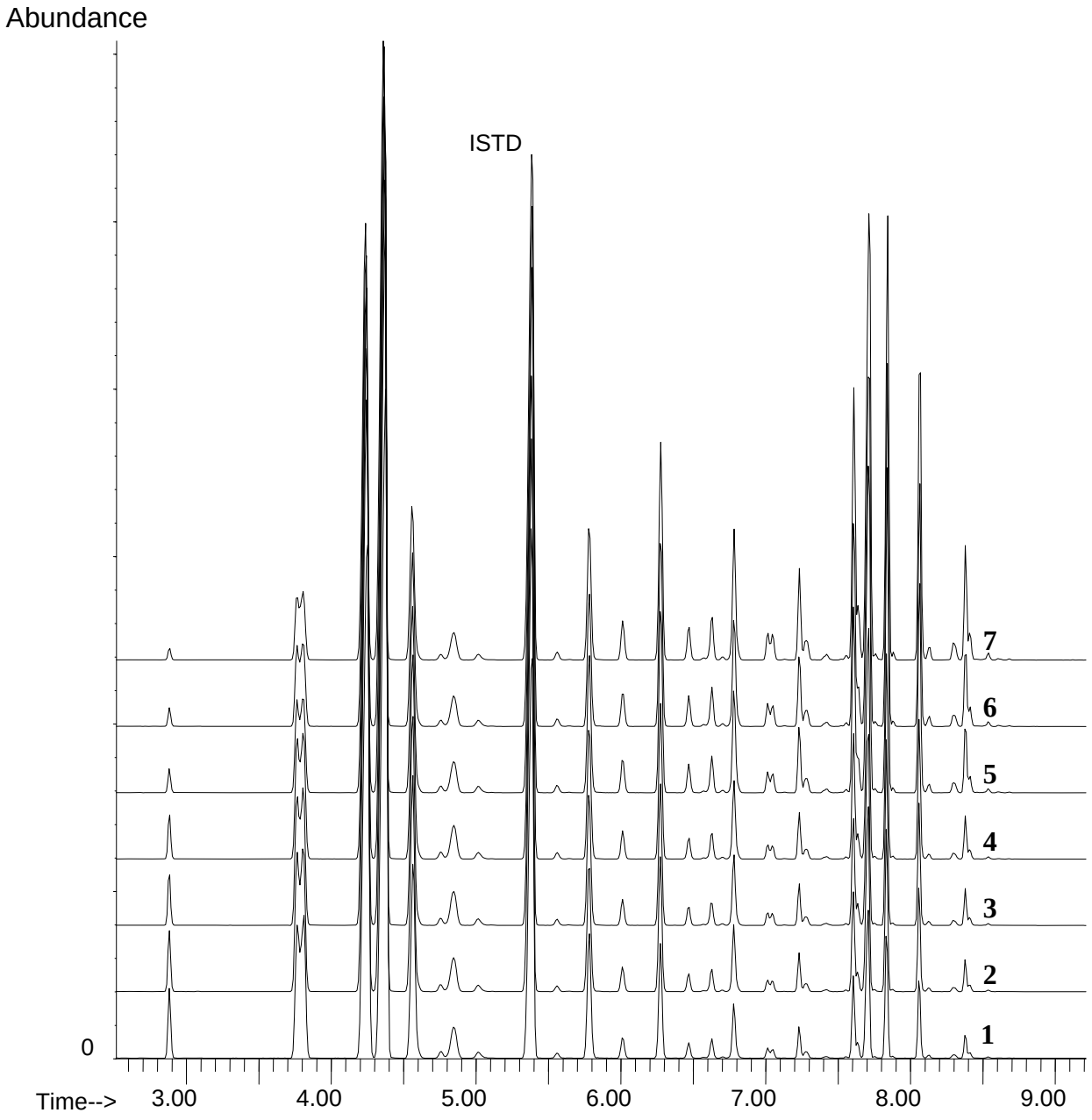


Figure 4.2 Chromatograms of the pure Zippo lighter fluid and evaporated Zippo lighter fluid to demonstrate gradual changes in major peaks abundance as the sample evaporated. 1= Neat, 2 = 10% evaporated, 3 = 25% evaporated, 4 = 50% evaporated, 5 = 75% evaporated, 6 = 90% evaporated and 7 = 95% evaporated.

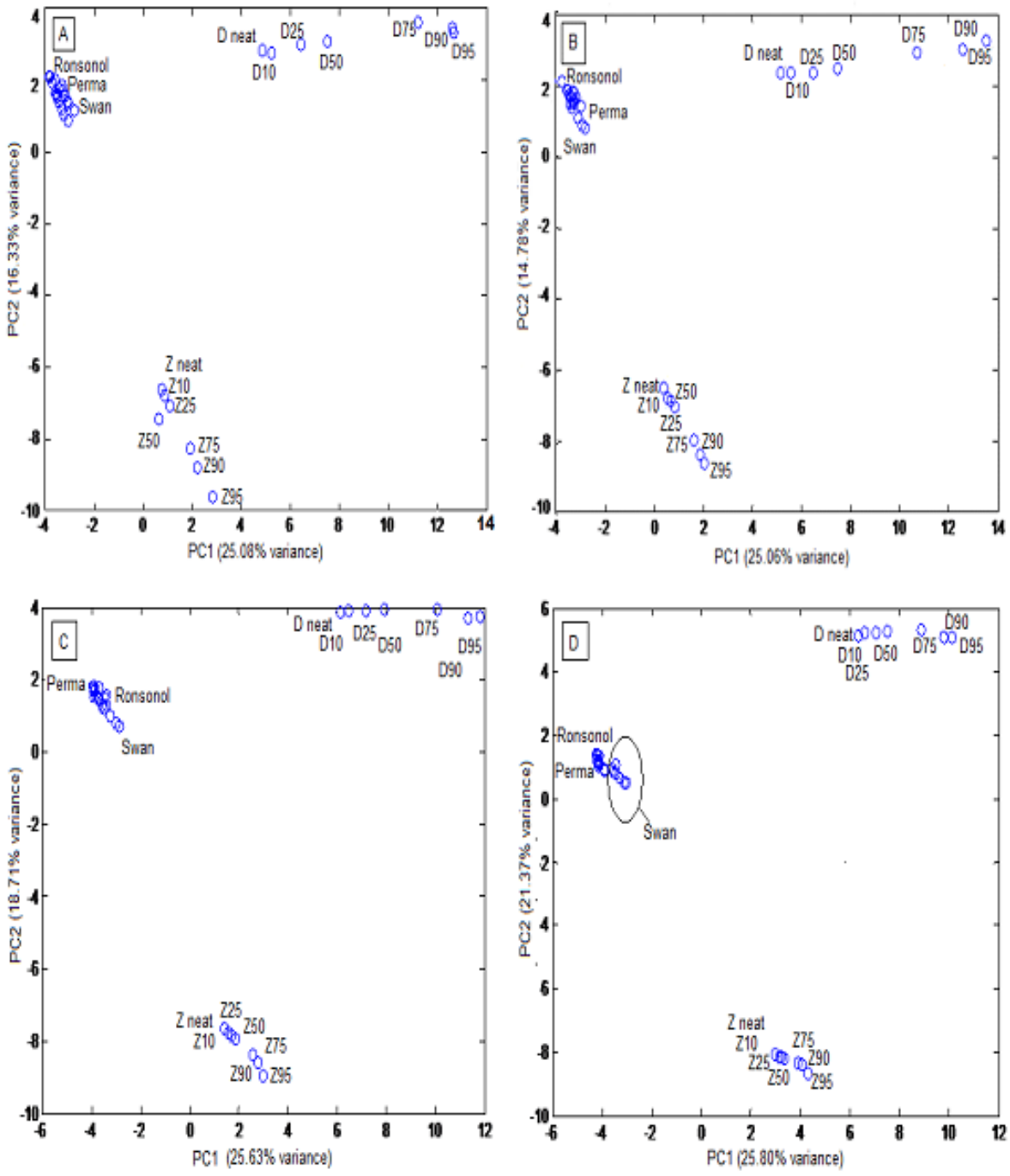


Figure 4.3 Principal component score plots of pure and evaporated samples. A, B, C and D represent plots of raw data, normalised, normalised square root and normalised fourth root transformation datasets respectively. (D=Dunhill, P=Perma, R=Ronsonol, S=Swan and Z=Zippo. Numbers represent degree of evaporation)

For all of data sets analysed only two brands of lighter fluid, Dunhill and Zippo were successfully resolved irrespective of the data pre-treatment method used. The rest of the lighter fluids samples (Perma, Ronsonol and Swan) were grouped into one group and as such PCA was not useful as a clustering technique.

4.4.3 HCA Classification

HCA was performed using the same four datasets used in PCA. Similarities between the samples were measured using complete linkage and the results are shown in Figure 4.4.

Like PCA, HCA did not correctly classify the samples by brand when using the raw data, normalised or normalised square root data sets. However, in contrast to PCA, the normalised fourth root data set produced a HCA classification which was capable of separating all of the pure and evaporated samples by brand and is shown in Figure 4.4 (D). The HCA classification suggests similarities between Swan and Perma brand samples as well as similarities between both of these samples and those of the Dunhill and Zippo brands; however this was not reflected in the visual comparison of the chromatographic profiles of these samples where clear differences were in evidence. The Ronsonol sample set was clustered away from all other sample sets including the Perma sample set even though their chromatographic profiles were visually similar.

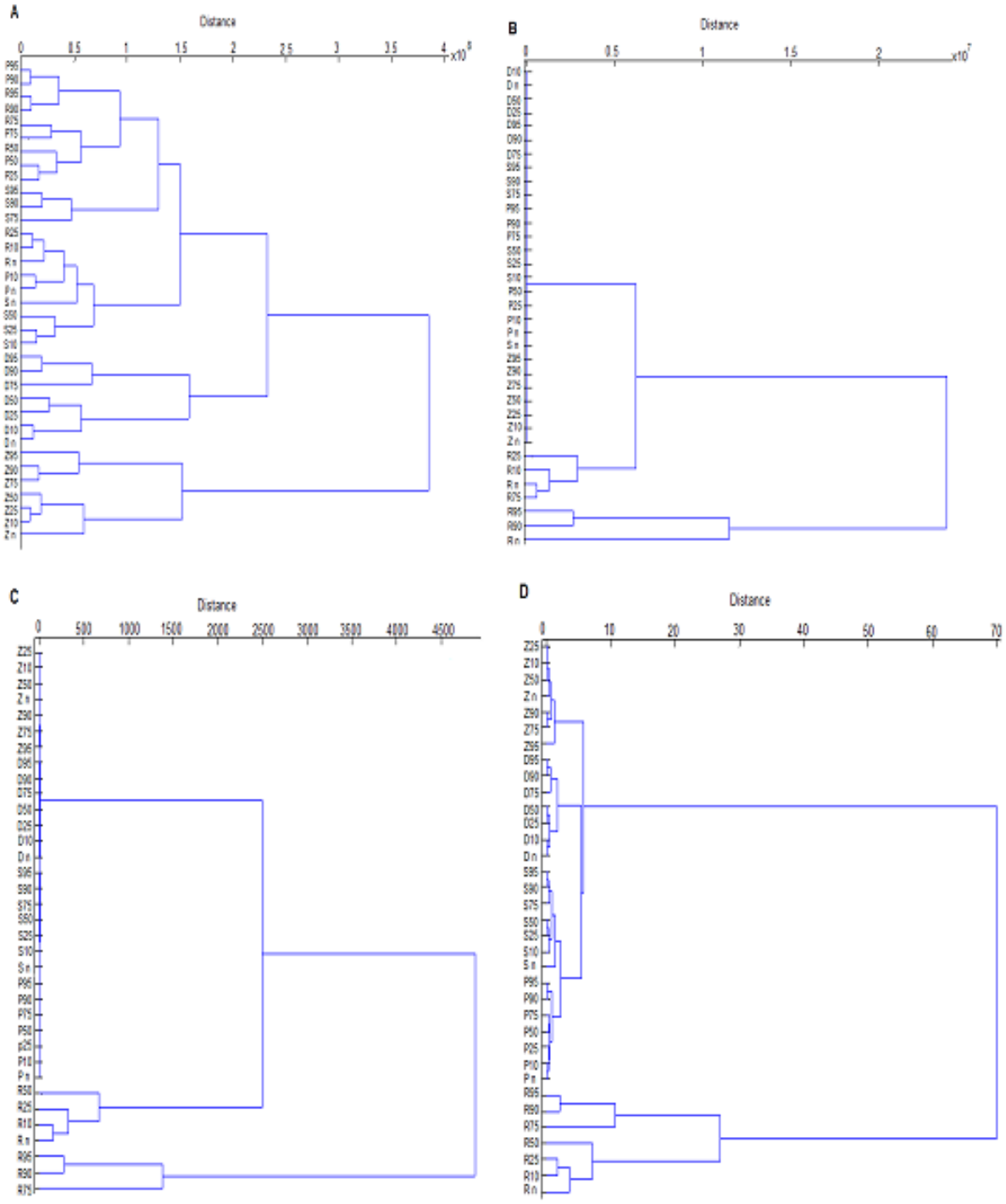


Figure 4.4 Hierarchical clustering of pure and evaporated samples. A, B, C and D represent dendrogram of raw data, normalised, normalised square root and normalised fourth root transformation datasets respectively. (D=Dunhill, P=Perma, R=Ronsonol, S=Swan and Z=Zippo. Numbers represent degree of evaporation)

4.4.4 SOFM Classification

The ability of SOFM to cluster related samples using the four data sets is shown in Figure 4.5 to 4.8. Data visualisation in SOFM can be accomplished by a number of techniques which can be a hit histogram, component planes and the U-Matrix display [24]. The one that is shown here follows the U-Matrix display where the groupings and the degree of dissimilarity between groupings are shown by the more intensely-coloured boundary lines, where darker boundary lines reflect a greater distance between the adjoining samples.

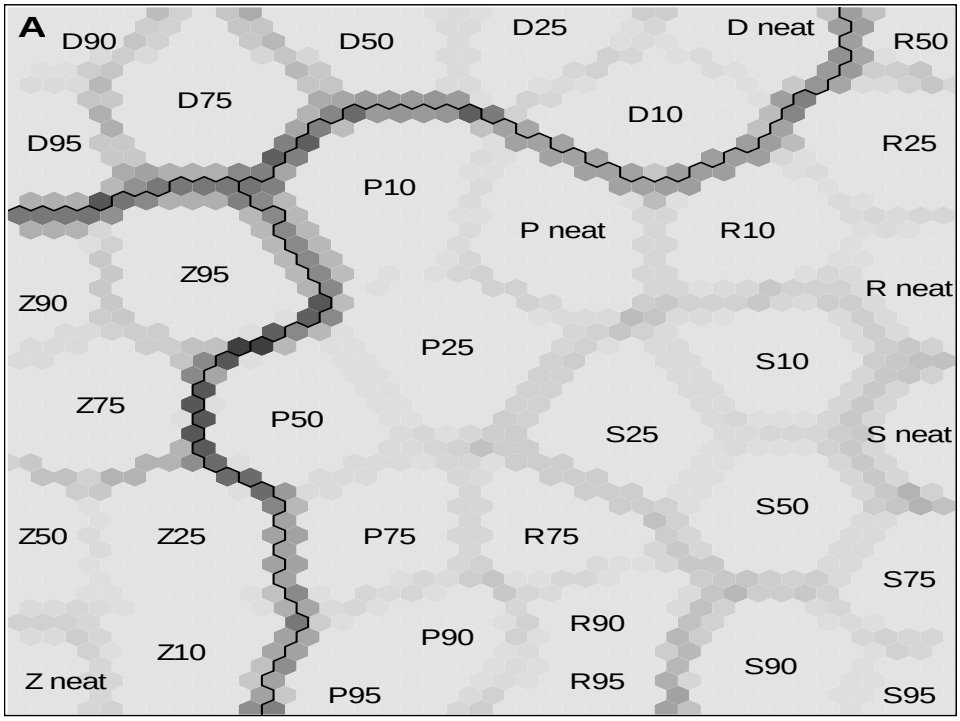


Figure 4.5 U-matrix displays of raw dataset, R = Ronsonol, Z = Zippo, S = Swan, D = Dunhill, P = Perma

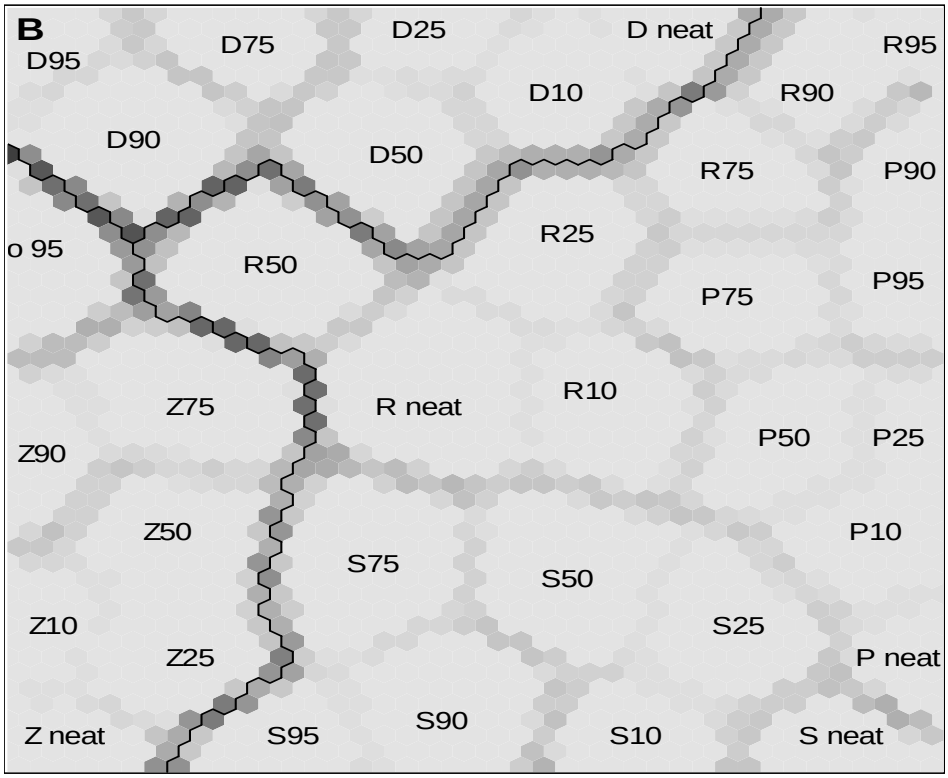


Figure 4.6 U-matrix displays of normalised dataset

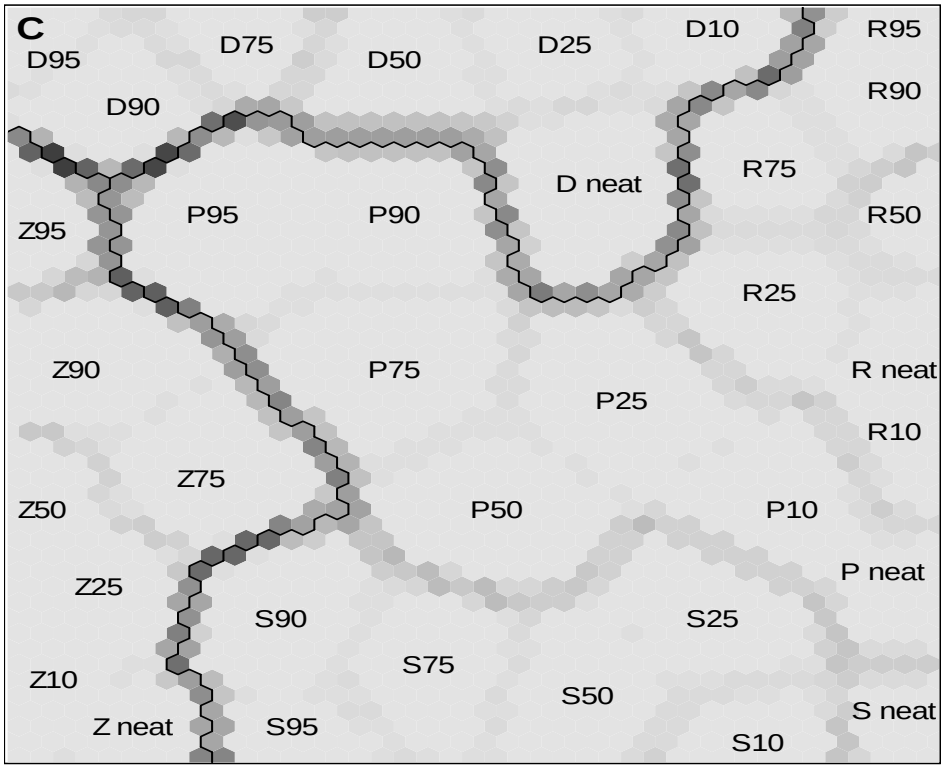


Figure 4.7 U-matrix displays of normalised squared root dataset

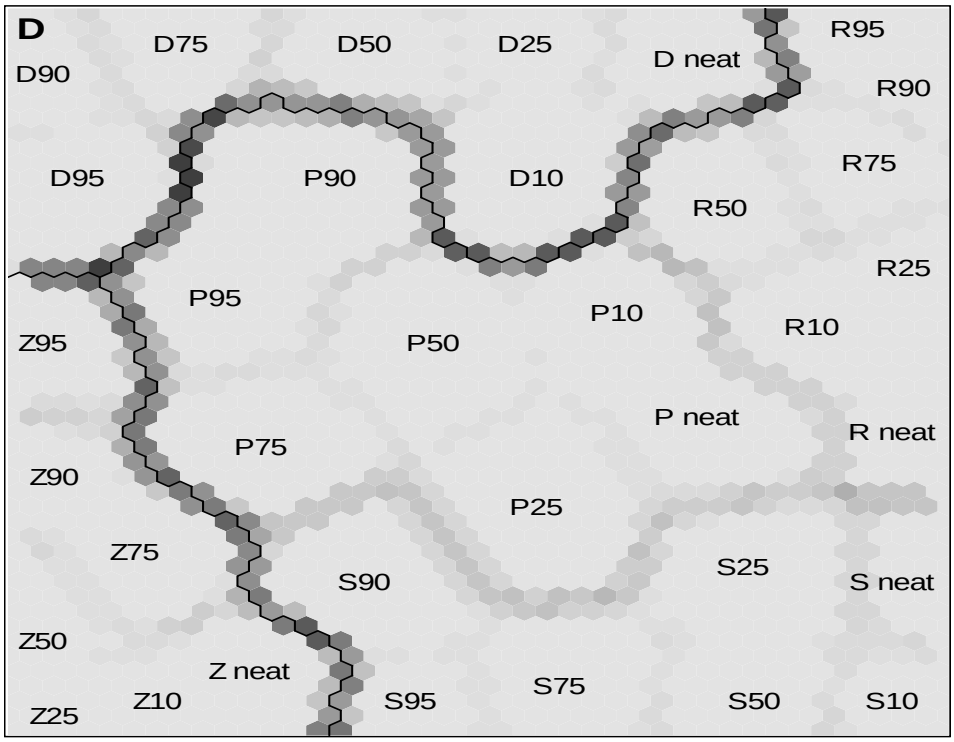


Figure 4.8 U-matrix displays of normalised fourth root dataset

As a result of using the U-Matrix visualisation method, groupings within and between sample brands and the cluster presentation was straight forward and easy to understand. The effectiveness of SOFM over both PCA and HCA was demonstrated by its ability to correctly classify samples using the normalised and pre-processed data sets, although as with HCA the best result was observed with normalised forth root data. In addition SOFM clearly illustrates samples that have similar chromatographic profiles (such as Ronsonol and Parma) by placing these sample clusters in close proximity with each other. This was not achievable using HCA.

4.5 CONCLUSIONS

The results presented demonstrate that pre-treated chromatographic data from analysed lighter fuel samples which have been evaporated to varying degrees can be interrogated using chemometric methods and successfully linked back to their parent pure ignitable liquid sample.

Data pre-treatment was essential in order to obtain accurate classifications. Three methods of data pre-treatment were applied and the best overall discrimination within and between samples was achieved with a normalised fourth root transformation data. It was determined that successful sample classification (by brand) was only achieved using HCA and SOFM where SOFM proved to have a more robust sample linkage capacity and confirmed visual similarities and differences between the samples in evidence in the chromatograms. This has demonstrated a potential means whereby pure and evaporated ignitable liquid samples can be linked successfully by brand and as such presents a powerful new means of interpreting chromatographic data retrieved from fire debris samples.

4.6 REFERENCES

1. *American Society for Testing and Materials Methods E1618-01. Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography/Mass Spectrometry*. 2001, American Society for Testing and Materials: West Conshohocken.

2. *American Society of Testing and Materials Method E1387-01. Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography*. 2001, American Society for Testing and Materials: West Conshohocken.

3. DeHaan, J.D., *Kirk's Fire Investigation*. 6th Ed. 2007, New York: Brady.

4. Almirall, J.R. and Furton, K.G., *Analysis and Interpretation of Fire Scene Evidence*. 2004, Boca Raton: CRC Press.

5. Newman, R., Gilbert, M. and Lothridge, K., *GC-MS Guide to Ignitable Liquids*. 1998, Boca Raton: CRC Press.

6. Vella, A.J., *Arson Investigation Using the Ion Trap Detector.* Journal of Forensic Science Society, 1992. **32**: p. 131 - 142.

7. Nowicki, J., *An Accelerant Classification Scheme Based on Analysis by Gas Chromatography/Mass Spectrometry (GC-MS).* Journal of Forensic Science, 1990. **35**: p. 1064 - 1086.

8. Nowicki, J., *Analysis of Fire Debris Sample by Gas Chromatography/Mass Spectrometry (GC-MS): Case Studies.* Journal of Forensic Science, 1991. **36**: p. 1536 - 1550.

9. Lennard, C.J., Rochaix, V.T. and Margot, P., *A GC-MS Dataset of Target Compound Chromatograms for the Identification of Arson Accelerants.* Science and Justice, 1995. **35**: p. 19 - 30.

10. Keto, R. and Wineman, P.L., *Detection of Petroleum-Based Accelerants in Fire Debris by Target Compound Gas Chromatography/Mass Spectrometry.* Analytical Chemistry, 1991. **63**: p. 1964 - 1971.

11. Wang, Z. and Fingas, M., *Differentiation of the Source of Spilled Oil and Monitoring of the Oil Weathering Process using GC-MS.* Journal of Chromatography A, 1995. **712**: p. 321 - 343.

12. Bertsch, W., *Volatiles from Carpet: A Source of Frequent Misinterpretation in Arson Analysis.* Journal of Chromatography A, 1994. **674**: p. 329 - 333.

13. DeHaan, J.D. and Bonarius, K., *Pyrolysis Products of Structure Fires.* Journal of Forensic Science Society, 1988. **28**: p. 299 - 309.

14. Lentini, J.J., *Scientific Protocols for Fire Investigation*. 2006, Boca Raton: CRC Press.

15. Lentini, J.J., *Differentiation of Asphalt and Smoke Condensates from Liquid Petroleum Distillates using GC/MS.* Journal of Forensic Science, 1998. **43**(1): p. 97 - 113.

16. Berrueta, L.A., *Supervised Pattern Recognition in Food Analysis.* Journal of Chromatography A, 2007. **1158**: p. 196 - 214.

17. Everitt, B.S. and Dunn, G., *Applied Multivariate Data Analysis*. 2001, London: Arnold.

18. Tan, B., Hardy, J.K. and Snavely, R.E., *Accelerant Classification by Gas Chromatography/Mass Spectrometry and Multivariate Pattern Recognition.* Analytica Chimica Acta, 2000. **422**: p. 37 - 46.

19. Bodle, E.S. and Hardy, J.K., *Multivariate Pattern Recognition of Petroleum Based Accelerants by Solid Phase Microextraction Gas Chromatography with Flame Ionisation Detector.* Analytica Chimica Acta, 2007. **59**: p. 247 - 254.

20. Doble, P., Sandercock, M., Pasquier, E.D., Petocz, P., Roux, C. and Dawson, M., *Classification of Premium and Regular Gasoline by Gas Chromatography/Mass Spectrometry, Principal Component Analysis and Artificial Neural Networks.* Forensic Science International, 2003. **132**: p. 235 - 240.

21. Lu, Y. and Harrington, P.B., *Forensic Application of Gas Chromatography-Differential Mobility Spectrometry with Two-Way Classification of Ignitable Liquids from Fire Debris.* Analytical Chemistry, 2007. **79**: p. 6752 - 6759.

22. Petraco, N.D., *Statistical Discrimination of Liquid Gasoline Samples from Casework.* Journal of Forensic Science, 2008. **53**: p. 1092 - 1101.

23. Kohonen, T., *Self Organising Maps*. 1991, Berlin: Springer-Verlag.

24. Brereton, R.G., *Chemometrics for Pattern Recognition*. 2009, West Sussex: John Wiley and Sons Ltd.

25. Borusiewicz, R., Zadora, G. and Paulus, J.Z., *Application of Head-Space Analysis with Passive Adsorption for Forensic Purposes in the Automated Thermal Desorption-Gas Chromatography-Mass Spectrometry System.* Chromatographia (Supplement), 2004. **60**: p. S133 - S142.

26. Sandercock, P.M.L. and Pasquier, E.D., *Chemical Fingerprinting of Unevaporated Automative Gasoline Samples.* Forensic Science International, 2003. **134**: p. 1 - 10.

27. Sandercock, P.M.L. and Pasquier, E.D., *Chemical Fingerprinting of Gasoline.* Forensic Science International, 2004. **140**: p. 43 - 59.

28. Wold, S., Sjostrom, M. and Erikksson, L., *PLS-Regression: A Basic Tool of Chemometrics.* Chemometrics and Intelligence Laboratory Systems, 2001. **58**: p. 109 - 130.

29. Rietjens, M., *Reduction of Error Propagation due to Normalisation: Effect of Error Propagation and Closure on Spurious Correlations.* Analytica Chimica Acta, 1995. **316**: p. 205 - 215.

30. Andersson, K., *Development of Harmonised Method for the Profiling of Amphetamines VI: Evaluation of Methods for Comparison of Amphetamine.* Forensic Science International, 2007. **169**: p. 86 - 99.

31. Stauffer, E., Dolan, J.A. and Newman, R., *Fire Debris Analysis*. 2008, Burlington, U.S.A: Academic Press.

32. Stauffer, E., *Concept of Pyrolysis for Fire Debris Analysis.* Science and Justice, 2003. **43**: p. 29 - 40.

33. Locke, A.K., Basara, G.J. and Sandercock, P.M.L., *Evaluation of Internal Standards for the Analysis of Ignitable Liquids in Fire Debris.* Journal of Forensic Science, 2009. **54**(2): p. 320 - 327.

34. *NIST/EPA/NIH Mass Spectral Library (NIST 08) and NIST Mass Spectral Search Program (Version 2.0f)*, in *Standard Reference Data Program*. 2008, U.S. Department of Commerce, National Institute of Standards and Technology (NIST): Gaithersburg.

35. Newman, R., in *Fire Investigation*, N.Nic Daeid, Editor. 2004, CRC Press: Boca Raton. p. 137 - 151.

CHAPTER 5: SUPERVISED PATTERN RECOGNITION TECHNIQUES APPLIED TO DATA DERIVED FROM THE ANALYSIS OF WAX BASED PRODUCTS

5.1 INTRODUCTION

This chapter describes the application of the multi-layer perceptron (MLP) neural network technique for the interpretation of the data derived from the analysis of the wax based products. MLP is a supervised pattern recognition neural network technique and as such differs from the mathematical methods outlined in the previous chapters as it is capable of being trained to recognise specific characteristics within the data matrix using different algorithms. This means that MLP has the potential to be used in a predictive way.

MLP can be trained using three different kinds of training algorithms. These are known as 'conventional' back-propagation (BP), conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithms. Each training mechanism provides a varying size of hidden neurons in the hidden layer. The performances of each MLP trained using each method will be compared for their applicability to the wax based product data.

5.2 Multi Layer Perceptron (MLP) Neural Network

Multi Layer Perceptron (MLP) neural network acts as 'universal function approximator' [1]. This means that it is capable of finding function(s) that are very close to the actual function(s) and that reflects or models a given dataset where the model describes the distribution of samples (of the given dataset) in its sample space.

According to Longman Dictionary of Contemporary English [2], a model can be defined as a small representation or copy of something, in other words, a small object, usually built to scale, that represent in detail another, often larger object. A mathematical representation of a model is given by,

$$Y = f(x)$$

An example of a mathematical model is,

$$Y = mX + C$$

A curve fitting equation is commonly used to model a distribution of objects or samples (by fitting a straight line) in two-dimensional object or sample space as illustrated in Figure 5.1 where *Y* is the value at Y-axis, *m* is the slope, *X* is the value at X-axis and *C* is the intercept.

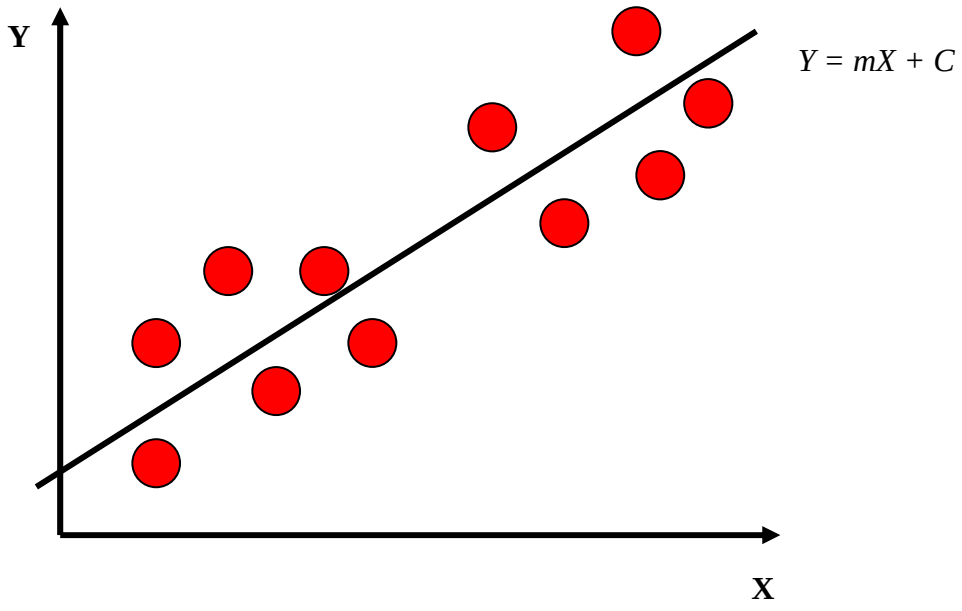


Figure 5.1 Distribution of objects or samples in two-dimensional sample space modelled by its corresponding *Y = mX + C* model.

The $Y = mX + C$ that models the distribution of objects or samples in the two-dimensional space as in Figure 5.1 has predictive ability. If given that, $X = 2$, $m = 2$ and $C = 2$, then Y (an unknown object or sample) can be predicted or determined by computing;

$$Y = mX + C$$
$$Y = 2(2) + 2$$
$$Y = 6$$

If classification rules (or boundaries) are further established (for the model in Figure 5.1) for example;

If $Y > 5$, then Y belongs to Group A,

If $Y < 5$, then Y belongs to Group B,

Therefore, by applying the classification rules and for $Y = 6$, the unknown belongs to Group B as illustrated in Figure 5.2.

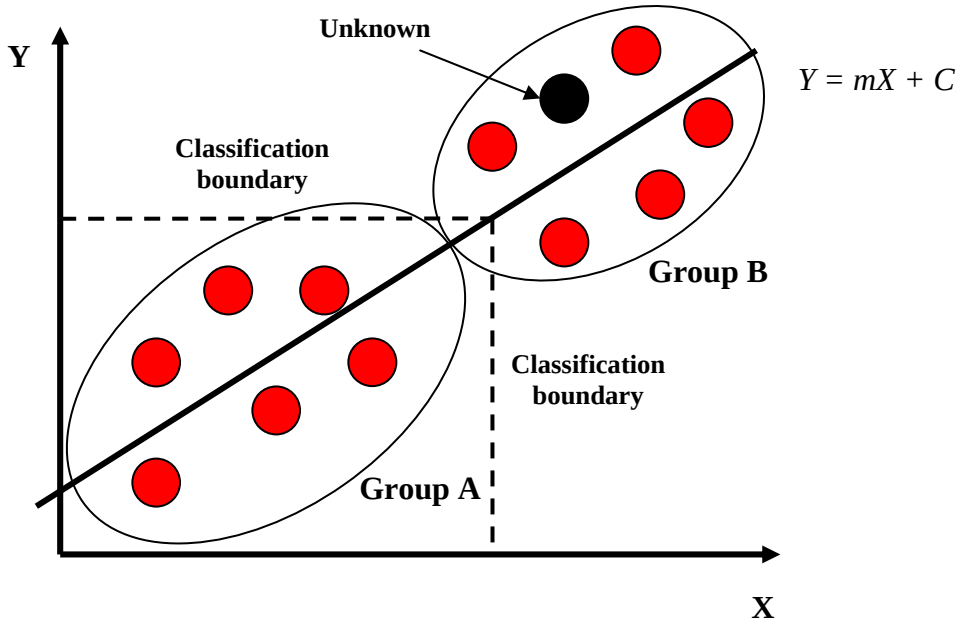


Figure 5.2 Classification of an unknown using the $Y = mX + C$ model and its classification rules

5.3 Limitations with Multi-Dimensional Datasets

For multidimensional datasets, estimating models that describe the dataset is quite a difficult task. Although the dataset can be broken down according to the number of variables, for example ten parts if there are ten variables (as illustrated in Figure 5.3), estimating each model (i.e. $Y = f(X)$) for each part, can be a complicated process.

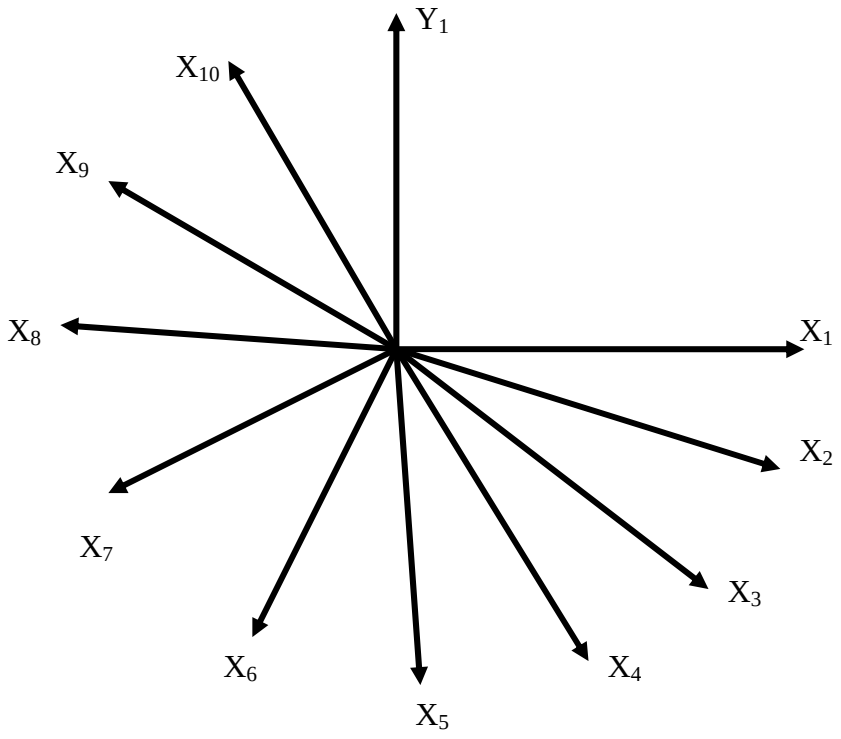


Figure 5.3 Representation of multidimensional dataset with ten variables projected at ten axes

A solution to this problem is to employ a multilayer perceptron (MLP) neural network which can estimate the 'best' function or model and classification rules or boundaries that describes the multidimensional dataset. Once found, the model can be employed as a predictive tool to interpret class membership of unknown(s). Prediction itself is a useful feature that has increasing significance in applications in forensic science [3].

5.4 EXPERIMENTAL

5.4.1 Dataset Preparation

Two sets of MLP neural networks were trained. One set was trained using the TLC combined with the MSP data for the wax based products (Dataset 1) while the other set was trained using UV/Vis combined with the GC data for the wax based products which was pre-treated using the square root method (Dataset 2). These datasets were chosen because discrimination was achieved with some of the unsupervised chemometrics and neural network techniques discussed in Chapter 3. Prior to data processing, each dataset was divided into a training, validation and test set in a 2: 1: 1 ratio of the samples available. The datasets were first prepared separately in the Microsoft® Excel spreadsheet and then imported to MATLAB® Version 7.6.0.324 R2008a (Math Works Inc.) where training of the MLP neural network was conducted using the Neural Network Toolbox (Version 6.0).

5.4.2 Input and Output Pairs of Dataset 1

Dataset 1 consisted of ninety nine (99) individual exemplars from the thirty three (33) samples (lipsticks and shoe polishes that produced any coloured spots) analysed in triplicate (33 x 3 = 99). Of these, eighty (80) exemplars (57 lipsticks exemplars and 23 shoe polishes exemplars) were used in the training. Nineteen (19) exemplars consisting of nine (9) lipsticks and ten (10) shoe polishes were kept aside and treated as unknowns. In addition to these samples, twelve (12) exemplars from three samples i.e. L7, L21 and SP2 in the weathering study (previously described in section 3.6.2) were also included.

The input vectors used were the measurements taken from TLC and MSP analysis of the samples. The data inputted were the retardation factor and the percentage reflectance of the coloured spot. Each of the output values for each of the sample type were coded as 1 (lipstick) and 0 (shoe polish) respectively.

5.4.3 Input and Output Pairs of Dataset 2

Dataset 2 consisted of one hundred and eighty nine (189) individual exemplars from 63 samples (from all lipsticks, lip balms and shoe polishes) analysed in triplicate (63 x 3 = 189). Of these, one hundred and fifty nine (159) exemplars (63 lipsticks exemplars, 54 lip balms exemplars and 42 shoe polishes exemplars) were used in the training. Thirty (30) exemplars consisting of twelve lipsticks, nine lip balms and nine shoe polishes were kept aside and treated as the unknowns set. In addition, twenty four (24) exemplars from six samples i.e. L7, L21, LB19, LB16, SP2 and SP17 in the weathering study (previously described in section 3.6.2) were also included in the unknowns set..

In this case the input vectors used were the measurements taken from the UV/Vis and GC analysis which were the absorbance and square root of the peak area recorded for each sample. Each of the output values for each sample type (lipstick, shoe polish or lip balm) were coded as 1 (lipstick), 2 (lip balm) and 3 (shoe polish) respectively.

5.4.4 Network Arrangement

In actual the size of an MLP neural network is not restricted to three layers (one input layer, one hidden layer and one output layer) and can grow by increasing the size of the hidden layer. Svozil [1] strongly suggested that a network with one hidden layer is used because with an increase in the size of the hidden layer, the training process becomes slow [1] and in most cases such a network is sufficient to learn the correlated patterns between a given input pattern and the associated output pattern [4]. Three layer networks were used in this study. The output values were paired with their corresponding input measurements for data processing. As such the MLP was ‘asked’ during the training process to achieve the target of 1 and 0 in Dataset 1 for lipstick and shoe polish respectively and 1, 2 and 3 for lipstick; lip balm and shoe polish respectively in Dataset 2. This is as illustrated in Figure 5.4.

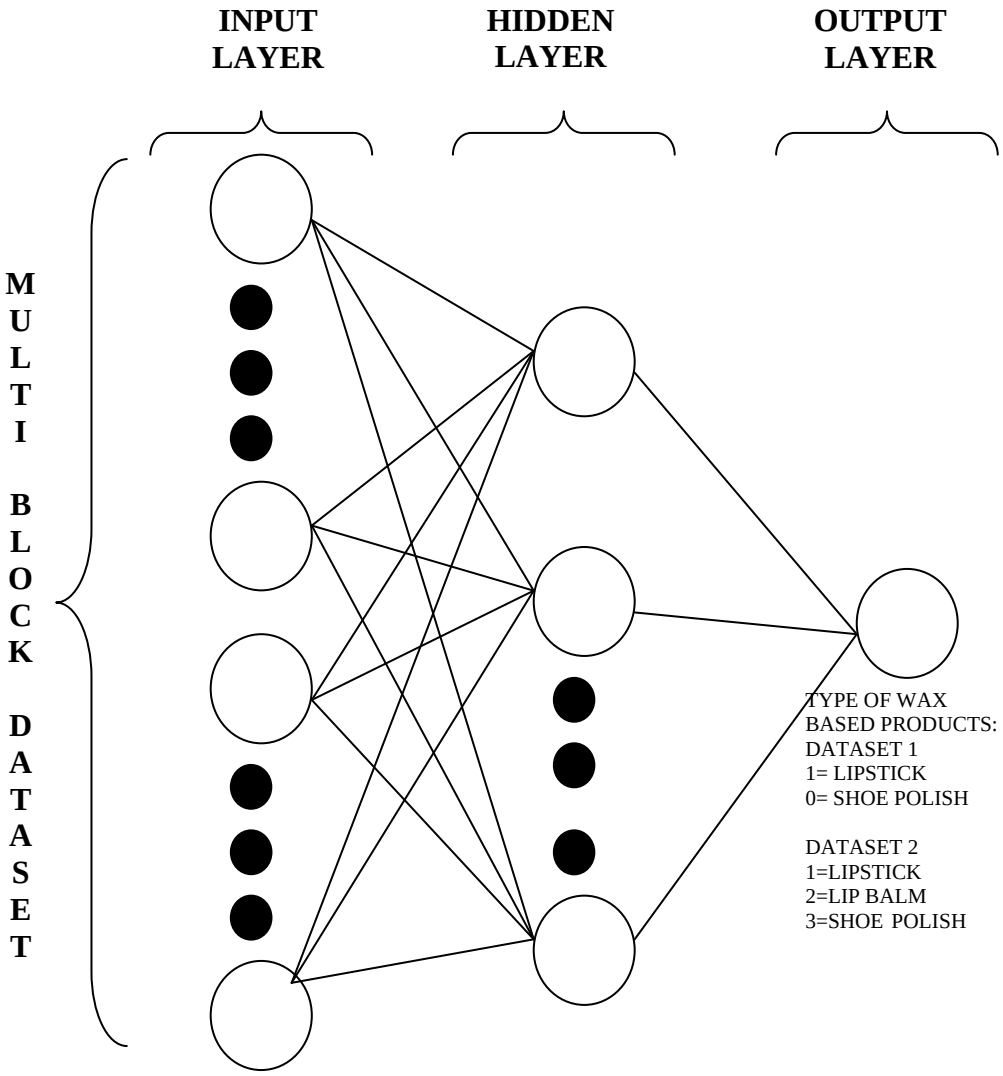


Figure 5.4 Layout of the MLP neural network

5.4.5 Learning Algorithms, Network Arrangement and Data Division

The most commonly used learning algorithm for MLP is back-propagation (BP) which was the first learning algorithm developed for multi-layer networks [5]. The algorithm determines the minimum error by examining changes in the error surface as a result of small alterations in the algorithm. This process can be slow which means that training of the MLP can be slow using this method.

To overcome this problem, other learning algorithms based on back propagation have been developed including the conjugate gradient (CG) and Lavenberg-Marquardt (LM) methods [6]. Both of these algorithms also determine the minimum error surface but in a manner which is faster than conventional back propagation.

For each learning algorithm, ten MLP neural networks with varying sizes of hidden neurons (from 2 – 11) were constructed and each was trained three times. This produced ninety trained MLP neural networks. Prior to training, the dataset was divided into a training, validation and test set in a 2: 1: 1 proportion as previously described. There is no specific rule in determining the best number of hidden neurons within the hidden layer for a given dataset and the best architecture is determined by varying the number of neurons sequentially [7]. The number of hidden neurons is important as a larger number may result in over-fitting, while a smaller number may not capture the information adequately [8].

Each MLP neural network began with a different random selection of weights associated with the connection between the neurons. At the end of each training process, each network achieved a different set of training errors which means that no MLP had exactly the same errors as any other. Training errors may also reflect on the ability of MLP neural network to generalise data and as such indicates its predictive quality.

5.4.6 MLP Neural Network Training

After the training of each MLP was completed, a classification matrix, a training performance plot (examples of which are presented in Figures 5.5 to 5.7) and a regression plot for the training, validation and test set were constructed. This information was used to evaluate the performance of each MLP and to select the optimum MLP neural network.

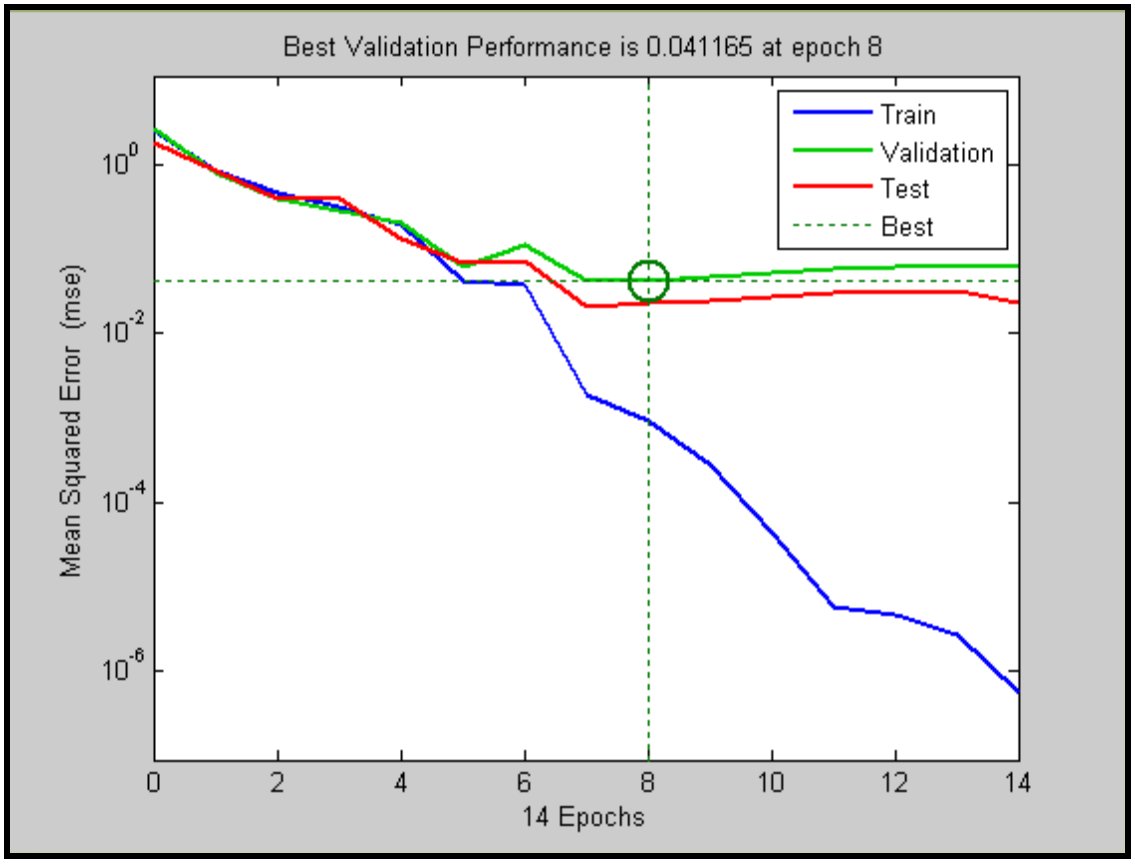


Figure 5.5 A typical performance plot

The training performance plot illustrates the performance (measured by the mean square error (MSE)) for the training, validation and test set respectively. In the example illustrated in Figure 5.5, the training was stopped after fourteen iterations (epochs) and the best validation performance was returned at epoch 8. From epoch 8 to 14, the

training error continued to decrease while the validation error began to increase slightly indicating that over fitting was occurring.

Regression plots for the training, validation, test set and for the entire dataset were also produced and are illustrated in Figures 5.6 to 5.7. The main use of these plots was to visualise the correlation between the input and the output pairs, in other words how well the outputs track the input. Ideally, the regression coefficient (R) that calculates the slope should be as close to 1 (unity) as possible demonstrating excellent correlation between each input and output pairs.

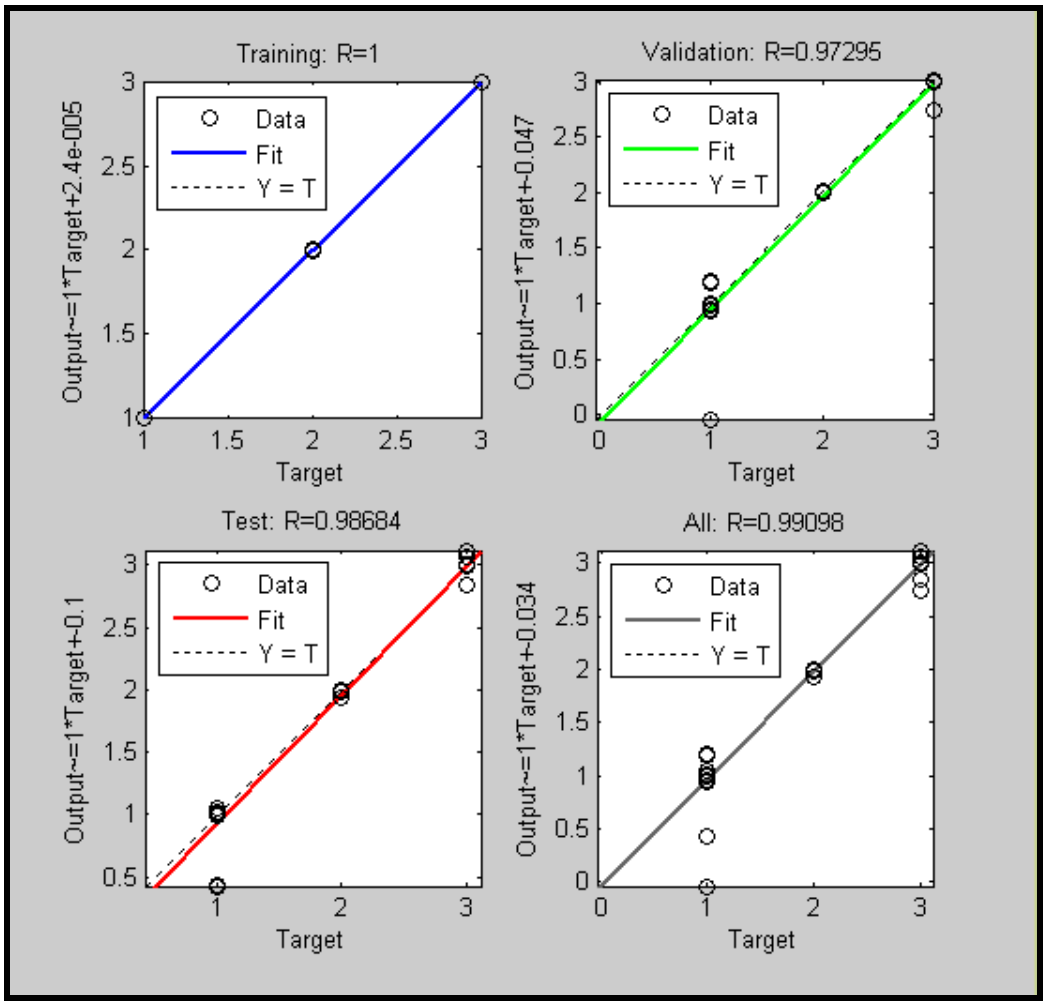


Figure 5.6 Example of regression plots for the training, validation and test set in a correctly training network.

All of the regression plots in Figure 5.6 have correlation coefficients very close to unity (R = 1 for the training set, R = 0.97295 for the validation set, R = 0.98684 for the test set and R = 0.99098 for the entire dataset). This means that, the outputs have tracked the input variables very well.

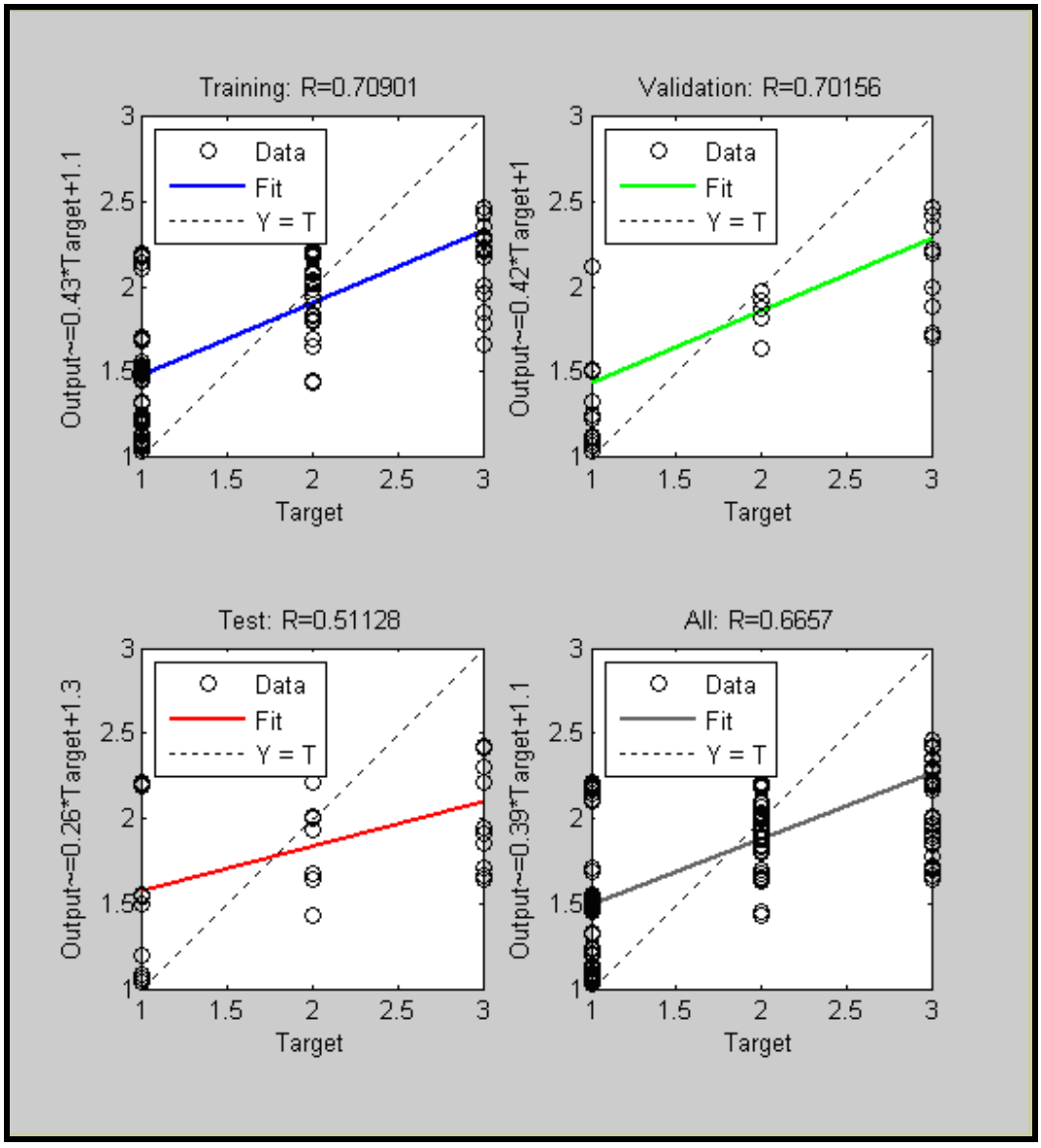


Figure 5.7 Regression plots for an under trained network.

Figure 5.7 depicts an example of an under trained MLP neural network. The correlation coefficients are all well below unity i.e. R = 0.70901 for the training set, R = 0.70156 for the validation set, R = 0.51128 for the test set and R = 0.6657 for the entire dataset. This means that the outputs have failed to track the input variables appropriately.

The final classification and misclassification matrix for each MLP is normally presented in the form of a table showing the percentage of correctly and wrongly classified samples.

5.5 RESULTS AND DISCUSSION

5.5.1 Dataset Preparation

The main target of the training process for an MLP is to 'achieve' networks that are capable of generalising well. There are a number of factors that can influence this, one of which is the number of samples used within the training set. When small sample sets are used to train the network it is possible that the network can become over trained rendering it incapable of generalising well [9]. To prevent this, replicates of each sample under study were used as the input vectors following the approach suggested by Goodacre *et. al* [10]. The performance plot, regression plot and classification matrix were used to select the optimum MLP neural networks that best described both datasets.

5.5.2 MLP Neural Networks Trained using Dataset 1

The best performances for the training, validation and test set for each training algorithm of the MLP neural networks trained using Dataset 1 are presented in Tables 5.1 to 5.3. The performance of a given MLP neural network is measured in terms of the mean square error (MSE) value. This value should be small indicating that network has reached a point where the difference between the target and input vectors is small.

Table 5.1 Training, validation and test performances for the MLP neural networks trained using back-propagation (BP) algorithm with Dataset 1

NETWORK	NO OF HIDDEN NEURONS	BEST PERFORMANCE – Mean Square Error (MSE)		
		Training	Validation	Test
BP102	2	0.0478	0.0675	0.0710
BP202	2	0.0491	0.1102	0.0959
BP302	2	0.0379	0.0361	0.0281
BP103	3	0.0183	0.0403	0.0522
BP203	3	0.0277	0.0304	0.0509
BP303	3	0.0235	0.0343	0.0412
BP104	4	0.0606	0.0781	0.0616
BP204	4	0.0439	0.0301	0.1563
BP304	4	0.0484	0.0735	0.0243
BP105	5	0.0200	0.0158	0.0111
BP205	5	0.0286	0.0158	0.0341
BP305	5	0.0263	0.0272	0.0123
BP106	6	0.0247	0.0357	0.0403
BP206	6	0.0200	0.0342	0.0101
BP306	6	0.0270	0.0248	0.0398
BP107	7	0.0210	0.0725	0.0840
BP207	7	0.0248	0.0444	0.0096
BP307	7	0.0517	0.0499	0.1048
BP108	8	0.0237	0.0052	0.0370
BP208	8	0.0202	0.0618	0.1078
BP308	8	0.0222	0.0476	0.0691
BP109	9	0.0151	0.0314	0.0628
BP209	9	0.0183	0.0141	0.0407
BP309	9	0.0180	0.0222	0.0212
BP110	10	0.0151	0.0314	0.0628
BP210	10	0.0269	0.0243	0.0295
BP310	10	0.0320	0.0659	0.1239
BP111	11	0.1080	0.0752	0.1101
BP211	11	0.0224	0.0823	0.0148
BP311	11	0.0210	0.0445	0.0647

Table 5.2 Training, validation and test performances for the MLP neural networks trained using conjugate gradient (CG) algorithm with Dataset 1

NETWORK	NO OF HIDDEN NEURONS	BEST PERFORMANCE – Mean Square Error (MSE)		
		Training	Validation	Test
CG 102	2	0.0189	0.0372	0.0241
CG 202	2	0.0009	0.0318	0.0518
CG 302	2	0.0187	0.0019	0.1030
CG 103	3	0.0212	0.0468	0.0400
CG 203	3	0.0007	0.0389	0.0017
CG 303	3	0.0002	0.0011	0.0013
CG 104	4	0.0095	0.0078	0.0070
CG 204	4	0.0004	0.0003	0.0098
CG 304	4	0.0008	0.0001	0.0002
CG 105	5	0.0044	0.0089	0.0035
CG 205	5	0.0064	0.0015	0.0097
CG 305	5	0.0013	0.0013	0.0015
CG 106	6	0.0060	0.0180	0.0200
CG 206	6	0.0004	0.0023	0.0022
CG 306	6	0.0001	0.0010	0.0001
CG 107	7	0.0001	0.0019	0.0007
CG 207	7	0.0025	0.0487	0.0782
CG 307	7	0.0010	0.0146	0.0102
CG 108	8	0.0001	0.0002	0.0001
CG 208	8	0.0137	0.0493	0.0469
CG 308	8	0.0001	0.0001	0.0005
CG 109	9	0.0046	0.0173	0.0205
CG 209	9	0.0005	0.0022	0.0015
CG 309	9	0.0010	0.0049	0.0039
CG 110	10	0.0003	0.0008	0.0051
CG 210	10	0.0001	0.0001	0.0004
CG 310	10	0.0020	0.0459	0.0072
CG 111	11	0.0119	0.0408	0.0831
CG 211	11	0.2660	0.0153	0.0055
CG 311	11	0.0010	0.0025	0.0026

Table 5.3 Training, validation and test performances for MLP neural networks trained using Lavenberg-Marquardt (LM) with Dataset 1

NETWORK	NO OF HIDDEN NEURONS	BEST PERFORMANCE – Mean Square Error (MSE)		
		Train	Validate	Test
LM102	2	5.7829E-16	4.76E-11	6.85E-11
LM202	2	7.91E-08	0.0067	0.0054
LM302	2	4.29E-07	1.63E-04	3.07E-04
LM103	3	3.40E-07	0.0772	0.0866
LM203	3	4.33E-08	0.0427	0.0038
LM303	3	7.48E-08	0.0285	0.0506
LM104	4	1.36E-07	0.009	0.0148
LM204	4	1.01E-24	8.34E-05	2.31E-04
LM304	4	1.44E-12	8.91E-07	3.93E-04
LM105	5	4.98E-16	2.34E-15	1.24E-15
LM205	5	1.35E-04	0.006	0.005
LM305	5	1.13E-16	0.0058	0.0016
LM106	6	3.04E-09	5.69E-05	3.04E-04
LM206	6	2.64E-09	4.50E-07	1.43E-04
LM306	6	9.89E-06	0.0192	2.34E-04
LM107	7	9.87E-12	7.53E-09	1.66E-06
LM207	7	3.60E-09	3.84E-04	6.23E-04
LM307	7	7.13E-25	0.0033	0.0055
LM108	8	1.10E-21	0.0033	5.31E-05
LM208	8	1.87E-09	1.66E-06	1.32E-05
LM308	8	8.40E-24	4.76E-04	2.89E-04
LM109	9	2.87E-25	0.001	1.73E-04
LM209	9	2.88E-23	0.0021	0.0022
LM309	9	2.71E-08	0.0023	6.14E-05
LM110	10	2.78E-07	4.17E-06	5.32E-05
LM210	10	3.10E-05	0.0206	0.0118
LM310	10	3.05E-24	0.0015	1.46E-04
LM111	11	3.23E-12	0.0068	0.003
LM211	11	8.79E-26	4.13E-04	2.11E-04
LM311	11	3.96E-23	6.75E-04	2.04E-04

As evident from Table 5.1 to 5.3, all networks (regardless of the training algorithm used) give the best performances with the MLP neural networks trained using LM algorithm as these provided the lowest MSE values. Table 5.4 to 5.6 illustrate the regression coefficient (R) of the training, validation, and test subset and for the overall dataset for the MLP neural networks trained using the back-propagation (BP), conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithm respectively for Dataset 1.

Table 5.4 Regression coefficients (R) of the training, validation and test subset and for the overall dataset for the MLP neural networks trained using BP algorithm with Dataset 1

NETWORK	REGRESSION COEFFICIENT (R)			
	Training	Validation	Test	All
BP102	0.95111	0.92784	0.88537	0.93022
BP202	0.91668	0.95454	0.93407	0.92977
BP302	0.94738	0.92805	0.88494	0.92809
BP103	0.97201	0.95557	0.95453	0.96533
BP203	0.96847	0.94439	0.93937	0.95309
BP303	0.92481	0.93955	0.97579	0.93481
BP104	0.83414	0.76827	0.89739	0.83453
BP204	0.89888	0.84801	0.60082	0.83167
BP304	0.86944	0.82135	0.95783	0.87434
BP105	0.95376	0.95790	0.97199	0.95750
BP205	0.93359	0.97496	0.94023	0.93529
BP305	0.93478	0.93096	0.97435	0.94206
BP106	0.93230	0.94078	0.93717	0.93004
BP206	0.94867	0.92577	0.97801	0.94947
BP306	0.93017	0.94096	0.90274	0.92642
BP107	0.94244	0.81543	0.87045	0.88843
BP207	0.93527	0.91012	0.97408	0.93583
BP307	0.93356	0.84733	0.95512	0.86854
BP108	0.94112	0.98554	0.92003	0.94342
BP208	0.94765	0.85514	0.72885	0.88231
BP308	0.94310	0.86941	0.88615	0.91212
BP109	0.95301	0.94220	0.85901	0.93011
BP209	0.95489	0.97506	0.90894	0.94685
BP309	0.95712	0.94262	0.95911	0.95285
BP110	0.82576	0.81358	0.44033	0.77163
BP210	0.93350	0.94519	0.95133	0.93454
BP310	0.91932	0.84330	0.72640	0.85579
BP111	0.71877	0.83307	0.70145	0.71723
BP211	0.93451	0.81973	0.96134	0.91639
BP311	0.94538	0.85082	0.86964	0.91236

Table 5.5 Regression coefficients (R) of the training, validation and test subset and for the overall dataset for the MLP neural networks trained using CG algorithm with Dataset 1

NETWORK	REGRESSION COEFFICIENT (R)			
	Training	Validation	Test	All
CG 102	0.95518	0.84049	0.88626	0.91947
CG 202	0.99789	0.87404	0.89153	0.94746
CG 302	0.95386	0.99194	0.76566	0.91241
CG 103	0.96042	0.53717	0.77586	0.88629
CG 203	0.99842	0.67174	0.95928	0.90260
CG 303	0.99919	0.99575	0.99569	0.99716
CG 104	0.97440	0.96869	0.89536	0.95610
CG 204	0.99918	0.99834	0.94833	0.98471
CG 304	0.99978	0.99968	0.99954	0.99969
CG 105	0.98990	0.97504	0.97092	0.98349
CG 205	0.98622	0.99775	0.92118	0.97322
CG 305	0.99694	0.99474	0.99737	0.99613
CG 106	0.98409	0.96055	0.96125	0.97307
CG 206	0.99908	0.99049	0.99313	0.99565
CG 306	0.99969	0.99749	0.99982	0.99920
CG 107	0.99974	0.99661	0.99931	0.99863
CG 207	0.99435	0.81371	0.79915	0.92779
CG 307	0.99728	0.97197	0.97312	0.98434
CG 108	0.99970	0.99912	0.99973	0.99961
CG 208	0.96438	0.89520	0.87758	0.92340
CG 308	0.99971	0.99947	0.99814	0.99936
CG 109	0.98837	0.95705	0.94124	0.97135
CG 209	0.99872	0.99547	0.99635	0.99731
CG 309	0.99726	0.99118	0.99364	0.99380
CG 110	0.99930	0.99792	0.98242	0.99554
CG 210	0.99962	0.99969	0.99912	0.99946
CG 310	0.99515	0.87795	0.94049	0.96097
CG 111	0.97519	0.82590	0.65044	0.90485
CG 211	0.96802	0.94985	0.89363	0.95273
CG 311	0.99744	0.99523	0.98245	0.99526

Table 5.6 Regression coefficients (R) of the training, validation and test subset and for the overall dataset for the MLP neural networks trained using LM algorithm with Dataset 1

NETWORK	REGRESSION COEFFICIENT (R)			
	Training	Validation	Test	All
LM102	1.00000	1.00000	1.00000	1.00000
LM202	1.00000	0.94564	0.98732	0.98536
LM302	1.00000	0.99854	0.99946	0.99957
LM103	1.00000	0.53756	0.61719	0.88341
LM203	1.00000	0.67182	0.98127	0.91825
LM303	1.00000	0.87082	0.88603	0.94630
LM104	1.00000	0.95978	0.97891	0.98588
LM204	1.00000	0.99984	0.99941	0.99985
LM304	1.00000	1.00000	0.99893	0.99974
LM105	1.00000	1.00000	1.00000	1.00000
LM205	0.99988	0.98145	0.98980	0.99368
LM305	1.00000	0.99040	0.99676	0.99669
LM106	1.00000	0.99974	0.99938	0.99978
LM206	1.00000	1.00000	0.99936	0.99993
LM306	0.99998	0.66854	0.99510	0.90943
LM107	1.00000	0.99999	1.00000	1.00000
LM207	1.00000	0.99691	0.99868	0.99950
LM307	1.00000	0.99309	0.99298	0.99648
LM108	1.00000	0.99370	0.99939	0.99814
LM208	1.00000	1.00000	0.99991	0.99998
LM308	1.00000	0.99913	0.99932	0.99962
LM109	1.00000	0.99798	0.99965	0.99941
LM209	1.00000	0.99415	0.99418	0.99795
LM309	1.00000	0.94747	0.99568	0.98596
LM110	1.00000	0.99978	0.99988	0.99991
LM210	0.99993	0.93310	0.95266	0.97848
LM310	1.00000	0.99795	0.99947	0.99918
LM111	1.00000	0.98592	0.98383	0.98560
LM211	1.00000	0.99893	0.99964	0.99970
LM311	1.00000	0.99891	0.99872	0.99944

Again with few exceptions, the best regression coefficients (closest to 1) were achieved using the Lavenberg-Marquardt algorithm. The classification matrix for each MLP neural networks trained using the back-propagation (BP), conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithm are presented in Table 5.7, 5.8 and 5.9 respectively. Networks which have misclassified the data are shown in red.

Table 5.7 Classification matrix for MLP neural networks trained using back-propagation (BP) algorithm with Dataset 1 which consisted of 57 lipstick and 23 shoe polishes exemplars.

NETWORK	SAMPLE	SAMPLE		SAMPLE (%)	
		L	SP	L	SP
BP102	L	54	3	67.50	3.75
	SP	3	20	3.75	25.00
BP202	L	54	3	67.50	3.75
	SP	4	19	5.00	23.75
BP302	L	57	0	71.25	0.00
	SP	3	20	3.75	25.00
BP103	L	57	0	71.25	0.00
	SP	2	21	2.50	26.25
BP203	**L**	**57**	**0**	**71.25**	**0.00**
	SP	**0**	**23**	**0.00**	**28.75**
BP303	**L**	**57**	**0**	**71.25**	**0.00**
	SP	**0**	**23**	**0.00**	**28.75**
BP104	L	54	3	67.50	3.75
	SP	2	21	2.50	26.25
BP204	L	56	1	70.00	1.25
	SP	4	19	5.00	23.75
BP304	L	54	3	70.13	3.90
	SP	0	23	0.00	29.87
BP105	**L**	**57**	**0**	**71.25**	**0.00**
	SP	**0**	**23**	**0.00**	**28.75**
BP205	**L**	**57**	**0**	**71.25**	**0.00**
	SP	**0**	**23**	**0.00**	**28.75**
BP305	**L**	**57**	**0**	**71.25**	**0.00**
	SP	**0**	**23**	**0.00**	**28.75**
BP106	L	54	3	67.50	3.75
	SP	12	11	15.00	13.75
BP206	**L**	**57**	**0**	**71.25**	**0.00**
	SP	**0**	**23**	**0.00**	**28.75**
BP306	L	56	1	70.00	1.25
	SP	3	20	3.75	25.00
BP107	L	57	0	70.37	0.00
	SP	6	18	7.41	22.22
BP207	L	57	0	71.25	0.00
	SP	2	21	2.50	26.25
BP307	L	55	2	68.75	2.50
	SP	3	20	3.75	25.00
BP108	**L**	**57**	**0**	**71.25**	**0.00**
	SP	**0**	**23**	**0.00**	**28.75**
BP208	L	54	3	67.50	3.75
	SP	2	21	2.50	26.25

Table 5.7 Continued

NETWORK	SAMPLE	SAMPLE		SAMPLE (%)	
		L	SP	L	SP
BP308	L	57	0	71.25	0.00
	SP	5	18	6.25	22.50
BP109	L	54	3	67.50	3.75
	SP	2	21	2.50	26.25
BP209	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
BP309	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
BP110	L	54	3	67.50	3.75
	SP	9	14	11.25	17.50
BP210	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
BP310	L	57	0	71.25	0.00
	SP	3	20	3.75	25.00
BP111	L	45	15	54.22	18.07
	SP	0	23	0.00	27.71
BP211	L	57	0	71.25	0.00
	SP	2	21	2.50	26.25
BP311	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75

Table 5.8 Classifications matrix for MLP neural networks trained using conjugate gradient (CG) algorithm with Dataset 1 which consisted of 57 lipstick and 23 shoe polishes exemplars.

NETWORK	SAMPLE	SAMPLE		SAMPLE (%)	
		L	SP	L	SP
CG102	L	55	2	68.75	2.50
	SP	3	20	3.75	25.00
CG202	L	54	3	67.50	3.75
	SP	0	23	0.00	28.75
CG302	L	54	3	67.50	3.75
	SP	0	23	0.00	28.75
CG103	L	52	5	65.00	6.25
	SP	3	20	3.75	25.00
CG203	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG303	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG104	L	54	3	67.50	3.75
	SP	0	23	0.00	28.75
CG204	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG304	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75

Table 5.8 Continued

NETWORK	SAMPLE	SAMPLE		SAMPLE (%)	
		L	SP	L	SP
CG105	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG205	L	55	2	68.75	2.50
	SP	0	23	0.00	28.75
CG305	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG106	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG206	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG306	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 107	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 207	L	54	3	67.50	3.75
	SP	0	23	0.00	28.75
CG 307	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 108	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 208	L	55	2	68.75	2.50
	SP	0	23	0.00	28.75
CG 308	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 109	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 209	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 309	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 110	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 210	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 310	L	57	0	71.25	0.00
	SP	3	20	3.75	25.00
CG 111	L	52	5	65.00	6.25
	SP	0	23	0.00	28.75
CG 211	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
CG 311	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75

Table 5.9 Classification matrix for MLP neural networks trained using Lavenberg-Marquardt (LM) with Dataset 1 consisting of 57 lipstick and 23 shoe polishes exemplars.

NETWORK	SAMPLE	SAMPLE		SAMPLE (%)	
		L	SP	L	SP
LM102	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM202	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM302	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM103	L	55	2	68.75	2.50
	SP	5	18	6.25	22.50
LM203	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM303	L	54	3	67.50	3.75
	SP	0	23	0.00	28.75
LM104	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM204	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM304	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM105	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM205	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM305	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM106	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM206	L	55	2	68.75	2.50
	SP	3	20	3.75	25.00
LM306	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM107	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM207	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM307	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM108	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM208	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75

Table 5.9 Continued

NETWORK	SAMPLE	SAMPLE		SAMPLE (%)	
		L	SP	L	SP
LM308	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM109	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM209	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM309	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM110	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM210	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM310	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM111	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM211	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75
LM311	L	57	0	71.25	0.00
	SP	0	23	0.00	28.75

The number of MLP neural networks that misclassified some of the exemplars (highlighted in red in the tables) were higher for networks trained using back-propagation (BP) algorithm (17 networks misclassified the samples) compared to the conjugate gradient (CG) (2 networks misclassified the samples) and Lavenberg-Marquardt (LM) algorithms (2 networks misclassified the samples). The greatest degree of misclassification also occurred with MLP neural networks trained using back-propagation (BP) algorithm, BP106 and BP111 each with fifteen misclassifications.

5.5.2.1 Selection of Optimum MLP Neural Networks Trained Using Dataset 1

All MLP neural networks with small validation and test performance errors (in Table 5.1 to 5.3) and regression coefficients close to 1 (in Table 5.4 to 5.6) as well as the networks with four or less misclassifications, representing 5% of misclassification across the data matrices (in Table 5.7 to 5.9), were used to classify the exemplars which were treated as the unknowns or test samples.

From the total of ninety MLP neural networks trained using back-propagation (BP), conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithm, nineteen MLP neural networks fit the set criteria (three trained using back-propagation (BP) and eight each trained using conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithms. The profiles of the nineteen networks are presented in Table 5.10.

Table 5.10 Profiles of the twenty three MLP neural networks for the classification of unknown exemplars

NETWORK	PERFORMANCE ERROR		TEST REGRESSION (R)	MISCLASSIFICATION (%)
	Validation	Test		
BP303	0.0343	0.0412	0.97579	0.00
BP105	0.0158	0.0111	0.97199	0.00
BP305	0.0273	0.0123	0.97435	0.00
CG206	0.0023	0.0022	0.99313	0.00
CG108	0.0002	0.0003	0.99973	0.00
CG209	0.0022	0.0015	0.99635	0.00
CG303	0.0011	0.0013	0.99569	0.00
CG304	0.0001	0.0002	0.99954	0.00
CG305	0.0013	0.0015	0.99737	0.00
CG309	0.0049	0.0039	0.99364	0.00
CG311	0.0025	0.0026	0.98245	0.00
LM102	4.76E-11	6.85E-11	1.00000	0.00
LM302	1.63E-4	3.07E-4	0.99946	0.00
LM105	2.34E-5	1.24E-15	1.00000	0.00
LM205	0.006	0.005	0.98980	0.00
LM308	4.76E-4	2.89E-4	0.99932	0.00
LM209	0.0021	0.0022	0.99418	0.00
LM211	4.13E-4	2.11E-4	0.99964	0.00
LM311	6.75E-4	2.04E-4	0.99872	0.00

5.5.2.2 Test Study MLP Dataset 1

Correctly trained MLP neural networks can give reasonable discrimination of samples into the correct outputs when presented with new input vectors (i.e. new analytical data). This ability can be exploited to present a discrimination of unknown wax based products by product type where no reference materials are available.

Each of the designated test sets previously discussed for each dataset were presented to the appropriate successful MLP neural networks as shown in Tables 5.10 and 5.20. The outcomes of the prediction by each the MLP neural networks trained with Dataset 1 are presented in Table 5.21 to 5.23 with misclassifications given in red.

Table 5.21 Prediction outcomes for the MLP neural network trained using back-propagation (BP) algorithm. 1 = lipstick, 0 = shoe polish. Black characters indicate a correct prediction

Actual	BP303	BP105	BP305
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
0	1	1	0
0	1	1	0
0	1	1	0
0	1	1	1
0	1	1	1
0	1	1	1
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	0	0
1	1	1	1
1	1	1	1
1	1	1	1
0	1	1	0
0	1	1	0
0	1	1	0
0	1	1	0

Table 5.22 Prediction outcomes for the MLP neural network trained using conjugate gradient (CG) algorithm. 1 = lipstick, 0 = shoe polish. Black characters indicate a correct prediction

Actual	CG206	CG108	CG209	CG303	CG104	CG305	CG309	CG311
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	0	1	1	1
1	1	1	1	1	0	1	1	1
1	1	1	1	1	0	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	1	1	0	1	1	0	1	1
0	1	1	1	1	1	0	1	1
0	1	1	1	1	1	0	1	1
0	0	0	1	1	1	0	1	0
0	0	1	1	1	1	0	1	0
0	0	1	1	1	1	0	1	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	0	0	1	0	0	0	0	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	0	1
0	1	1	1	1	1	1	0	1
0	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	0	1

Table 5.23 Prediction outcomes for the MLP neural network trained using Lavenberg-Marquardt (LM) algorithm. 1 = lipstick, 0 = shoe polish. Black characters indicate a correct prediction

Actual	LM202	LM302	LM105	LM205	LM308	LM209	LM211	LM311
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	1	1	1	0	0	1	1	1
0	1	1	1	0	0	1	1	1
0	1	1	1	0	0	1	1	1
0	1	1	0	1	0	1	1	0
0	1	1	0	1	0	0	1	0
0	1	1	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	0	0	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	1	1	0	1	0	1	1	1
0	1	1	0	0	0	1	1	1
0	1	1	1	1	0	1	1	0
0	1	1	0	0	0	0	1	0

Of the twenty three MLP neural networks retained, only one, LM308 (in green), correctly predicted the entire test set. This network, consisting of eight hidden neurons in its hidden layer produced good performances in terms of both the validation and test error as presented in Figure 5.8. The regression coefficient (R) for the test set was also excellent and is presented in Figure 5.9.

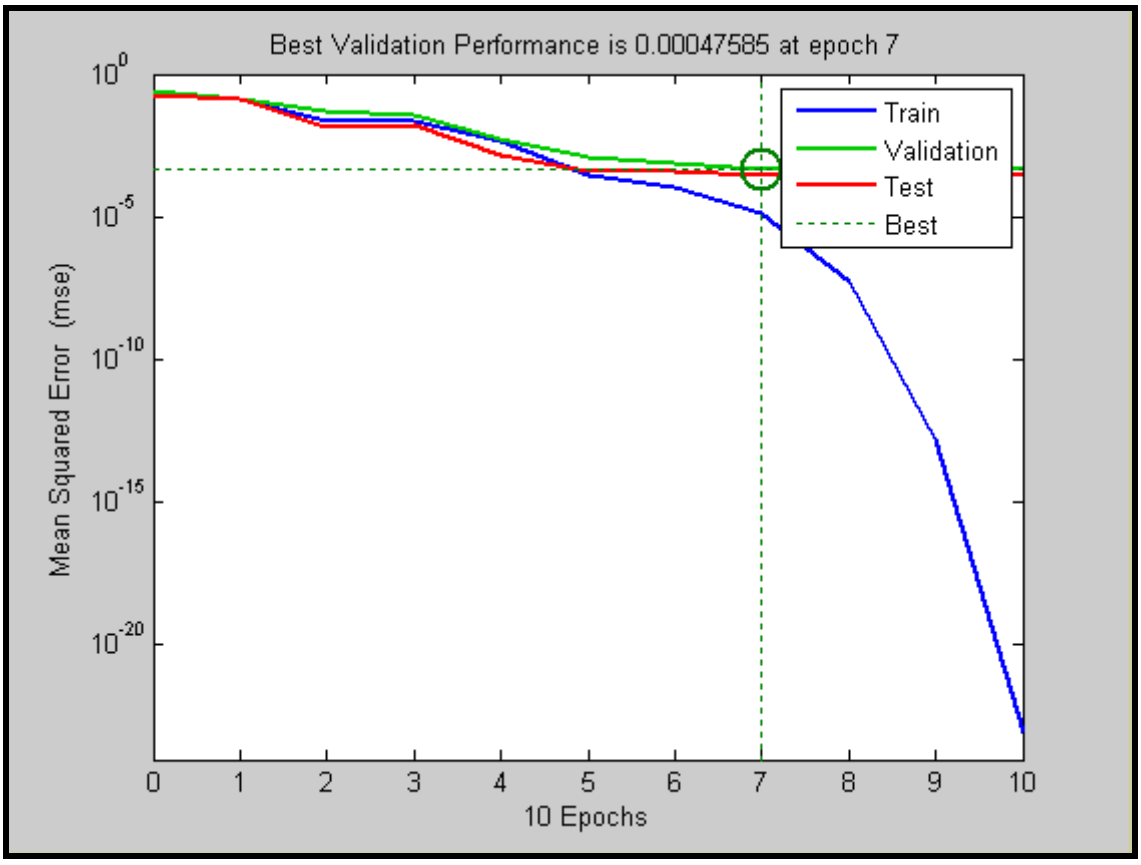


Figure 5.8 Performance plot for network LM308

For this particular network, training was stopped at ten epochs and the best validation performance was returned at epoch seven. As it is evident from Figure 5.8, all plots (training, validation and test plots) followed the same general trends moving downwards towards lowest mean square error (MSE) value. At epoch ten, the training was eventually stopped before the network began to over-fit the dataset. From epoch seven to ten, the training error continued to decrease while the validation error began to increase indicating that the optimum network performance occurred at epoch seven.

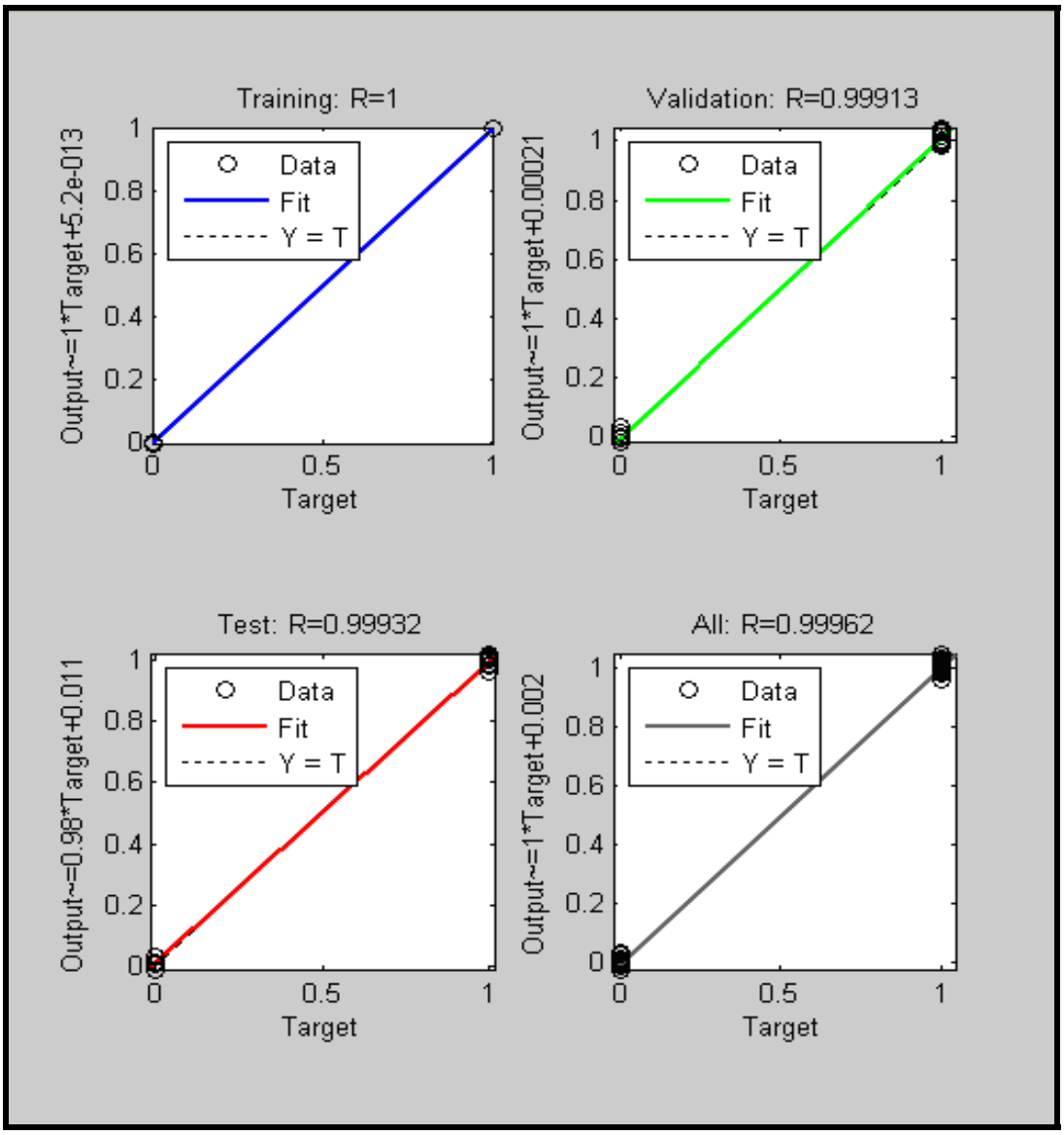


Figure 5.9 Regression plot for network LM308

All regression plots for network LM308 provided correlation coefficients close to one (R = 1 for the training subset, R = 0.99913 for the validation subset, R = 0.99932 for the test subset and R = 0.99962 for the entire dataset). This indicates that the output tracked the test target very well.

5.5.3 MLP Neural Networks Trained Using Dataset 2

The best performances for the training, validation and test set for each training algorithm for the MLP neural networks trained using Dataset 2 are presented in Tables 5.11 to 5.13.

Table 5.11 Training, validation and test performances for MLP neural network trained using back-propagation (BP) algorithm with Dataset 2

NETWORK	NO OF HIDDEN NEURONS	BEST PERFORMANCE		
		Training	Validation	Test
BP102	2	0.0436	0.0506	0.0595
BP202	2	0.0446	0.1258	0.0857
BP302	2	0.1214	0.1808	0.1979
BP103	3	0.0333	0.0691	0.0495
BP203	3	0.0795	0.1385	0.1158
BP303	3	0.6027	0.4914	0.6709
BP104	4	0.1265	0.1626	0.1642
BP204	4	0.5965	0.4844	0.5507
BP304	4	0.0594	0.1067	0.1337
BP105	5	0.0247	0.0726	0.0754
BP205	5	0.1567	0.5222	0.2870
BP305	5	0.1769	0.6443	0.3148
BP106	6	0.0482	0.0784	0.0413
BP206	6	0.0514	0.1116	0.1306
BP306	6	0.0237	0.0410	0.0409
BP107	7	0.0399	0.1026	0.0635
BP207	7	0.0267	0.0477	0.0607
BP307	7	0.0343	0.0767	0.0751
BP108	8	0.0197	0.0412	0.0615
BP208	8	0.0201	0.0425	0.0445
BP308	8	0.0311	0.0775	0.0318
BP109	9	0.0431	0.0685	0.1029
BP209	9	0.0658	0.0640	0.1347
BP309	9	0.0374	0.0904	0.0621
BP110	10	0.0220	0.0234	0.0303
BP210	10	0.0194	0.0666	0.1112
BP310	10	0.0333	0.1291	0.0585
BP111	11	0.0214	0.0170	0.0634
BP211	11	0.0224	0.1281	0.0985
BP311	11	0.0221	0.0891	0.0794

Table 5.12 Training, validation and test performances for MLP neural network trained using conjugate gradient (CG) algorithm with Dataset 2

NETWORK	NO OF HIDDEN NEURONS	BEST PERFORMANCE		
		Training	Validation	Test
CG102	2	8.94E-4	0.0068	0.0092
CG202	2	2.93E-4	2.54E-4	4.77E-4
CG302	2	0.0023	0.0280	0.0099
CG103	3	0.0013	0.0123	0.0230
CG203	3	9.24E-4	0.1300	0.1690
CG303	3	0.0091	0.0818	0.1423
CG104	4	0.0018	0.0118	0.0183
CG204	4	0.0017	0.0285	0.0437
CG304	4	5.63E-4	0.0013	0.0033
CG105	5	2.15E-4	0.0044	0.0020
CG205	5	0.0015	0.0097	0.0086
CG305	5	1.64E-4	5.12E-4	0.0015
CG106	6	2.29E-4	8.93E-4	0.0186
CG206	6	0.0084	0.0630	0.0561
CG306	6	0.0128	0.0534	0.0674
CG107	7	0.0022	0.0799	0.0912
CG207	7	2.19E-4	7.16E-4	0.0012
CG307	7	0.0083	0.0812	0.1424
CG108	8	0.0118	0.0539	0.1403
CG208	8	0.0048	0.0226	0.0202
CG308	8	0.0148	0.0487	0.0242
CG109	9	3.04E-4	3.21E-4	4.84E-4
CG209	9	1.28E-4	3.69E-4	6.69E-4
CG309	9	0.0017	0.0121	0.0161
CG110	10	0.0053	0.0137	0.0188
CG210	10	6.30E-4	0.0211	0.0205
CG310	10	0.0080	0.0643	0.0904
CG111	11	0.0016	0.0108	0.0146
CG211	11	0.0165	0.0788	0.0369
CG311	11	0.0074	0.0615	0.1011

Table 5.13 Training, validation and test performances for MLP neural network trained using Lavenberg-Marquardt (LM) algorithm with Dataset 2

NETWORK	NO OF HIDDEN NEURONS	BEST PERFORMANCE		
		Training	Validation	Test
LM102	2	7.79E-14	2.03E-4	5.59E-5
LM202	2	0.0535	0.1534	0.1489
LM302	2	7.55E-5	0.0191	0.0135
LM103	3	8.21E-5	0.0332	0.0190
LM203	3	6.85E-9	5.91E-4	0.0313
LM303	3	0.1826	0.1549	0.1628
LM104	4	7.94E-8	0.0139	0.0229
LM204	4	1.09E-5	0.0168	0.0145
LM304	4	3.44E-8	0.0545	0.0684
LM105	5	8.2086	0.2139	0.0163
LM205	5	9.19E-10	5.11E-6	0.0370
LM305	5	1.94E-8	0.0209	0.0542
LM106	6	3.21E-7	0.0481	0.1224
LM206	6	1.92E-6	0.0028	0.0020
LM306	6	1.52E-30	0.0207	0.0139
LM107	7	1.10E-30	0.0866	0.1164
LM207	7	1.45E-11	0.0021	0.0281
LM307	7	1.89E-6	0.1099	0.0166
LM108	8	6.48E-31	0.0495	0.0238
LM208	8	9.23E-22	0.0060	0.0056
LM308	8	9.68E-7	0.0132	0.0862
LM109	9	2.93E-6	0.2130	0.0175
LM209	9	1.51E-31	0.0326	0.0203
LM309	9	7.74E-12	0.0145	0.0087
LM110	10	7.66E-29	0.0867	0.0401
LM210	10	1.06E-22	0.0303	0.1127
LM310	10	2.28E-31	0.0802	0.0603
LM111	11	2.94E-31	0.0072	0.0030
LM211	11	2.11E-31	0.0032	0.0026
LM311	11	6.3999e-8	0.0816	0.0317

As evident from Table 5.11 to 5.13, all networks (regardless of the training algorithm used) give good performances. The MLP neural networks trained using the LM algorithm again provided the lowest values followed by those trained using CG and the BP algorithm. These findings are similar to the results revealed with the MLP neural networks trained with Dataset 1.

Table 5.14 to 5.16 illustrate the regression coefficient (R) of the training, validation, and test subset and for the overall dataset for the MLP neural networks trained using the back-propagation (BP), conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithm.

Table 5.14 Regression coefficients (R) of the training, validation and test subset and for the overall dataset for the MLP neural networks trained using BP algorithm with Dataset 2

NETWORK	REGRESSION COEFFICENT (R)			
	Training	Validation	Test	All
BP102	0.96550	0.96992	0.95198	0.96364
BP202	0.96694	0.88656	0.92659	0.94542
BP302	0.90415	0.85854	0.84260	0.88050
BP103	0.97300	0.94958	0.95867	0.96582
BP203	0.93572	0.91144	0.88074	0.92183
BP303	0.32982	0.25353	0.41605	0.29732
BP104	0.90077	0.83917	0.85380	0.88372
BP204	0.34838	0.25869	0.56472	0.36448
BP304	0.95153	0.91456	0.89597	0.93333
BP105	0.98061	0.95290	0.93788	0.96603
BP205	0.96702	0.92267	0.94148	0.95417
BP305	0.84035	0.53001	0.76669	0.76117
BP106	0.96036	0.95887	0.96663	0.95887
BP206	0.95805	0.92093	0.89516	0.93730
BP306	0.92841	0.96696	0.95835	0.97610
BP107	0.96773	0.92014	0.94999	0.95520
BP207	0.98012	0.95698	0.94359	0.97057
BP307	0.97387	0.94936	0.92616	0.96014
BP108	0.98477	0.96346	0.95392	0.97480
BP208	0.98512	0.96364	0.96099	0.97715
BP308	0.97422	0.95972	0.97236	0.96871
BP109	0.96367	0.94754	0.93076	0.95297
BP209	0.95143	0.94307	0.89171	0.93715
BP309	0.97078	0.94050	0.95700	0.96034
BP110	0.98275	0.98220	0.97842	0.98157
BP210	0.98382	0.95188	0.94613	0.95626
BP310	0.97459	0.90017	0.95744	0.95535
BP111	0.98427	0.98552	0.93634	0.97742
BP211	0.98087	0.91216	0.92951	0.95424
BP311	0.98237	0.92335	0.94637	0.96368

Table 5.15 Regression coefficients (R) of the training, validation and test subset and for the overall dataset for the MLP neural networks trained using CG algorithm with Dataset 2

NETWORK	REGRESSION COEFFICENT (R)			
	Training	Validation	Test	All
CG102	0.99930	0.99407	0.99021	0.99649
CG202	0.99976	0.99982	0.99971	0.99975
CG302	0.99817	0.97195	0.99243	0.99183
CG103	0.99895	0.99077	0.98620	0.99387
CG203	0.99929	0.90646	0.80108	0.94214
CG303	0.99273	0.92846	0.89713	0.96087
CG104	0.99845	0.99227	0.98542	0.99397
CG204	0.99859	0.96990	0.97432	0.98740
CG304	0.99955	0.99913	0.99769	0.99905
CG105	0.99983	0.99615	0.99850	0.99895
CG205	0.99888	0.99097	0.99321	0.99600
CG305	0.99988	0.99957	0.99879	0.99962
CG106	0.98957	0.96090	0.92944	0.96881
CG206	0.99356	0.94112	0.92813	0.96876
CG306	0.99976	0.99919	0.98413	0.99596
CG107	0.99823	0.94910	0.89811	0.97020
CG207	0.99985	0.99925	0.99919	0.99960
CG307	0.99337	0.92521	0.81589	0.95727
CG108	0.98983	0.95719	0.91659	0.96359
CG208	0.99629	0.98923	0.98313	0.99099
CG308	0.96800	0.96943	0.95781	0.97441
CG109	0.99976	0.99969	0.99965	0.99973
CG209	0.99990	0.99970	0.99942	0.99978
CG309	0.99865	0.98855	0.98781	0.99458
CG110	0.99954	0.98459	0.98247	0.99351
CG210	0.99580	0.98918	0.98692	0.99230
CG310	0.99402	0.93929	0.92627	0.97310
CG111	0.99884	0.99050	0.98590	0.99583
CG211	0.98683	0.93892	0.93801	0.96723
CG311	0.99452	0.96591	0.93861	0.97508

Table 5.16 Regression coefficients (R) of the training, validation and test subset and for the overall dataset for the MLP neural networks trained using LM algorithm with Dataset 2

NETWORK	REGRESSION COEFFICENT (R)			
	Training	Validation	Test	All
LM102	1.00000	0.99984	0.99991	0.99995
LM202	0.96141	0.72098	0.85822	0.90256
LM302	0.99995	0.96798	0.98844	0.99155
LM103	0.99993	0.94473	0.98661	0.98268
LM203	1.00000	0.99800	0.95475	0.98837
LM303	0.84565	0.86674	0.88723	0.85589
LM104	1.00000	0.89542	0.98152	0.97525
LM204	0.99999	0.92132	0.98718	0.97984
LM304	1.00000	0.86677	0.95385	0.96518
LM105	0.99999	0.77797	0.97469	0.94793
LM205	1.00000	0.99999	0.95401	0.98974
LM305	1.00000	0.97968	0.94170	0.98606
LM106	1.00000	0.96931	0.84142	0.96418
LM206	1.00000	0.97708	0.99844	0.99919
LM306	1.00000	0.97592	0.98333	0.99134
LM107	1.00000	0.90292	0.89392	0.96398
LM207	1.00000	0.99826	0.96371	0.99318
LM307	1.00000	0.93934	0.96614	0.97602
LM108	1.00000	0.96595	0.97726	0.98842
LM208	1.00000	0.99570	0.99542	0.99819
LM308	1.00000	0.97489	0.92741	0.97938
LM109	1.00000	0.73670	0.87016	0.92386
LM209	1.00000	0.97410	0.98101	0.99198
LM309	1.00000	0.98186	0.98925	0.99378
LM110	1.00000	0.92808	0.96385	0.98091
LM210	1.00000	0.97239	0.86240	0.97602
LM310	1.00000	0.94735	0.95968	0.97928
LM111	1.00000	0.99346	0.99753	0.99824
LM211	1.00000	0.99760	0.99776	0.99910
LM311	1.00000	0.93324	0.98028	0.98320

The majority of the networks with the exception of BP302, BP303, BP104, BP204, BP205, BP209, CG203, CG303, CG307, LM202, LM303, LM304, LM107, LM109 and LM210, have correlation coefficients (R) very close to unity for the training, validation, test subset and for the overall dataset indicating that the targets have been very well tracked by the networks. The classification matrix for each MLP neural network trained

using the back-propagation (BP), conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithm are presented in Tables 5.17 to 5.19 respectively. Networks which have misclassified the data are shown in red.

Table 5.17 Classification matrix for MLP neural networks trained using back-propagation (BP) algorithm with Dataset 2 which consisted of 63 lipsticks, 54 lip balms and 42 shoe polishes.

NETWORK	SAMPLE	SAMPLE			SAMPLE (%)		
		L	LB	SP	L	LB	SP
BP102	L	51	32	0	32.08	20.13	0.00
	LB	0	38	0	0.00	23.90	0.00
	SP	0	17	21	0.00	10.69	13.21
BP202	L	48	24	0	30.19	15.09	0.00
	LB	3	42	3	1.89	26.42	1.89
	SP	0	15	24	0.00	9.43	15.09
BP302	L	31	39	0	19.50	24.53	0.00
	LB	15	29	0	9.43	18.24	0.00
	SP	0	22	23	0.00	13.84	14.47
BP103	L	44	36	0	27.67	22.64	0.00
	LB	7	29	1	4.40	18.24	0.63
	SP	0	21	21	0.00	13.21	13.21
BP203	L	20	31	0	12.58	19.50	0.00
	LB	14	36	0	8.81	22.64	0.00
	SP	0	35	23	0.00	22.01	14.47
BP303	L	26	18	0	16.35	11.32	0.00
	LB	18	35	0	11.32	22.01	0.00
	SP	0	62	0	0.00	38.99	0.00
BP104	L	37	48	0	23.27	30.19	0.00
	LB	9	23	0	5.66	14.47	0.00
	SP	0	24	18	0.00	15.09	11.32
BP204	L	43	15	0	27.04	9.43	0.00
	LB	9	55	0	5.66	34.59	0.00
	SP	0	46	0	0.00	28.93	0.00
BP304	L	26	29	0	16.35	18.24	0.00
	LB	12	38	3	7.55	23.90	1.89
	SP	0	30	21	0.00	18.87	13.21
BP105	L	39	25	0	24.53	15.72	0.00
	LB	9	45	1	5.66	28.30	0.63
	SP	0	18	22	0.00	11.32	13.84
BP205	L	34	25	0	21.38	15.72	0.00
	LB	10	46	0	6.29	28.93	0.00
	SP	0	19	25	0.00	11.95	15.72

Table 5.17 Continued

NETWORK	SAMPLE	SAMPLE			SAMPLE (%)		
		L	LB	SP	L	LB	SP
BP305	L	14	30	0	8.81	18.87	0.00
	LB	9	57	2	5.66	35.85	1.26
	SP	0	38	9	0.00	23.90	5.66
BP106	L	36	28	0	22.64	17.61	0.00
	LB	12	34	3	7.55	21.38	1.89
	SP	0	27	19	0.00	16.98	11.95
BP206	L	34	31	0	21.38	19.50	0.00
	LB	6	34	0	3.77	21.38	0.00
	SP	2	28	24	1.26	17.61	15.09
BP306	L	48	17	0	30.19	10.69	0.00
	LB	7	38	0	4.40	23.90	0.00
	SP	0	14	35	0.00	8.81	22.01
BP107	L	40	10	0	25.16	6.29	0.00
	LB	7	52	2	4.40	32.70	1.26
	SP	2	20	26	1.26	12.58	16.35
BP207	L	35	19	0	22.01	11.95	0.00
	LB	8	32	1	5.03	20.13	0.63
	SP	0	30	34	0.00	18.87	21.38
BP307	L	36	22	0	22.64	13.84	0.00
	LB	5	36	3	3.14	22.64	1.89
	SP	0	33	24	0.00	20.75	15.09
BP108	L	38	15	0	23.90	9.43	0.00
	LB	12	50	3	7.55	31.45	1.89
	SP	0	11	30	0.00	6.92	18.87
BP208	L	50	18	0	31.45	11.32	0.00
	LB	3	35	0	1.89	22.01	0.00
	SP	0	20	33	0.00	12.58	20.75
BP308	L	41	29	0	25.79	18.24	0.00
	LB	8	40	2	5.03	25.16	1.26
	SP	0	19	20	0.00	11.95	12.58
BP109	L	35	27	0	22.01	16.98	0.00
	LB	7	44	0	4.40	27.67	0.00
	SP	0	22	24	0.00	13.84	15.09
BP209	L	27	40	0	16.98	25.16	0.00
	LB	12	36	6	7.55	22.64	3.77
	SP	0	26	12	0.00	16.35	7.55
BP309	L	27	30	0	16.98	18.87	0.00
	LB	4	42	3	2.52	26.42	1.89
	SP	2	27	24	1.26	16.98	15.09
BP110	L	49	17	0	30.82	10.69	0.00
	LB	8	46	0	5.03	28.93	0.00
	SP	0	9	30	0.00	5.66	18.87

Table 5.17 Continued

NETWORK	SAMPLE	SAMPLE			SAMPLE (%)		
		L	LB	SP	L	LB	SP
BP210	L	42	18	0	26.42	11.32	0.00
	LB	0	46	2	0.00	28.93	1.26
	SP	0	20	31	0.00	12.58	19.50
BP310	L	40	11	0	25.16	6.92	**0.00**
	LB	6	54	6	3.77	33.96	**3.77**
	SP	0	14	28	0.00	8.81	**17.61**
BP111	L	43	16	0	27.04	10.06	**0.00**
	LB	2	45	2	1.26	28.30	**1.26**
	SP	3	17	31	1.89	10.69	**19.50**
BP211	L	43	16	0	27.04	10.06	**0.00**
	LB	8	53	1	5.03	33.33	**0.63**
	SP	0	15	23	0.00	9.43	**14.47**
BP311	L	47	18	0	29.56	11.32	**0.00**
	LB	3	47	5	1.89	29.56	**3.14**
	SP	**0**	**19**	**20**	**0.00**	**11.95**	**12.58**

Table 5.18 Classifications matrix for MLP neural networks trained using conjugate gradient (CG) algorithm with Dataset 2 consisting of 63 lipsticks, 54 lip balms and 42 shoe polishes

NETWORK	SAMPLE	SAMPLE			SAMPLE (%)		
		L	LB	SP	L	LB	SP
CG102	L	57	0	0	35.85	0.00	0.00
	LB	3	54	0	1.89	33.96	0.00
	SP	0	3	42	0.00	1.89	26.42
CG202	L	**63**	**0**	**0**	**39.62**	**0.00**	**0.00**
	LB	**0**	**54**	**0**	**0.00**	**33.96**	**0.00**
	SP	**0**	**0**	**42**	**0.00**	**0.00**	**26.42**
CG302	L	63	9	0	39.62	5.66	0.00
	LB	0	48	0	0.00	30.19	0.00
	SP	0	0	39	0.00	0.00	24.53
CG103	L	60	3	0	37.74	1.89	0.00
	LB	2	51	0	1.26	32.08	0.00
	SP	1	3	39	0.63	1.89	24.53
CG203	L	60	3	3	37.74	1.89	1.89
	LB	0	51	0	0.00	32.08	0.00
	SP	0	0	42	0.00	0.00	26.42
CG303	L	57	18	0	35.85	11.32	0.00
	LB	0	42	0	0.00	26.42	0.00
	SP	3	3	36	1.89	1.89	22.64
CG104	L	62	7	0	38.99	4.40	0.00
	LB	0	53	0	0.00	33.33	0.00
	SP	0	1	36	0.00	0.63	22.64

Table 5.18 Continued

NETWORK	SAMPLE	SAMPLE			SAMPLE (%)		
		L	LB	SP	L	LB	SP
CG204	L	57	5	0	35.85	3.14	0.00
	LB	0	58	3	0.00	36.48	1.89
	SP	0	3	33	0.00	1.89	20.75
CG304	L	63	0	0	39.62	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	0	42	0.00	0.00	26.42
CG105	L	60	0	0	37.74	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	3	42	0.00	1.89	26.42
CG205	L	63	0	0	37.74	0.00	0.00
	LB	0	54	0	0.00	32.70	1.89
	SP	0	0	42	0.00	3.14	24.53
CG305	L	63	0	0	39.62	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	0	42	0.00	0.00	26.42
CG106	L	35	16	0	22.01	10.06	0.00
	LB	12	49	0	7.55	30.82	0.00
	SP	3	15	29	1.89	9.43	18.24
CG206	L	51	6	0	32.08	3.77	0.00
	LB	6	47	3	3.77	29.56	1.89
	SP	0	13	33	0.00	8.18	20.75
CG306	L	62	0	0	38.99	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	1	42	0.00	0.63	26.42
CG107	L	60	7	0	37.74	4.40	0.00
	LB	0	55	0	0.00	34.59	0.00
	SP	0	6	31	0.00	3.77	19.50
CG207	L	63	0	0	39.62	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	0	42	0.00	0.00	26.42
CG307	L	54	13	0	33.96	8.18	0.00
	LB	3	50	2	1.89	31.45	1.26
	SP	0	10	27	0.00	6.29	16.98
CG108	L	45	23	0	28.30	14.47	0.00
	LB	10	49	3	6.29	30.82	1.89
	SP	0	5	24	0.00	3.14	15.09
CG208	L	54	10	0	33.96	6.29	0.00
	LB	3	45	3	1.89	28.30	1.89
	SP	0	9	35	0.00	5.66	22.01
CG308	L	40	20	2	25.16	12.58	1.26
	LB	12	51	1	7.55	32.08	0.63
	SP	0	15	18	0.00	9.43	11.32

Table 5.18 Continued

NETWORK	SAMPLE	SAMPLE			SAMPLE (%)		
		L	LB	SP	L	LB	SP
CG109	L	63	0	0	39.62	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	0	42	0.00	0.00	26.42
CG209	L	63	0	0	39.62	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	0	42	0.00	0.00	26.42
CG309	L	59	6	0	37.11	3.77	0.00
	LB	0	51	0	0.00	32.08	0.00
	SP	0	4	39	0.00	2.52	24.53
CG110	L	59	6	0	37.11	3.77	0.00
	LB	3	48	0	1.89	30.19	0.00
	SP	0	1	42	0.00	0.63	26.42
CG210	L	55	13	0	34.59	8.18	0.00
	LB	6	47	0	3.77	29.56	0.00
	SP	0	2	36	0.00	1.26	22.64
CG310	L	42	21	0	26.42	13.21	0.00
	LB	6	42	4	3.77	26.42	2.52
	SP	0	15	29	0.00	9.43	18.24
CG111	L	60	1	0	37.74	0.63	0.00
	LB	0	55	0	0.00	34.59	0.00
	SP	0	2	41	0.00	1.26	25.79
CG211	L	21	17	0	13.21	10.69	0.00
	LB	20	41	3	12.58	25.79	1.89
	SP	0	30	27	0.00	18.87	16.98
CG311	L	46	9	0	28.93	5.66	0.00
	LB	3	50	0	1.89	31.45	0.00
	SP	3	9	36	1.89	5.66	22.64

Table 5.19 Classifications matrix for MLP neural networks trained using Lavenberg-Marquardt (LM) algorithm with Dataset 2 consisting of 63 lipsticks, 54 lip balms and 42 shoe polishes

NETWORK	SAMPLE	SAMPLE			SAMPLE (%)		
		L	LB	SP	L	LB	SP
LM102	L	63	0	0	39.62	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	0	42	0.00	0.00	26.42
LM202	L	50	11	0	31.45	6.92	0.00
	LB	0	50	3	0.00	31.45	1.89
	SP	0	13	32	0.00	8.18	20.13
LM302	L	63	5	0	39.62	3.14	0.00
	LB	1	46	0	0.63	28.93	0.00
	SP	0	2	42	0.00	1.26	26.42
LM103	L	58	0	1	36.48	0	0.63
	LB	0	57	0	0	35.85	0
	SP	0	5	39	0	3.14	24.53
LM203	L	63	0	0	39.62	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	0	42	0.00	0.00	26.42
LM303	L	61	24	0	38.36	15.09	0
	LB	0	22	0	1.89	13.84	0
	SP	0	16	33	0	10.06	20.75
LM104	L	63	1	0	39.62	0.63	0.00
	LB	0	51	0	0.00	32.08	0.00
	SP	0	0	44	0.00	0.00	27.67
LM204	L	62	1	0	38.99	0.63	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	3	39	0.00	1.89	24.53
LM304	L	60	10	0	37.74	6.29	0.00
	LB	0	43	0	0.00	27.04	0.00
	SP	0	4	42	0.00	2.52	26.42
LM105	L	53	23	0	33.33	14.47	0.00
	LB	1	30	0	0.63	18.87	0.00
	SP	0	31	21	0.00	19.50	13.21
LM205	L	63	1	0	39.62	0.63	0.00
	LB	1	52	0	0.63	32.70	0.00
	SP	0	0	42	0.00	0.00	26.42
LM305	L	60	6	0	37.74	3.77	0.00
	LB	0	55	0	0.00	34.59	0.00
	SP	0	2	36	0.00	1.26	22.64
LM106	L	48	0	0	30.19	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	3	9	45	1.89	5.66	28.30
LM206	L	61	0	0	38.36	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	2	42	0.00	1.26	26.42

Table 5.19 Continued

NETWORK	SAMPLE	SAMPLE			SAMPLE (%)		
		L	LB	SP	L	LB	SP
LM306	L	59	0	0	37.11	0.00	0.00
	LB	0	57	1	0.00	35.85	0.63
	SP	0	4	38	0.00	2.52	23.90
LM107	L	59	4	0	37.11	2.52	0.00
	LB	0	56	0	0.00	35.22	0.00
	SP	0	7	33	0.00	4.40	20.75
LM207	L	62	2	0	38.99	1.26	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	0	41	0.00	0.00	25.79
LM307	L	51	39	0	32.08	24.53	0.00
	LB	9	32	0	5.66	20.13	0.00
	SP	3	9	16	1.89	5.66	10.06
LM108	L	63	0	0	39.62	0.00	0.00
	LB	0	56	3	0.00	35.22	1.89
	SP	0	1	36	0.00	0.63	22.64
LM208	L	60	0	0	37.74	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	3	42	0.00	1.89	26.42
LM308	L	49	3	0	30.82	1.89	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	11	42	0.00	6.92	26.42
LM109	L	50	20	0	31.45	12.58	0.00
	LB	5	35	0	3.14	22.01	0.00
	SP	0	23	26	0.00	14.47	16.35
LM209	L	63	0	0	39.62	0.00	0.00
	LB	0	51	0	0.00	32.08	0.00
	SP	0	3	42	0.00	1.89	26.42
LM210	L	53	4	0	33.33	2.52	0.00
	LB	0	47	2	0.00	29.56	1.26
	SP	0	13	40	0.00	8.18	25.16
LM310	L	63	3	0	39.62	1.89	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	3	36	0.00	1.89	22.64
LM111	L	62	3	0	38.99	1.89	0.00
	LB	0	51	0	0.00	32.08	0.00
	SP	0	1	42	0.00	0.63	26.42
LM211	L	61	0	0	38.36	0.00	0.00
	LB	0	54	0	0.00	33.96	0.00
	SP	0	2	42	0.00	1.26	26.42
LM311	L	57	6	0	35.85	3.77	0.00
	LB	0	51	0	0.00	32.08	0.00
	SP	0	3	42	0.00	1.89	26.42

All of the MLP neural networks trained using the back-propagation (BP) algorithm misclassified all of the exemplars. Out of thirty networks trained using the conjugate gradient (CG) and the Lavenberg-Marquardt (LM) algorithms, twenty three (23) and twenty eight (28) networks misclassified the exemplars respectively. The higher number of networks that misclassified the exemplars was most likely due the complexity of Dataset 2 over Dataset 1. In the case of Dataset 2 the conjugate gradient (CG) algorithm returned the greatest number of functional neural networks.

5.5.3.1 Selection of Optimum MLP Neural Networks trained using Dataset 2

All MLP neural networks with small differences between their validation and test performances (in Table 5.11 to 5.13) and regression coefficients close to 1 (in Table 5.14 to 5.16) and neural networks with eight or less misclassifications representing 5% of misclassification across the data matrices (in Table 5.17 to 5.19) were used to classify the unknown test set.

From a total of ninety MLP neural networks trained using back-propagation (BP), conjugate gradient (CG) and Lavenberg-Marquardt (LM) algorithm, ten MLP neural networks fit the set criteria (five were trained using conjugate gradient (CG) algorithm and the rest were trained using Lavenberg-Marquardt (LM) algorithm). The profiles of the ten MLP neural networks are presented in Table 5.20.

Table 5.20 Profiles of the ten MLP neural networks from Dataset 2 for the classification of unknown exemplars

NETWORK	PERFORMANCE ERROR		TEST REGRESSION (R)	MISCLASSIFICATION (%)
	Validation	Test		
CG104	0.01180	0.0183	0.98542	5.00
CG205	0.00970	0.0086	0.99321	0.00
CG306	0.00534	0.0674	0.98431	0.63
CG109	3.21E-4	4.84E-4	0.99965	0.00
CG111	0.0108	0.0146	0.98590	1.89
LM204	0.0168	0.0145	0.98718	2.52
LM206	0.0028	0.0020	0.99844	1.26
LM306	0.0207	0.0139	0.98333	3.14
LM209	0.0326	0.0203	0.98101	1.89
LM211	0.0032	0.0026	0.99776	1.26

5.5.3.2 Test Study MLP Dataset 2

The outcomes of the prediction by each of the MLP neural networks trained with Dataset 2 are presented in Table 5.24 to 5.25. Misclassifications are given in red.

Table 5.24 Prediction outcomes for the MLP neural network trained using conjugate gradient (CG) algorithm. 1 = lipstick, 2 = lip balm, 3 = shoe polish. Black characters indicate a correct prediction

Actual	CG104	CG205	CG306	CG109	CG111
1	2	2	1	2	2
1	2	2	1	2	2
1	2	2	1	2	2
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	2	1	1	1
1	1	2	1	1	1
1	1	2	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
2	1	2	2	2	2
2	1	2	2	2	2
2	1	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2

Table 5.24 Continued

Actual	CG104	CG205	CG306	CG109	CG111
2	2	2	3	2	2
3	2	3	3	3	2
3	2	3	3	3	2
3	2	3	3	3	2
3	2	2	3	3	2
3	2	2	3	3	2
3	2	2	3	3	2
3	3	2	3	2	1
3	2	2	3	2	1
3	3	2	3	2	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	2	1
1	0	1	1	2	1
1	1	1	1	1	1
1	1	1	1	2	1
3	3	3	3	3	3
3	3	3	3	3	3
3	3	3	3	3	3
3	3	3	3	3	3
3	2	2	3	2	2
3	2	2	3	2	2
3	3	3	3	3	3
3	3	3	3	3	3
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2

Table 5.25 Prediction outcomes for the MLP neural network trained using Lavenberg-Marquardt (LM) algorithm. 1 = lipstick, 2 = lip balm, 3 = shoe polish. Black characters indicate a correct prediction

Actual	LM204	LM206	LM306	LM209	LM211
1	3	2	2	2	2
1	3	2	2	2	2
1	3	2	2	2	2
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	2	2	2	3
1	1	2	2	2	2
1	1	2	2	2	3
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
2	2	2	2	1	1
2	2	2	2	1	1
2	2	2	2	1	1
2	2	2	2	1	2
2	2	2	2	1	2
2	3	2	2	1	2
2	2	2	2	2	2
2	2	2	2	2	2
2	2	2	2	2	2
3	3	3	3	3	3
3	3	3	3	3	3
3	3	3	3	3	3
3	1	2	2	3	3
3	2	2	2	3	3
3	1	2	2	3	3
3	2	2	2	2	1
3	2	2	2	2	1
3	2	2	2	2	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	2	1	1	2
1	1	1	1	1	2
1	1	1	1	1	2
3	3	2	2	3	3
3	3	3	3	3	2

Table 5.25 Continued

Actual	LM204	LM206	LM306	LM209	LM211
3	3	3	3	3	3
3	3	2	3	3	3
3	2	3	3	3	3
3	3	1	2	2	3
3	3	2	2	2	3
3	3	3	3	3	3
2	2	2	3	2	2
2	2	2	3	2	2
2	2	2	2	2	2
2	1	2	2	2	2
2	1	2	2	2	2
2	1	2	2	2	2
2	1	2	2	3	3
2	1	2	2	2	2

Of the twenty three MLP neural networks retained for dataset 2, only one network, CG306 (in green in Table 5.24) correctly predicted the entire unknowns.. This network which contained six hidden neurons in the hidden layer produced good performances in terms of validation and test error which as presented in Figure 5.10. The regression coefficient (R) for the test set was also excellent and is presented in Figure 5.11.

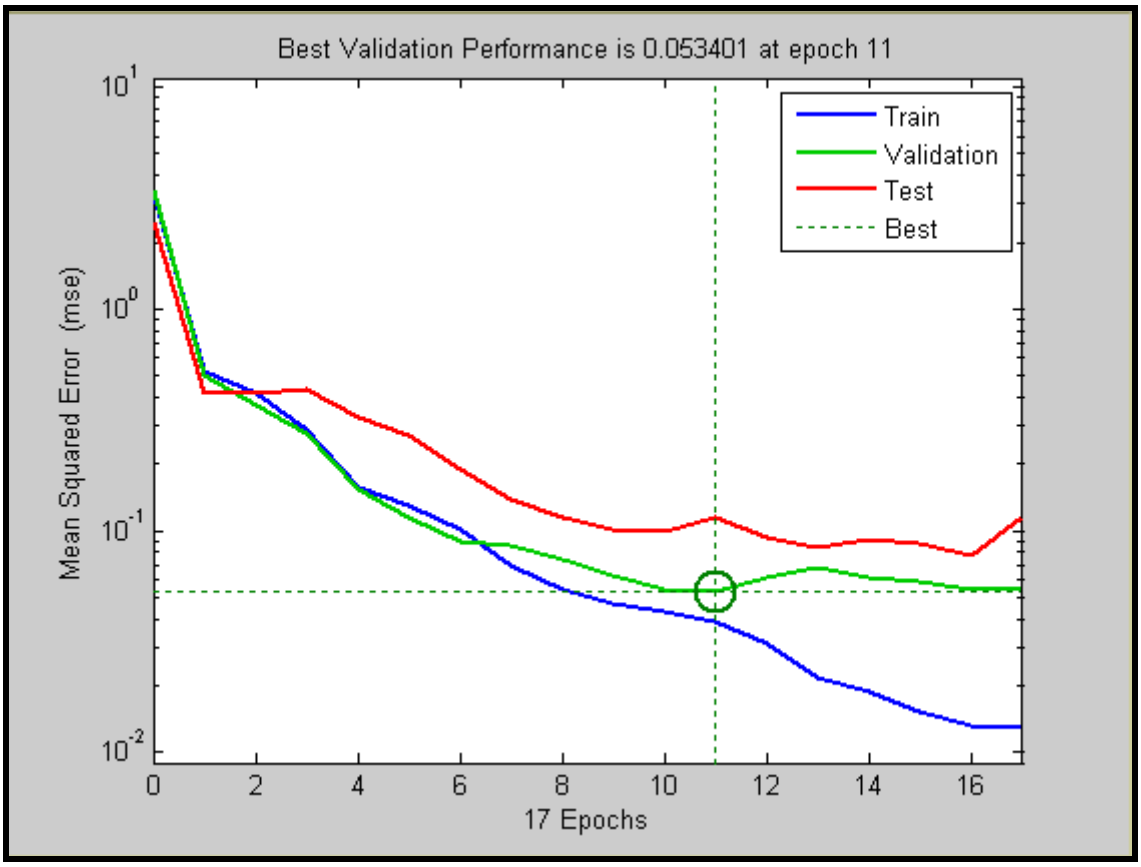


Figure 5.10 Performance plot for network CG306

Training was stopped at eighteen epochs and the best validation performance was returned at epoch eleven. At epoch eighteen, the training was stopped to prevent over-fitting of the dataset. From epoch eleven to eighteen, the training error continued to decrease while the validation error began to increase indicating that the optimum network performance occurred at epoch eleven.

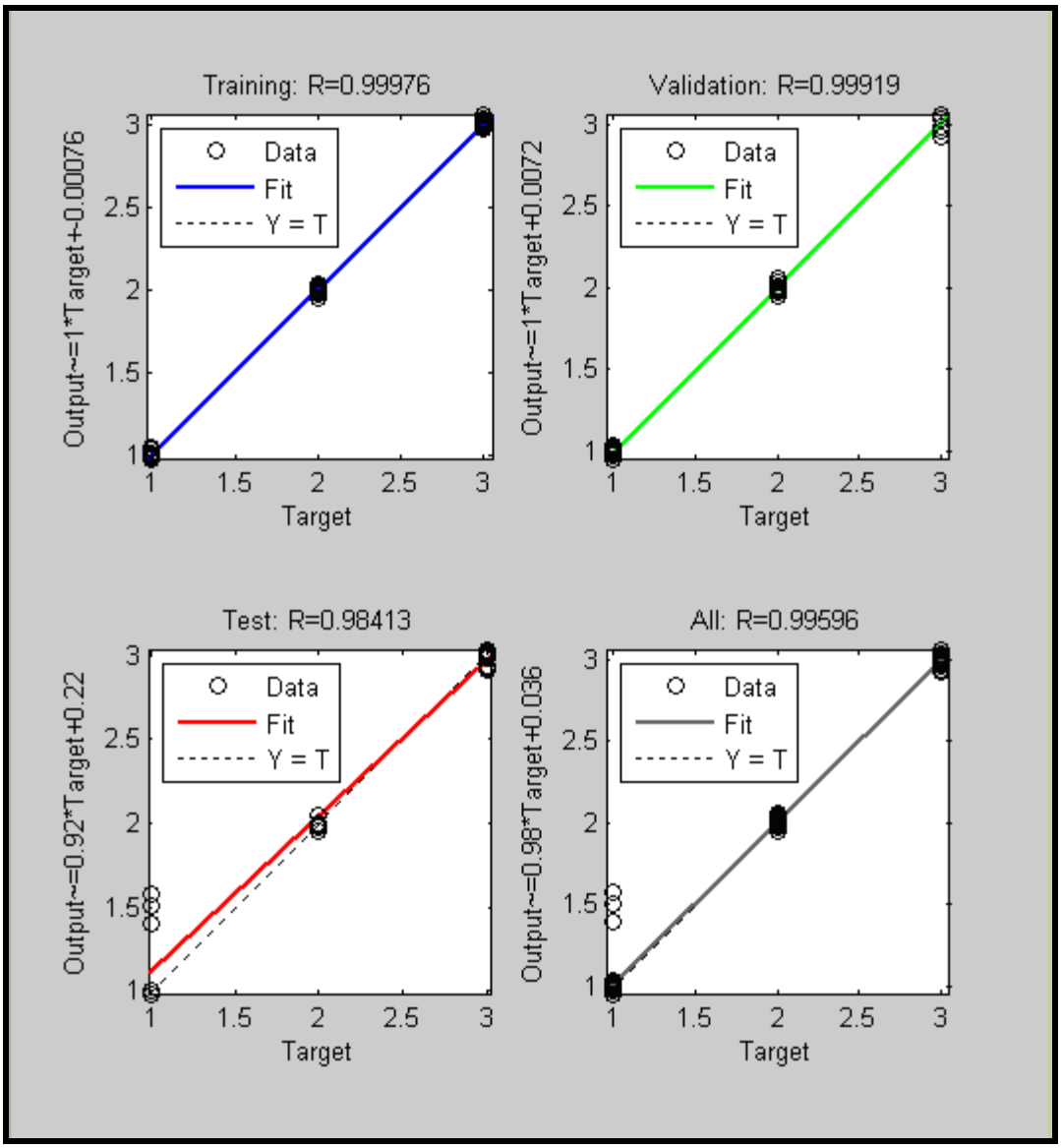


Figure 5.11 Regression plot for network CG306

All regression plots for network CG306 provided correlation coefficients close to one (R = 0.99976 for the training subset, R = 0.99919 for the validation subset, R = 0.98413 for the test subset and R = 0.99596 for the entire dataset) indicating that the output tracked the test target quite well.

Surprisingly, during training, network CG306 had a misclassification rate of 0.63% (compared to network LM308 which had 0.00% misclassification rate) despite correctly

classifying the training set. This suggests that the misclassification rate does not necessarily reflect the actual ability of the MLP neural network to correctly classify unknown samples however choosing a network with a low misclassification rate (as set in this study i.e. 5%) is a good approach in identifying network that is correctly trained.

5.6 CONCLUSION

The ability of MLP neural networks to learn and generalise can be exploited for use as a predictive tool, in this case, of wax based product. The MLP neural networks were developed using the Neural Network Toolbox in MATLAB®. A number of training algorithm platforms (back-propagation (BP), conjugate gradient (CG) and Lavenberg-Marquardt (LM)) were evaluated for the efficacy in differentiating and thereby correctly classifying the wax based products into their individual categories.

The MLP neural network trained with the Lavenberg-Marquardt (LM) algorithm which incorporated eight hidden neurons in its hidden layer (LM308) was the optimum neural network to model Dataset 1 (the TLC and MSP data alone for coloured samples) where it correctly classified thirty one test samples into their appropriate product types. The MLP network trained with the conjugate gradient (CG) algorithm which incorporated six hidden neurons in its hidden layer (CG306) was the optimum neural network to model Dataset 2 (the UV/Vis and pre-treated GC data for the complete dataset). Despite a 0.63% misclassification error during training, the network correctly classified fifty four test samples.

In both cases the MLP neural networks also correctly classified weathered samples into their correct products types despite the fact that no weathered exemplars were used in the training process. This demonstrates that there is certainly a potential for the use of MLP networks for the classification of unknown samples of data of this type.

5.7 REFERENCES

1. Svozil, D., Kvasnicka, V. and Pospichal, J., *Introduction to Multi-Layer Feed-Forward Neural Networks.* Chemometrics and Intelligence Laboratory Systems, 1997. **39**: p. 43 - 62.

2. *Longman Dictionary of Contemporary English*. 2nd Ed. 1987, Essex: Longman Group UK Limited.

3. Rossmo, D.K., *Geographic Profiling*. 2000, Washington: CRC Press.

4. Lek, S. and Guegan, J.F., *Artificial Neural Network as A Tool in Ecological Modelling: An Introduction.* Ecological Modelling, 1999. **120**: p. 65 - 73.

5. Wassermann, P.D., *Neural Computing Theory and Practice*. 1989, New York: Von Nostrand Reinhold.

6. Hagan, M.T., Demuth, H.B. and Beale, M., *Neural Network Design*. 1996, Boston: PWS Publishing Company.

7. Doble, P., Sandercock, M., Pasquier, E.D., Petocz, P., Roux, C. and Dawson, M., *Classification of Premium and Regular Gasoline by Gas Chromatography/Mass Spectrometry, Principal Component Analysis and Artificial Neural Networks.* Forensic Science International, 2003. **132**: p. 235 - 240.

8. Singh, K.P., Basant, A., Malik, A. and Jain, G., *Artificial Neural Network Modelling of the River Water Quality - A Case Study*. Ecological Modelling, 2009. **220**: p. 888-895.

9. Patterson, D.W., *Artificial Neural Networks - Theory and Applications*. 1996, London: Prentice Hall.

10. Goodacre, R., Hammond, D. and Kell, D.B., *Quantitative Analysis of the Adulteration of Orange Juice with Sucrose using Pyrolysis Mass Spectrometry and Chemometrics.* Journal of Analytical and Applied Pyrolysis, 1997. **40 - 41**: p. 135 - 158.

CHAPTER 6: GENERAL CONCLUSIONS AND FUTHER WORK

6.1 Pattern Recognition Techniques

Two different types of data were explored during this work using a variety of mathematical pattern recognition techniques. The wax based product data set presented a conventional challenge of sample discrimination using a combination of spectroscopic and chromatographic data. The lighter fluid data presented a more complex sample linkage problem where both brand differences and sample degradation were required to be taken into account.

In the investigation of data derived from wax based products and lighter fuels analysis, a number of attributes are examined and compared. Although such an approach has been successful and helpful, it can only deal with a limited number of samples and variables at one time. In a case where the attributes constitute patterns, such as UV/Vis spectra (in wax based products) and gas chromatographic profiles (in wax based products and lighter fuels) the result is a highly multidimensional dataset. Deciphering useful information from the dataset through direct examination is a daunting process which can be both time consuming and laborious to perform particularly when there are large number of samples or objects to deal with. The answer to this problem is the application of statistical techniques known collectively as pattern recognition techniques which are capable of identifying patterns (both similar and non-similar) within the dataset.

Pattern recognition itself is a term used to cover all stages of investigation of data starting from problem formulation, data collection through to discrimination, assessment of results and its interpretation [1]. Chemometric procedures and artificial neural networks (ANNs) are amongst pattern recognition techniques commonly employed to gather and visualise useful information from multidimensional datasets. Of the many pattern recognition techniques available, Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) are the most commonly used in the evaluation of data generated from analysis of samples of relevance to forensic science. In contrast, the

use of artificial neural networks (ANNs) such as Multi Layer Perceptron (MLP) and Self Organising Feature Maps (SOFM) are relatively limited in the forensic science related literature eventhough the techniques have been around for almost 30 years [2]. As this work has demonstrated, their potential use within forensic science is promising and may be of great significance.

Pattern recognition techniques, as a whole, can be divided into two main groups known as supervised and unsupervised techniques. The former requires an input and output pattern pair where the output pattern acts as a template from which the outcome of the pair is derived. Unsupervised pattern recognition on the other hand, does not require any output patterns. The outcome is natural grouping or clustering of similar input patterns whose properties can be explored and summarised as demonstrated in this study.

Whilst the mathematical outlines of HCA and SOFM are relatively easy to understand, PCA involves more complex computation of correlation or covariance matrices followed by computation of principal component loadings through eigenanalysis. Having said that, HCA has a wide number of linkage strategies each of which can produce a different outcome and therefore often require to be systematically tested on a given dataset [3]. Furthermore, despite being visualisation techniques, both fail to fully utilise the visualisation space [4] i.e. the score plot and dendrogram. The output map in SOFM can be regarded as equivalent to score plot and dendrogram and one obvious advantage of the output map is that it does not suffer from 'overcrowding of the visualisation space' as commonly experience in a score plot and dendrogram. Further to this, the SOFM component maps provide considerable opportunity for data mining to reveal the impact of specific variables on the overall sample groupings obtained.

PCA, HCA and SOFM are unsupervised pattern recognition techniques which have been employed in this study to find natural groupings or clustering within the data sets presented. Whilst both PCA and HCA failed to perform even with dataset obtained from basic analysis for example TLC and MSP (in the wax based product data set), SOFM

successfully attributed samples to type demonstrating a potentially greater applicability to data of this type. Similar results were obtained on more complex chromatographic data.

The supervised pattern recognition technique employed in this study was the Multi Layer Perceptron. MLP acts as universal function approximator which means that it can approximate underlying function(s) of a given dataset. As with any other mathematical function, the approximated function(s) has a predictive ability which means that it can be employed to determine the relationship of unknown inputs with known outputs. To be able to do this, the MLP must first be trained in order to capture any hidden information in a given dataset. This is done by evaluating and adjusting the connection weights between neurons in the MLP arrangement or architecture using systematic learning instruction or algorithm.

There are a number of learning algorithms available for MLP. Trial and error using datasets of known provenance is commonly used to resolve which algorithm is most effective for the dataset. Of the problems associated with MLP training, choosing the 'right' number of hidden neurons in the hidden layer for a given dataset is difficult and using trial and error or a heuristic process is found to be the most effective mechanism for this. Under-training and over-learning (over-fitting) can also occur however there are ways in which these can be prevented.

6.2 Datasets from Wax Based Products Analysis

Because a priori knowledge of the distribution of the wax based products under study was available, the optimum grouping or clustering of the samples was known and the ability of each data pre-processing techniques to correctly discriminate the samples could be evaluated.

Both simple and complex data were available from the wax based product analysis. A simple dataset was generated from the TLC analysis with maximum of two reference points i.e. the retardation factor of the first and second development of the coloured spot observed. More complicated datasets were generated from the MSP, UV/Vis and GC-FID analysis with multiple reference points for each sample under study.

PCA, HCA and SOFM applied to the simple dataset did not facilitate correct grouping of the samples in any case. However when the TLC data was combined with the MSP data (to further represent the coloured wax based products), SOFM successfully grouped the wax based products according to its type, an outcome which was not achieved using PCA and HCA techniques. This finding demonstrates the effectiveness of the combination of TLC and subsequent MSP analysis of the resultant spots as a discriminating methodology for coloured wax based products.

For the dataset which contained both non-coloured and coloured samples, the best data facilitating discrimination was presented as a combination of UV/Vis and GC using the square root transformation. All samples irrespective of their colour were discriminated into their separate product types using SOFM only. Furthermore the component maps elucidated with SOFM provided an insight into the strength of association provided by individual variables such as specific retention time or wavelength. Again this provides valuable mathematical evidence of the validity of SOFM in the discrimination of wax based products.

When degraded samples were examined, SOFM again facilitated the grouping by product type using the TLC and MSP combined dataset for coloured samples and UV/Vis and GC-FID square root combined dataset for both coloured and non-coloured samples.

MLP treated with both TLC and MSP combined dataset (Dataset 1) and UV/Vis and GC square root combined dataset (Dataset 2) provided good classification (according to product type) of the wax based products under study. The findings indicated that information i.e. in terms of type for coloured wax based products could be successfully determined using information gathered from TLC combined with MSP analysis. For both coloured and non-coloured wax based products, the type could also be successfully determined using information gathered from UV/Vis and GC-FID analysis.

Different algorithms settle on different minimum error surface (whether it be local or global minimum error surface) within a given dataset. The target in MLP training is to find the global minimum error surface, the point where the difference in error between the target and actual output is the lowest. In this study, MLP neural networks trained using BP algorithm showed poor results compared to those trained using CG (conjugate gradient) and LM (Lavenberg-Marquardt) algorithm. This is hardly surprising since BP is never assured of finding the global minimum error surface [5] compared to CG and LM algorithm. Since global minimum error surface is not known in advance and also which type of learning algorithm is 'capable of finding' the global minimum error surface (of a given dataset), it is therefore advisable to train MLP neural networks using different types of algorithms and select the one that gives the most promising outcomes.

6.3 Datasets from Lighter Fluid Analysis

The data generated from lighter fluid analysis provided a more significant challenge than the datasets generated from wax based product analysis. This was because of the number of target peaks within the chromatogram which were included within the data matrix and the effect of sample degradation on the chromatographic profile of the various samples analysed. PCA failed to correctly group all lighter fluid samples according to their brands after evaporation. HCA and SOFM on the other hand, successfully grouped the samples together even after the samples had undergone extreme evaporation resulting in the loss of 95% of its original components. However HCA also suggested groupings between branded samples that were clearly not

evidenced by their chromatograms. SOFM resolved this issue clearly separating all samples by brand. This finding provides valuable mathematical evidence of the validity of SOFM not only in sample discrimination but also in the linking of samples after they have undergone significant degradation.

6.4 Suggestions for Future Work

Pattern recognition techniques particularly the SOFM and MLP are not limited to datasets from chemical analysis. In fact, any kind of datasets can be 'explored' using both techniques as long as the datasets can be represented in the form of a set of numerical values. Datasets of qualitative or subjective data therefore must be converted to numerical values prior to its interpretation.

Further extension of SOFM and MLP to other datasets of forensic science relevance where pattern recognition is required would be warranted. Both artificial neural network systems have demonstrated considerable advantages over conventional chemometric systems. Given that appropriate conversions of subjective data could be made, SOFM in particular, has the potential to provide significant advances in the pattern recognition required for the comparison of evidence types such as toolmarks, ballistics, footwear marks, handwriting analysis, vein pattern analysis and so on. Similarly, for well defined datasets such as those associated with chromatographic and spectroscopic data this project has demonstrated the capabilities of the neural network techniques to facilitate sample and degraded sample linkages. Further applications could include drug impurity profiling, environmental forensic data analysis, ignitable liquid sample linkage, ink analysis including ink ageing.

The optimisation of chromatographic systems of forensic science interest (HPLC, GC-FID, GC-MS, GC-GC-TOF and electrophoresis systems) using MLP neural networks in conjunction with factorial design techniques would also be of interest particularly in the development of analytical methods.

6.5 REFERENCES

1. Webb, A., *Statistical Pattern Recognition*. 2nd Ed. 2002, West Sussex: John Wiley and Sons Ltd.

2. Wassermann, P.D., *Neural Computing Theory and Practice*. 1989, New York: Von Nostrand Reinhold.

3. Everitt, B.S. and Dunn, G., *Applied Multivariate Data Analysis*. 2001, London: Arnold.

4. Brereton, R.G., *Chemometrics for Pattern Recognition*. 2009, West Sussex: John Wiley and Sons Ltd.

5. Patterson, D.W., *Artificial Neural Networks - Theory and Applications*. 1996, London: Prentice Hall.

APPENDIX A

APPENDIX A: BACKGROUND OF THE ANALYTICAL TECHNIQUES

A.1 INTRODUCTION

This appendix describes the theoretical background of the analytical techniques used to generate the data in this project. The techniques used reside within two areas of analytical chemistry measurement, namely chromatographic and spectroscopic analysis. The background theory for both is presented together with an explanation of the basis of thin layer chromatography (TLC), microspectrophotometry (MSP), ultra-violet visible spectrometry (UV/Vis) and gas chromatography with either flame ionisation detection (GC-FID) or mass spectrometry (GC-MS).

A.2 Chromatography

According to the International Union of Pure and Applied Chemistry (IUPAC), chromatography is defined as 'a method used primarily for the separation of the components of samples, in which the components are distributed between two phases, one of which is stationary while the other moves. The stationary phase may be a solid, or a liquid supported on a solid, or a gel. The stationary phase may be packed in a column, spread as a layer, or distributed as a film, etc; in these the term chromatographic bed is used as a general term to denote any of the different forms in which the stationary phase may be used where the mobile phase may be of gaseous or liquid'[1].

The term "chromatography" was first used by a Russian chemist, Tswett (1872 – 1919), whose interest was in the separation of pigments [2]. He had packed powdered adsorbents such as alumina, silica, chalk and sucrose into a glass tube and eluted it with various solvents. This process resulted in the formation of discrete coloured bands along the glass tube, hence the derivation of the name of the technique, i.e. chroma which means colour [2]. After his death, this particular chromatographic technique was not used again until the 1930's, when it was revised by Zechmeister, Cholkony, Ruzicka and others [2].

Tswett's early form of column chromatography could not be successfully applied for separation of water soluble analytes and this lead to the development of partition chromatography in 1941 by Martin and Synge [2]. The technique utilised two liquid phases for the separation of amino acids. However the method was not particularly robust and to overcome this Martin and co-workers developed a paper chromatography system in the form of filter paper sheets impregnated with water or other liquids and used this to facilitate the separation of analytes. The process was slow and the filter paper was later replaced by thin layer of adsorbent applied on glass plate, a technique which became known as thin layer chromatography (TLC) [3]. Martin and Synge had also anticipated gas chromatography (GC) i.e. a chromatographic technique utilising gas as the moving phase [3]. GC however was not exploited until 1952 when Martin and James published their work on the separation of volatile fatty acids [4].

The development of GC led to the establishment of the relationships (postulated by van Deemter) which discussed the theoretical and physicochemical aspects of the chromatographic process [3]. The van Deemter equation is presented in Equation A.1.

$$H = A + \frac{B}{\bar{u}} + C\bar{u} \quad \text{..... equation A.1}$$

Where:

H = plate height

A = eddy diffusion

B = longitudinal molecular diffusion

C = mass transfer in the stationary liquid phase and;

$\bar{u}$ = average linear gas velocity

The van Deemter equation is commonly illustrated by a two dimensional plot of plate height versus the average linear gas velocity as presented in Figure A.1. The minimum value in the plot indicates the optimum velocity of mobile phase or carrier gas which

will provide the greatest efficiency of separation with least peak broadening [5] in a given chromatographic system.

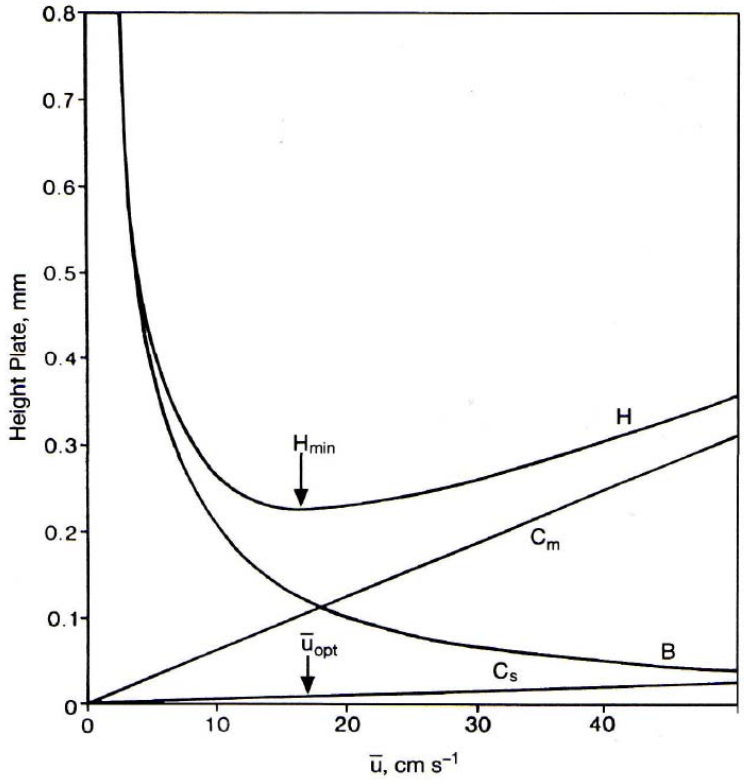


Figure A.1 van Deemter plot. Reproduced from McNair [5]

High performance liquid chromatography (HPLC) was developed at a later stage after GC [3]. It was initially called high-pressure liquid chromatography because the viscosity of the liquids used in the system required a higher pressure to push them through the instruments. It was later exchanged to high performance because smaller particles and shorter columns could be used reducing the necessity for very high pressures to be used [3].

In forensic science, chromatographic techniques, particularly thin layer chromatography, GC and HPLC are routinely applied to samples of forensic relevance such as drugs of abuse [6], inks in questioned documents [7, 8] and dyes in fibres analysis [9, 10]. Gas

chromatography coupled to a flame ionisation detector (GC-FID) is suitable for use with samples obtained from biological fluids [4], drug seizures [11, 12] and fire debris [13, 14]. HPLC has been used mainly in forensic toxicology [15, 16] although it has also been used in drug analysis [6, 17].

A.2.1 Principles of Chromatographic Separation

The principle of chromatographic separation is the same irrespective of the chromatographic technique employed and is based on the establishment of an equilibrium for the target analytes between the mobile and stationary phases. In any chromatographic technique, the process begins with the introduction of a mixture of components (the analytes) into the chromatographic system using either simple or specialised equipment. The sample is carried through the chromatographic system via a mobile phase (either a flowing liquid or a gas).

Separation of the mixture of analytes occurs through their interaction with the stationary phase within the chromatographic system. The degree of interaction between the components and the stationary phase depends on the nature of the components in terms of analyte polarity and size for example, as well as the analytes affinity for both the mobile and stationary phase matrices. The greater the interactions with the stationary phase the slower the analyte moves through the chromatographic system. As the various analytes migrate they undergo a series of equilibration steps between the stationary and mobile phases so that the separation becomes more pronounced as compounds progress through the chromatographic system [18]. In an ideal situation, the differences in interaction are sufficient to allow all the analytes in the sample to be completely separated from each other [1].

A.2.2 Modes of Chromatography

There are a number of modes of chromatography and the exact mode in a given application depends largely on the nature of the stationary phase used. The most common means by which analyte separation is achieved are via adsorption or partition

mechanisms. According to Braithwaite [1], whilst there may be one dominant mechanism, the means of separation are not mutually exclusive and may be the result of two or more mechanisms [19].

A.2.2.1 Liquid – Solid Adsorption Chromatography

Adsorption chromatography is regarded as the oldest type of chromatographic separation mechanism and is considered as liquid-solid chromatography (LSC) where a solvent or mixture of solvents is used as the mobile phase and the stationary phase is composed of either porous fine solid or porous layer beads packed within a column [20]. The former are commonly used as the stationary phases in LSC and normally utilise active materials such as silica gel. Other stationary phases available include alumina, carbon and magnesia and their applications are dependent on the type of analytes to be separated.

Chromatographic grade silica is commonly prepared by the reaction of sodium silicate and mineral acids such as hydrochloric and sulphuric acid followed by washing and drying. The final product is 'polysilicic acid' which is a stable, porous solid and terminated at its surfaces with either silanol (Si-OH) or siloxane (Si-O-Si) bonds. The silanol or surface hydroxyl group provides the means by which analyte-stationary phase interaction can occur. Stationary phases such as these are called 'straight or normal phases' stationary phases [21]. The silica material can be modified through the addition of non polar groups (most usually octadecyl groups) to create non polar interaction sites on the silica and are termed 'reversed phase'.

Separation of analytical samples into individual analytes in adsorption chromatography depends largely on the degree of polarity of the analyte, where the more polar analytes interact to a greater degree with the silanol groups exposed on the surface of a normal stationary phase, thus retaining these compounds for longer. The analyte adsorbed to the stationary phase will then be displaced by polar modifier molecules that are added to the mobile phase to facilitate movement of the analyte through the chromatographic system and the analyte passes through the chromatographic system through a series of

adsorption and release steps. Reversed stationary phases retain non polar analytes to a greater degree.

A.2.2.2 Liquid – Liquid Partition Chromatography

Unlike LSC which uses solid particles as the stationary phase, liquid – liquid partition chromatography (LLC) uses a liquid as the stationary phase where the liquid is either adsorbed or chemically bonded onto an inert support such as glass. LLC is similar to solvent extraction techniques and solvent extraction data can be used to predict LLC partition coefficients. In solvent extraction, when a mixture of analytes is added into two immiscible solvents such as water (aqueous) and hexane (organic phase) some analytes will settle in the aqueous phase and some in the organic phase. The concentration of the analytes in each phase is dependent on the solubility of the individual analyte within each phase.

The same type of partitioning occurs in LLC. When a mixture of analytes is introduced into an LLC system, they will distribute themselves between the two immiscible liquid phases. Separation in LLC relies on the partition coefficient of the analyte which is given as the ratio of the concentration of the analyte in the stationary phase to its concentration in the mobile phase as in Equation A.2. Thus,

$$K = \frac{C\ stationary}{C\ mobile} \qquad \text{..... equation A.2}$$

Where,

K = partition coefficient

C stationary = concentration of analyte in the stationary phase

C mobile = concentration of analyte in the mobile phase

The difference in the partition coefficients of the analyte determines their separation. An analyte having high partition coefficient will reside for a longer period in the stationary phase than those having a low partition coefficient and as such will be retained longer and eluted more slowly from the chromatographic system. As for LSC, both normal

phase and reversed phase chromatography is possible through modification of the stationary phase.

A.2.2.3 Ion Exchange Chromatography

In Ion Exchange Chromatography (IEC), the stationary phase is made up of ion-bearing solid particles known as an ion exchange resin and provides sites where the exchange of ions within the analytes can be facilitated. To necessitate the separation of anions and cations, two types of ion exchange resins are available. Cation exchange resins are made up of functional groups such as SO_3H (sulphonic) or COOH (carboxylic). Anion exchange resins contain functional groups such as tertiary $–CH_2\text{-}NR_2$ or quaternary $–CH_2\text{-}NR_3$- amines.

In IEC, separation is achieved when an electrolyte solution is brought into contact with the ion exchange resin. If for example, the resin is cation exchanger, a positively charged analyte will be attracted to and retained at the exchange site. The presence of ions in the mobile phase having the same charge as the analyte will create competition for the exchange site. The analyte ion retained at the exchange site will be displaced by the competing ion provided that the analyte ion is not strongly adsorbed to the exchange site. The exchange of ions at the exchange site is the basis of ion exchange chromatography. The retention and separation of analytes are highly dependent upon the pH and ionic strength.

A.2.2.4 Size Exclusion Chromatography

Size Exclusion Chromatography (SEC) is also known as Gel Chromatography, Gel Filtration or Gel Permeation Chromatography (GPC). In this type of chromatographic technique, the stationary phase or the column packing is made up of a porous polymer matrix where the pores on the surface are completely filled with the mobile phase. A number of column packings are available for example cross-linked polystyrene, used for the separation of organic polymers or polyacrylamide gels for the separation of water soluble materials.

The separation in SEC is based on the relative size of the molecule of the analytes. When a sample is introduced, analyte having large molecular size will pass through the stationary phase without being retained within the pores whereas analyte having small molecular size will be retained in the pores and will elute more slowly from the chromatographic system.

A.2.3 Thin Layer Chromatography (TLC)

Thin Layer Chromatography (TLC) is a chromatographic technique where the separation of components within an analyte mixture occurs on a finely divided thin layer of adsorbent and was originally suggested by Izmailov and Shraiber in 1938 [22]. The stationary phase is coated onto a rigid support such as glass, thin plastic or an aluminium sheet. Modern TLC has developed from extensive study by Stahl [22] who first explored the possibilities of techniques which could be used to produce uniform thin layers of various powdered materials.

A.2.3.1TLC Apparatus

The basic apparatus of TLC is the chromatographic plate (either purchased or prepared), a separation chamber and a solvent system which acts as the mobile phase. Figure A.2 shows the schematic of a typical TLC set up.

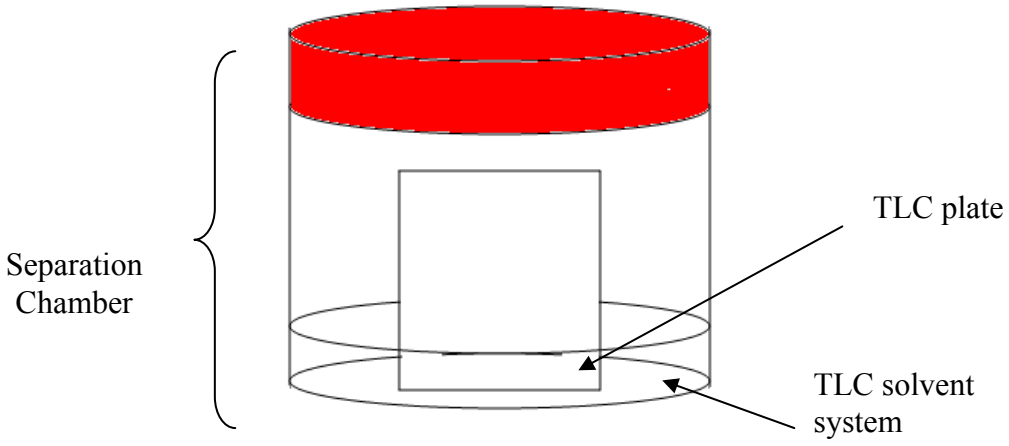


Figure A.2 Schematic of a TLC set up

A.2.3.2TLC Plate

The TLC plate is generally a sheet of glass, metal or plastic which is coated with a thin layer of the adsorbent. Although there are many types of adsorbent materials which can be used for TLC, the most commonly used adsorbents are silica gel, alumina, kieselguhr and cellulose [22]. In this study, TLC plates with silica gel as the adsorbent material are used. The dominant separation method is via adsorption.

A.2.3.3Solvent Systems

The solvent system will 'push' the analytes up the TLC plate therefore choice of solvent system is important as incomplete resolution, distortion or tailing of spots or bands can occur as a result of poor solvent choice [1]. In general, non-polar or low polarity solvent systems are used. Table A.1 shows the list of the eluotropic series of solvents which are arranged in decreasing order of polarity [1] which can be used as a guideline for TLC solvent system selection.

Table A.1 Eluotropic series of solvent in decreasing order (from top to bottom)

SOLVENT
Pure Water
Methanol
Ethanol
Propanol
Acetone
Ethyl Acetate
Diethyl ether
Chloroform
Dichloromethane
Benzene
Toluene
Trichloroethylene
Carbon Tetrachloride
Cyclohexane
Hexane

A.2.3.4Preparation of the TLC Plate

Prior to spotting the analyte mixture onto the TLC plate, a line is drawn in pencil parallel to the bottom of the plate and approximately 1 cm from the bottom. Depending on the type of development used, a predefined distance is then marked at the side of the plate. A gap of about 6 mm is normally left at each side of the plate. The parallel line is then marked with a dot using a pencil where the analyte mixture will be spotted using a micro-syringe or pulled capillary. If there are a number of sample mixtures to be analysed, the marks are made such that they will give enough space between the samples. Negative controls of the extracting solvent are also spotted onto the TLC plate. This is as shown in Figure A.3.

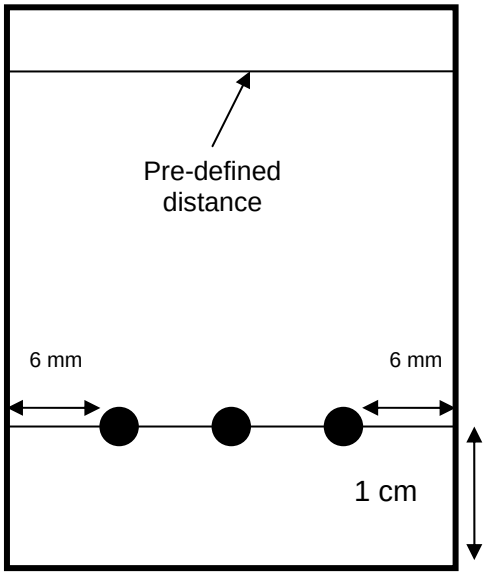


Figure A.3 Preparation of TLC plate

A.2.3.5 The TLC Separation Process

The TLC plate is then placed in a shallow pool of the mobile phase in the developing chamber so that only the very bottom of the plate is in contact with the mobile phase. The mobile phase raises up the TLC plate by capillary action carrying with it the analytes and by the interaction of the analytes with the adsorbent material, the sample

mixture can be resolved into normally a series of discrete spots on the plate. When the solvent has traversed to a predetermined distance, the plate is removed from the tank and excess mobile phase is allowed to evaporate.

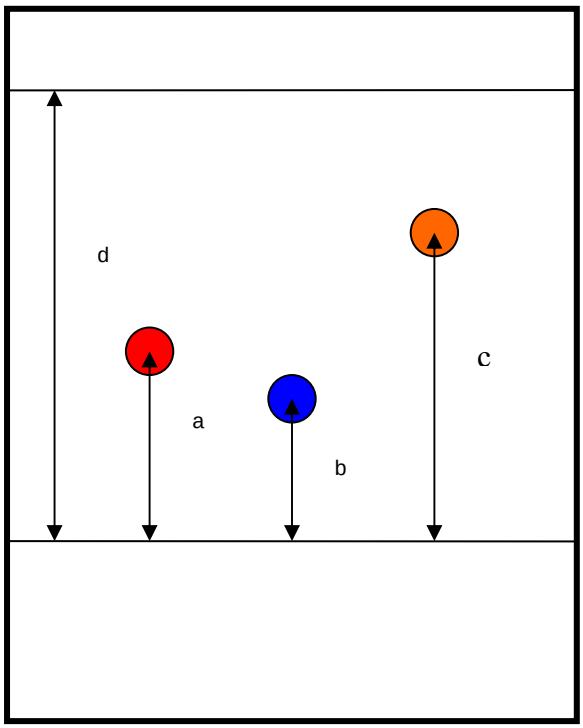
A.2.3.6Visualisation and Documentation

The simplest visualisation is by the naked eye however this is only suitable for coloured spots. In some cases where the spots cannot be seen, specialised visualisation techniques are used, involving spray reagents or observing the plate under UV lamps or by a combination of these techniques. Once the spots have been located they are marked by circling them using a pencil.

A.2.3.7Retardation Factor

The R_f value or the retardation factor is simply the distance moved by the analyte spot relative to the distance moved by the solvent as given in the Equation A.3. The R_f value is not a definite value of a component since two different components can have similar R_f values however it can aid in the identification of the analytes. Apart from the R_f value, the colour of the analyte spot and its characteristics under the UV lamps i.e. fluorescence or absorbance can also be used in its identification. An example of calculation for R_f is presented in Figure A.4.

$$\text{Retardation factor} = \frac{\text{distance of sample spot from origin}}{\text{distance of solvent front from origin}} \quad \text{.....equation A.3}$$



$$Rf_a = \frac{a}{d}$$

$$Rf_b = \frac{b}{d}$$

$$Rf_c = \frac{c}{d}$$

Figure A.4 Example of calculation of retardation factors, d is the distance traversed by the solvent front from the origin whilst a, b and c are the distances traversed by the spots from the origin.

A.2.4 Gas Chromatography

In 1952, Martin and James developed Gas Chromatography (GC) nearly 50 years after Tswett's development of chromatographic methods in 1903 [1]. GC evolved from earlier works on the adsorption of gases on various materials and Martin and James refined these earlier works to develop Gas Liquid Chromatography (GLC), introducing a new perspective to sample separation into analytical chemistry.

A.2.4.1GC Instrumentation

Gas chromatography has particular application in the separation of volatile organic compounds. These compounds are separated due to their differences in partitioning behaviour between the mobile and stationary phase within the analytical column. A typical GC comprises a carrier gas supply, sample introduction or injection system,

oven, chromatographic column, detector and data processing system or integrator. The schematic of a typical gas chromatograph is illustrated in Figure A.5.

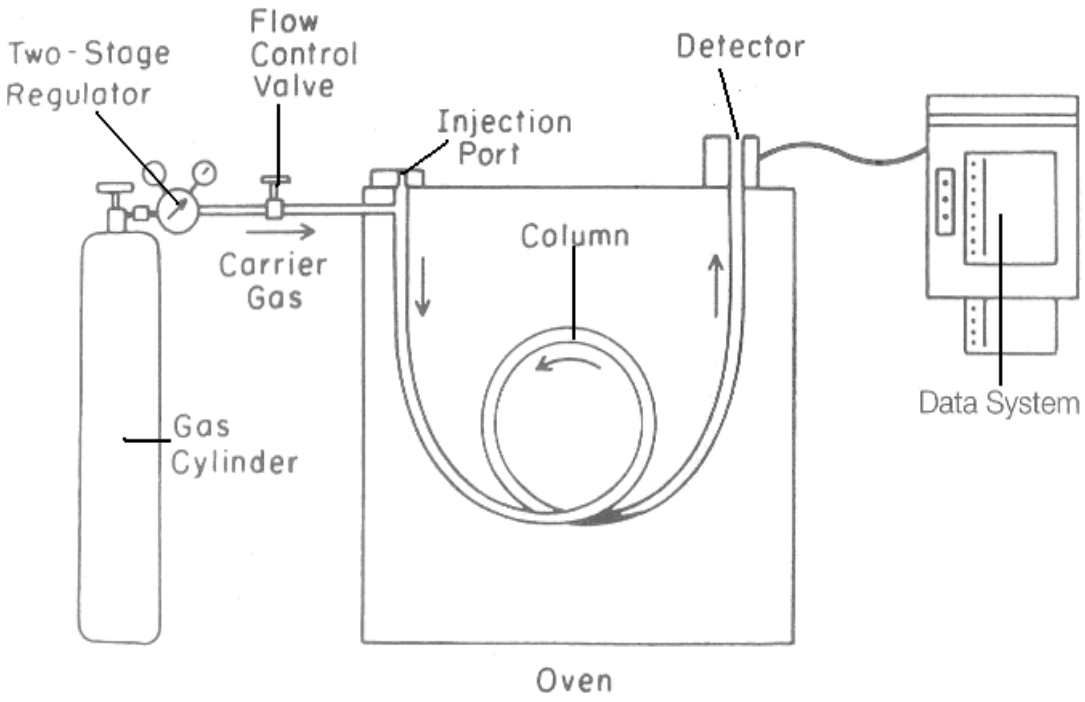


Figure A.5 Schematic of a gas chromatograph. Reproduced from McNair [5]

A.2.4.2 Carrier Gas Supply and Control

The carrier gas, functions as the mobile phase and transports the sample components from the injection port into the column and finally to the detector. The choice of carrier gas affects the column separation processes and the detector performance and has to be inert such that it does not react with the analytes or the column materials. The most common carrier gases are helium, argon and nitrogen.

High purity carrier gas is normally used, since impurities such as oxygen and water can chemically attack the liquid phase in the column and damage it [5]. Trace amounts of water can also desorb other column contaminants and cause high detector background and "ghost peaks". Trace hydrocarbons present in a carrier gas can cause high detector

background especially with ionisation detectors [5]. Nitrogen is normally used as a carrier gas mobile phase with Flame Ionisation Detector (FID). To ensure a uniform gas pressure to the column inlet, a pressure regulator and flow controllers are used [1].

A.2.4.3Injection Block

Samples can be introduced into GC in the form of gas or liquid. The former is normally introduced using a gas tight syringe or gas sampling valve and the latter is introduced normally by a micro-syringe. Capillary columns require strict sample introduction where the injection profile is narrow and the quantity is small, typically less than one microgram on column [5]. This is because the capillary columns only contain a small amount of stationary phase and as such only a small volume of sample is necessary.

A number of injection techniques have been devised for capillary columns but the most commonly used technique is a split injection. Figure A.6 shows the cross section of a typical split injector. Normally 1 μL of sample is injected through a heated injection port where it vapourises. Only a small percentage of the vapour enters the column while the rest of the vapour passes out of the injector through a split vent. The injection can also be made in splitless mode where the entire sample is introduced to the analytical column.

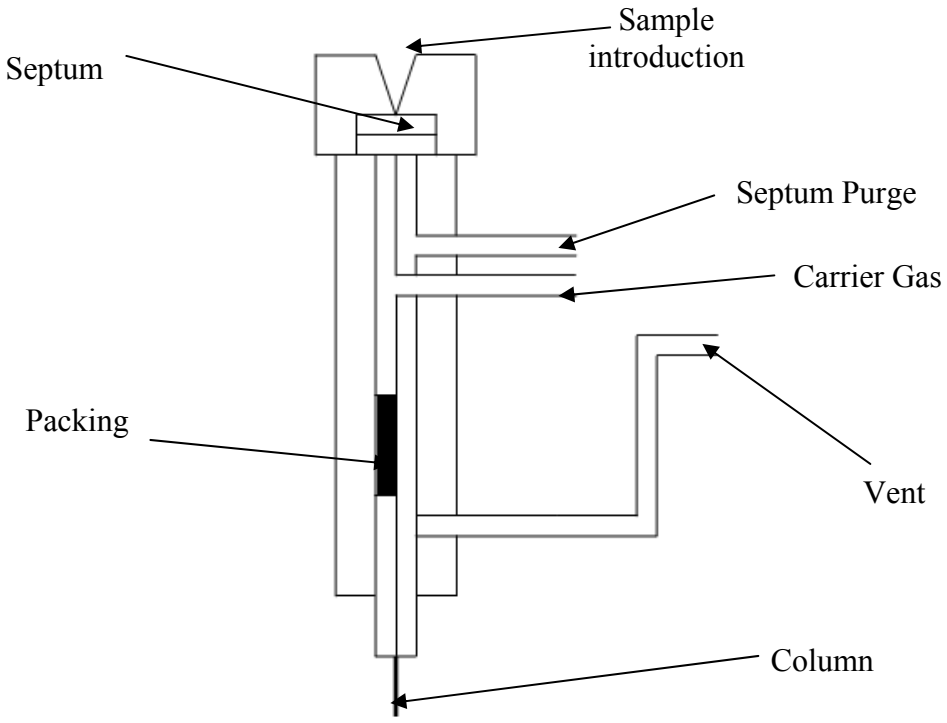


Figure A.6 Typical split injector

A.2.4.4 GC Oven

A GC oven within which the analytical column is situated is essentially a box equipped with a heating element which distributes heat to the entire oven by means of a fan. Temperature sensors are normally positioned at different locations in the oven to regulate the temperature. The output from these sensors is compared with the expected temperature and adjustments are made by means of microprocessors to the heating element.

A.2.4.5Chromatographic Column

The chromatographic column can be made from metal, glass, fused silica or plastic where fused silica is the most common. The column contains the stationary phase. In general there are two types of chromatographic column available which are either

packed columns or capillary columns. The former is normally 1 - 3 m in length with 2 - 6 mm internal diameter (i.d) and the latter is 10 - 100 m long with 0.2 - 0.5 mm i.d.

According to McNair *et al.* [5], the first commercial GC instruments used packed columns and as a result, many of the early literature reports only packed column separations. Capillary columns were introduced in 1959 as a consequence of the work of Golay [23], but were only used widely from about 1980.

There are a number of types of capillary columns available which are wall-coated open tubular capillary column (WCOT), porous-layer open tubular (PLOT), support-coated open tubular (SCOT) and micro-packed capillary columns. In WCOT, a thin layer of polymer stationary phase is coated on to the wall of the capillary tube. PLOT columns contain a porous layer of solid material such as alumina or porous polymer coated on to the capillary tube wall. In SCOT columns, a liquid stationary phase is coated on to particles of solid support materials where the latter is then coated on to the wall of the capillary tubes. Micro-packed columns are packed columns with the dimensions of capillary columns [23]. The WCOT column is the most common type of capillary column used at present [5].

A.2.2.3.5 Stationary Phase for Capillary Columns

Silicone polymers are the most commonly used stationary phases in GC. The polymer backbone has good flexibility which allows high diffusion rates of the solute molecules into the polymer [23]. They also have good thermal stability (300 to 350^{o}C), good film-forming properties, are resistive towards oxidation and can be synthesised to provide a range of selectivities [23].

Dimethylsilicone, whose structure is presented in Figure A.7, is the least polar silicone polymer whilst diphenylmethylsilicone, whose structure is presented in Figure A.8, is a medium polar silicone polymer. It should be noted that the polarity of silicone polymers change with the changes of organic groups within its structure.

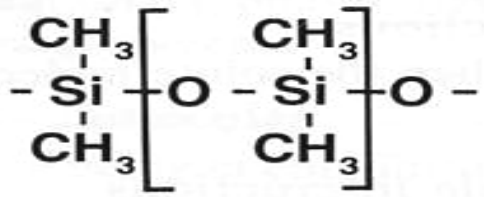


Figure A.7 Structure of Dimethylsilicone. Reproduced from McNair [5]

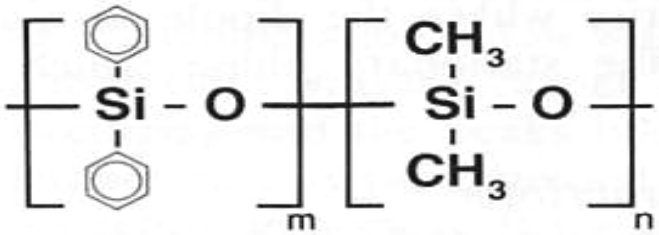


Figure A.8 Structure of Diphenylmethylsilicone. Reproduced from McNair [5]

Another stationary phase that is commonly encountered is polyethylene glycol, presented in Figure A.9, and is commonly used for the analysis of polar compounds.

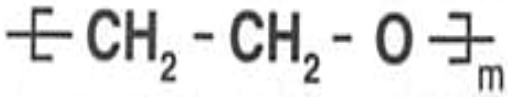


Figure A.9 Structure of polyethylene glycol. Reproduced from McNair [5]

A.2.4.6Detector

The detector measures the amount of the separated component as it elutes from the system. As mentioned earlier, there are many detectors that can be used in gas chromatography but for the purpose of this study, only the flame ionisation detector (FID) is discussed.

The FID is often considered as the most ideal and 'nearly universal' detector for GC and has been used in most routine and general purpose analysis [1]. One of the reasons for this is that it detects a wide variety of organic compounds over a wide linear range [5]. Figure A.10 shows a typical design of an FID. It consists of a small hydrogen-air flame burning at a small metal jet [1]. Hydrogen is introduced at the end of the column where it is mixed thoroughly with the column eluent before emerging at the jet into an air stream where it is ignited and produces ions. These ions are deflected to collector electrodes and produce a signal which is the detector output signal. The greater the component producing the ions, the greater the electric current and the greater the intensity of the signal produced.

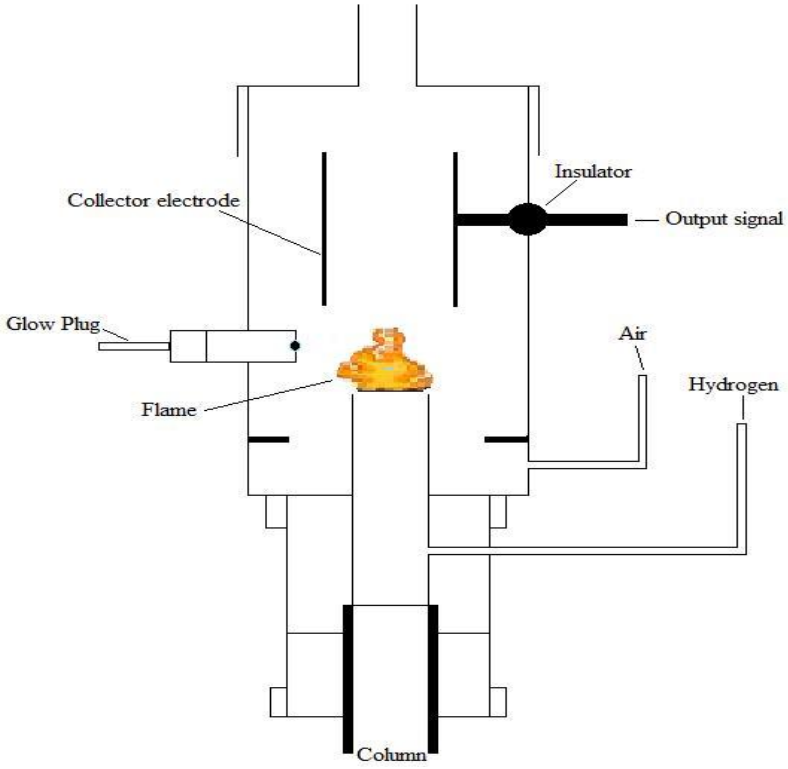


Figure A.10 A Typical design of an FID

A.2.4.7Sample Resolution in Gas Chromatography

In chromatography, the term resolution R_S is used to express the degree to which adjacent peaks are separated [5]. The resolution of two solutes, A and B is given in equation A.4;

$$R_S = \frac{2d}{(w_b)_A + (w_b)_B} \quad \text{.....equation A.4}$$

Where ***d*** is the distance between the peak maxima for two solutes A and B as depicted in Figure A.11

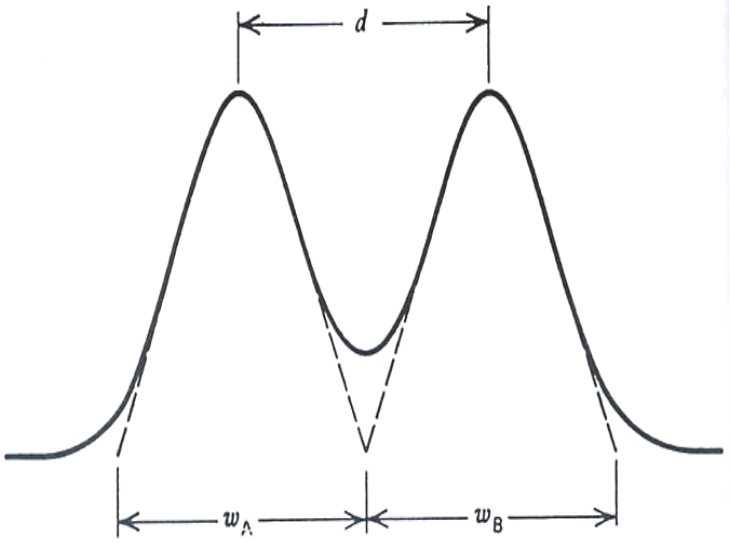


Figure A.11 Resolutions in GC. Reproduced from McNair [5]

Resolution can also be represented in term of retention time and efficiency, capacity factor and efficiency and capacity factor, efficiency and selectivity factor. The selectivity factor, ***α*** is given as in Equation A.5, which is the ratio of capacity factor of two retained peaks in the column.

$$\alpha = \frac{k_B}{k_A} \quad \text{..... equation A.5}$$

Resolution expressed in term of the capacity factor, efficiency and selectivity factor is given in Equation A.6.

$$R_S = \left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k'_B}{1 + k'_B}\right) \times \frac{\sqrt{N}}{4} \quad \text{..... equation A.6}$$

Where,

α is the selectivity factor, k is the capacity factor and N is the theoretical number of plates in the column.

A.3 Spectroscopy

Spectroscopy is defined as the study of matter and its properties by investigating light, sound or particles that are emitted, absorbed or scattered by the matter under investigation [24]. The electromagnetic spectrum displays the distribution of electromagnetic radiation which is a range of energy that is transmitted through space at various velocities. Electromagnetic radiation includes energy of varying wavelengths for example visible light, X–ray, ultra violet, radio and microwave radiation [25]. The various regions of the electromagnetic spectrum are shown in Figure A.12.

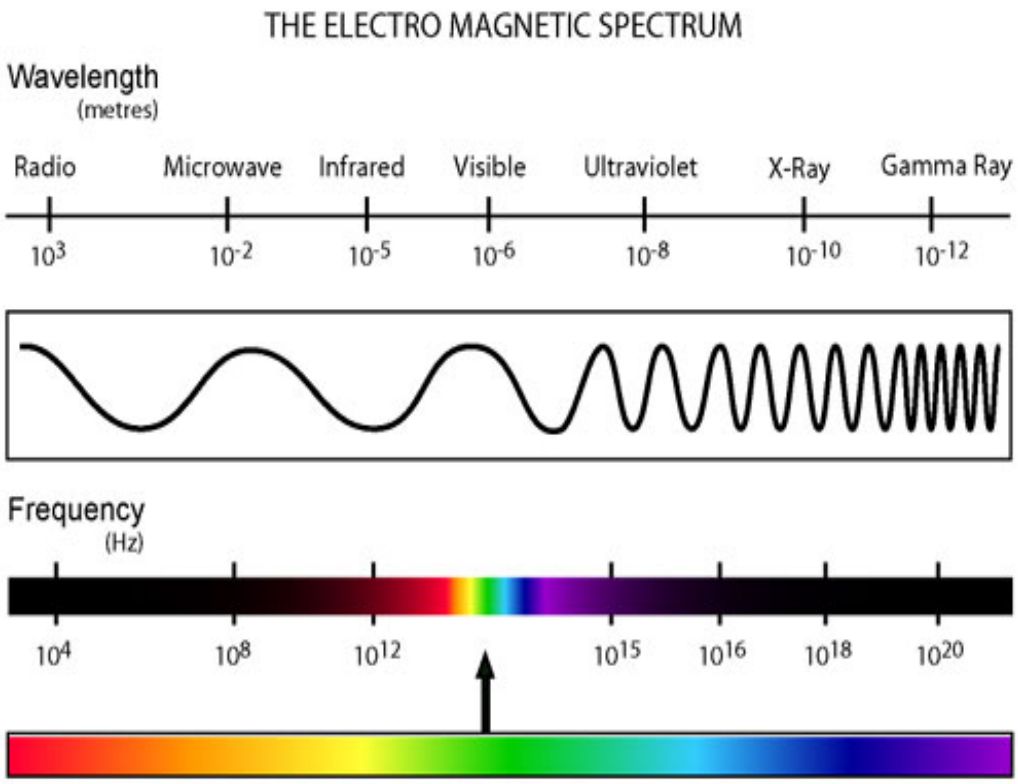


Figure A.12 The electromagnetic spectrum. Reproduced from Skoog [26]

There are a number of spectroscopic techniques available but for the purpose of this study only ultra violet and visible (UV/Vis) and microspectrophotometry (MSP) will be discussed.

A.3.1 UV/Vis Spectroscopy

As it can be seen, the ultra violet and visible (UV/Vis) regions correspond with absorption of light in the wavelength region from 100 – 390 and 400 – 800 nm of the electromagnetic spectrum respectively. The UV region can be sub-divided according to colour i.e. blue, yellow and red.

A.3.1.1 Molecular Absorption

Molecules undergo three types of quantised transitions which are rotational, vibrational, and electronic transitions. The relative energy levels of these transitions are in the order

of electronic > vibrational > rotational. The first two transitions i.e. rotational and vibrational involve absorption of microwave or far infra-red and near infra-red radiation respectively and are beyond the scope of this study.

An electronic transition occurs when a molecule absorbs energy from ultra violet and visible radiation, the consequence of which is to promote electrons residing in a low energy molecular or atomic orbital to a higher-energy orbital [25]. Such a transition of electrons between the two orbitals is known as an electronic transition and the absorption process is known as electronic absorption.

Figure A.13 shows the energy-level diagram illustrating the energy changes that occurs during absorption of infra-red, visible and ultra violet radiation respectively. E_0 in the figure represents the electronic ground state where E_1 and E_2 represent the electronic excited state.

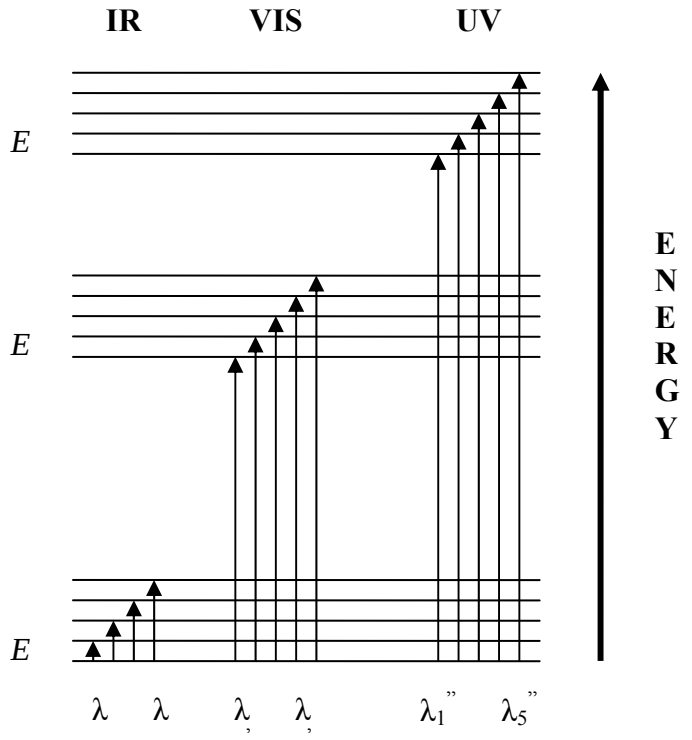


Figure A.13 Energy level diagrams. E_0, E_1 and E_2 are the electronic excited states. IR = infra red, VIS = visible and UV = ultra violet. The λ represents the wavelength.

A.3.1.2 Relaxation Mechanisms

The lifetime of excited states of molecules is short and the molecules will eventually lose their excitation energy and return to their ground state. Two of the mechanisms where excited molecules can give up their excess energy are non-radiative and fluorescent relaxation. In the non-radiative relaxation the excess energy is lost in the form of heat from collision of the excited molecules with the molecules of the solvent [25] whereas in the fluorescent relaxation, the energy is lost by the emission of a photon or light.

A.3.2 UV/Vis Instrumentation

A UV/Vis spectrophotometer consists of a source, or sources, of radiation, a monochromator for selecting wavelengths from the radiation sources, a facility for holding the sample i.e. the sample cell or cuvette and a detector for converting radiant energy into electrical current. Figure A.14 shows the block diagram of a typical UV/Vis spectrophotometer.

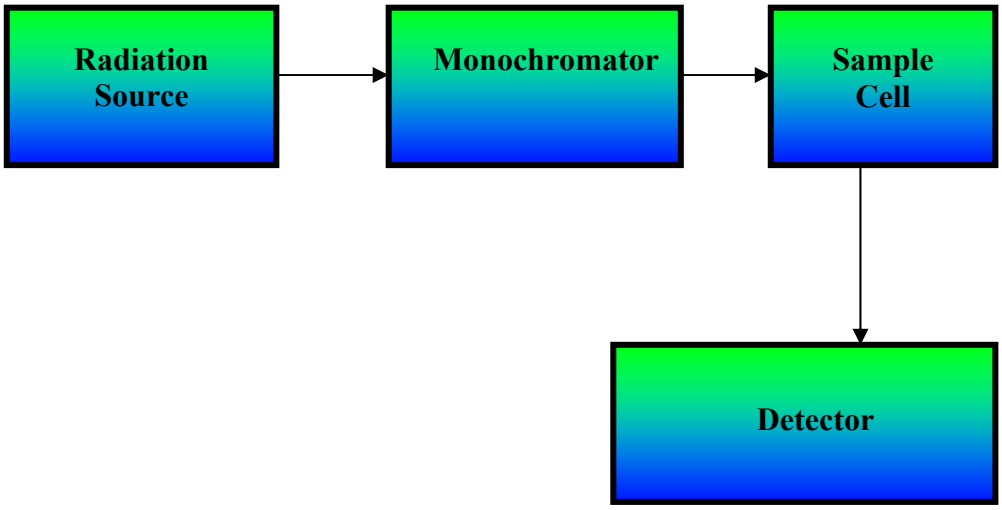


Figure A.14 Block diagram of a typical UV/Vis spectrophotometer

A.3.2.1 Radiation Sources

There are two radiation sources generally used in UV/Vis spectrophotometers which cover the required range of the ultra-violet and visible regions of the electromagnetic spectrum. These are deuterium lamps and tungsten lamps. The former emit electromagnetic radiation in the ultra-violet region while the latter source emits electromagnetic radiation in the visible region.

A.3.2.2 Monochromator

A monochromator consists of lenses or mirrors to focus the radiation, entrance and exit slits to restrict unwanted radiation and a dispersing medium to separate the wavelengths of the polychromatic radiation from the source [26]. In general, there are two types of monochromator, one that uses a reflection grating and the other one that uses a prism to disperse radiation into its component wavelengths [27].

A.3.2.3 Reflection Grating Based Monochromator

A diagram of the mechanism of a reflection grating-based monochromator is presented in Figure A.15 and consists of an entrance slit, a concave mirror, reflection grating and an exit slit. Polychromatic radiation i.e. radiation of more than one wavelength enters the monochromator through the entrance slit. The concave mirror collimates the radiation which produces a parallel beam which strike the reflection grating [27]. The beam is then split into its individual wavelengths where it is then focussed or collected onto another concave mirror. The individual wavelength exit the monochromator through the exit slit which is accomplished either by moving the grating or the exit slit.

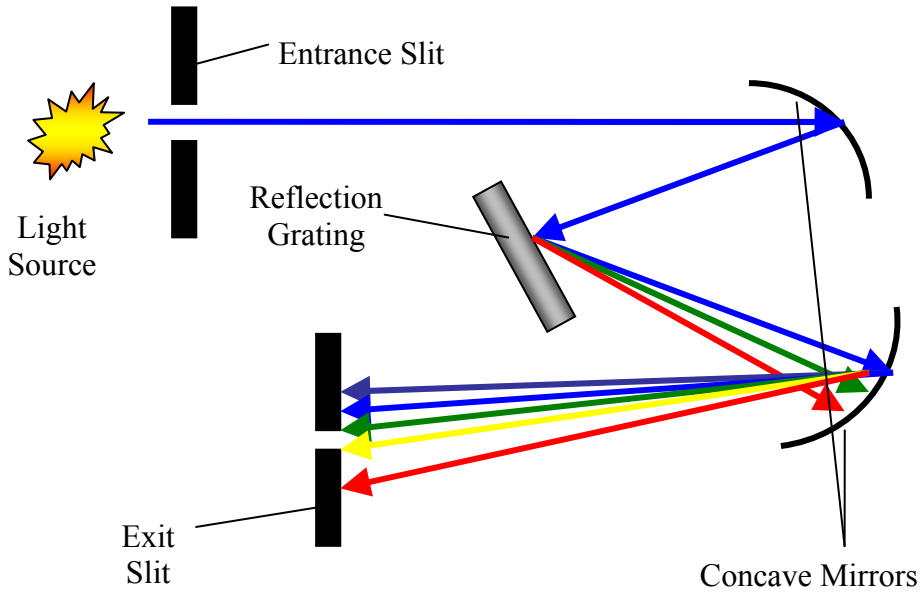


Figure A.15 Schematic of a reflection based monochromator

A.3.2.4Prism Based Monochromator

A typical prism-based monochromator is illustrated in Figure A.16. It consists of an entrance slit, collimating lens, a prism, focusing lens and an exit slit. Similar to the reflection grating-based monochromator, polychromatic radiation enters the monochromator through the entrance slit where it is collimated by the collimating lens into a parallel beam which enters the prism. When the beam passes through the prism, it is refracted into different wavelengths. Individual wavelengths are focused onto the focusing lens and leave the monochromator by the rotation of the prism or the movement of the exit slit.

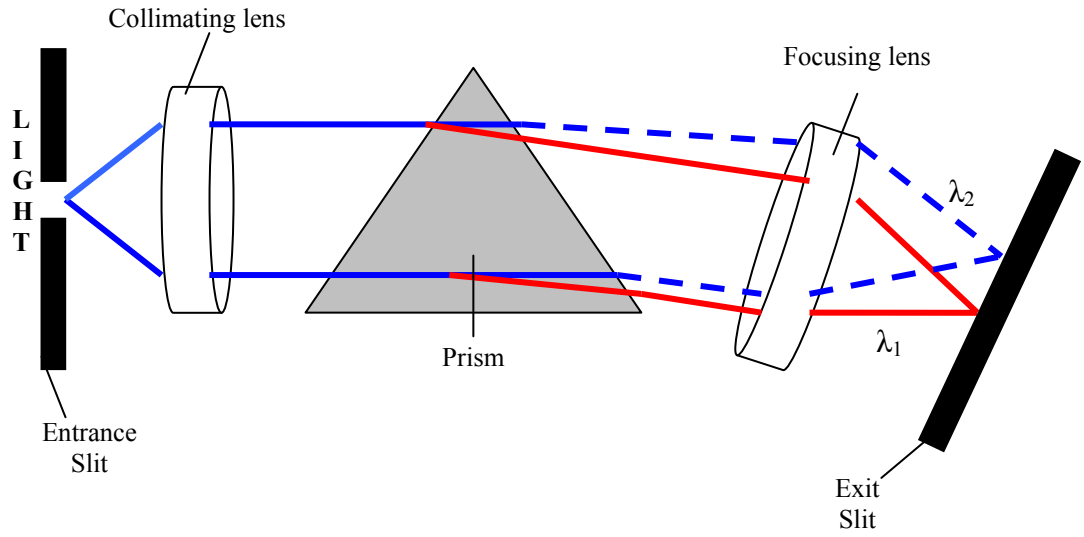


Figure A.16 Prism based monochromator

A.3.2.5 Sample Cell or Cuvette

The sample cell or cuvette is made from material which is transparent to the electromagnetic radiation being used. Quartz cuvettes are used for both the ultra-violet and visible regions while glass or plastic cuvettes are used for the visible region. Glass/plastic cuvettes are not suitable for the ultra-violet region because they absorb ultra-violet radiation. Although cuvettes of different volumes can be used, the most common for ultra-violet and visible is of 1 cm width [27].

A.3.2.6 Detector

There are a number of detectors available for UV/Vis spectrophotometers but for the purpose of this study only the photomultiplier (PMT) tube will be discussed. A typical PMT consists of a photo-emissive cathode, a photocathode, followed by focusing electrodes, an electron multiplier and an electron collector, (an anode in a vacuum tube), as illustrated in Figure A.17. When light enters and strikes the photocathode, the photocathode emits photoelectrons into the vacuum. These photoelectrons are then directed by the focusing electrode towards the electron multiplier where the electrons are

multiplied by the process known as secondary emission. The multiplied electrons are then collected by the anode as an output signal.

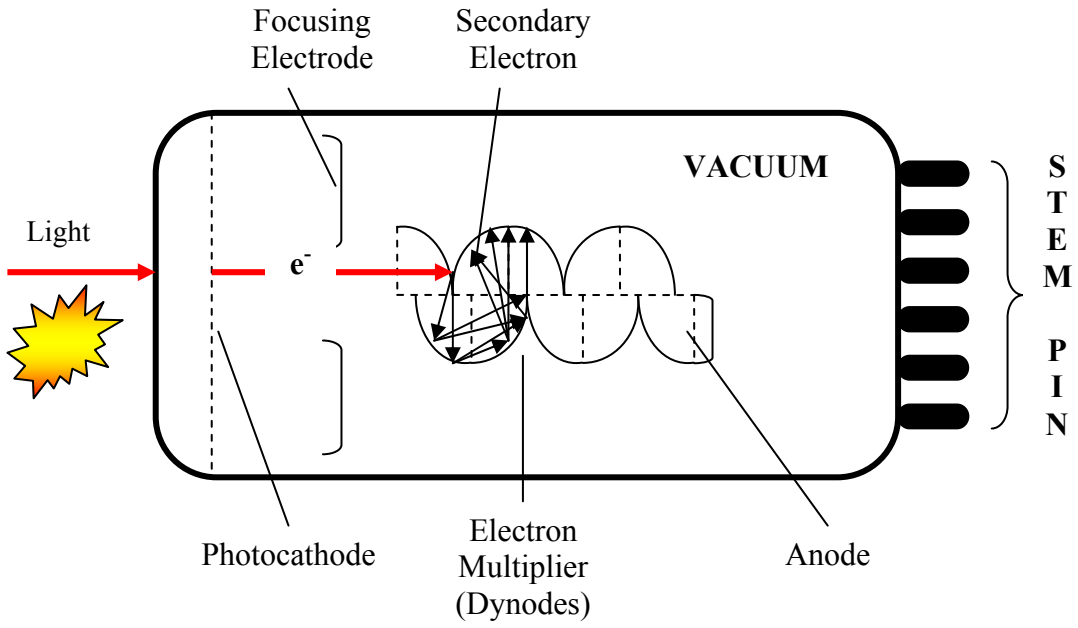


Figure A.17 A photomultiplier tube

A.3.2.7 Types of UV/Vis Spectrophotometer

There are two types of UV/Vis spectrophotometer commercially available which are single and double beam spectrophotometers. Although the set up of these spectrophotometers may vary between different manufacturers, their basic design follows that of the diagram presented in Figure A.14.

A typical arrangement of a single beam spectrophotometer is presented in Figure A.18. This type of spectrophotometer operates in a single beam sequence which means that the reference sample is measured separately from the test sample. The spectrophotometer has the advantages of being robust, easy to operate and cheap to maintain.

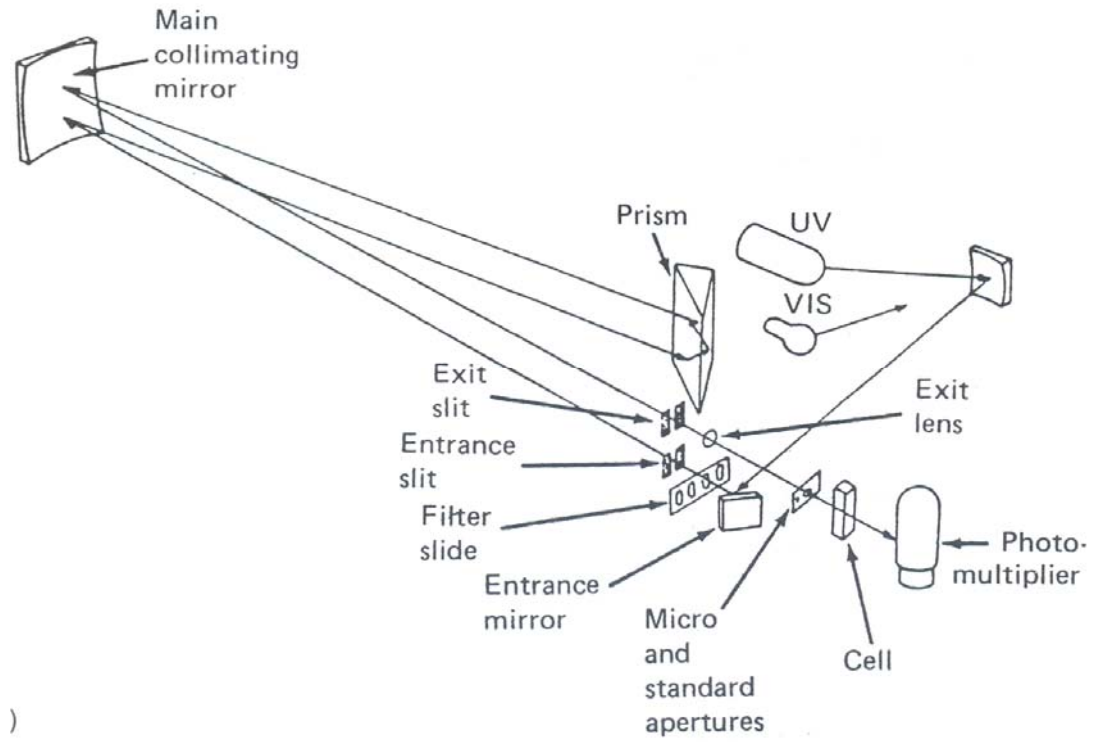


Figure A.18 Typical arrangement of a single beam spectrophotometer. Reproduced from Clark [28]

A double beam spectrophotometer is illustrated in Figure A.19. In this case monochromatic light passes alternately into the sample and the reference cells. This is achieved by the action of a motor which rotates a chopper into and out of the light path. When the chopper is placed in between the monochromator and the sample cell, the light is directed to another mirror, which reflects the light into the reference cell. When the chopper is rotated out of the light path, the light then travels into the sample cell. This action enables simultaneous measurement of the sample and reference.

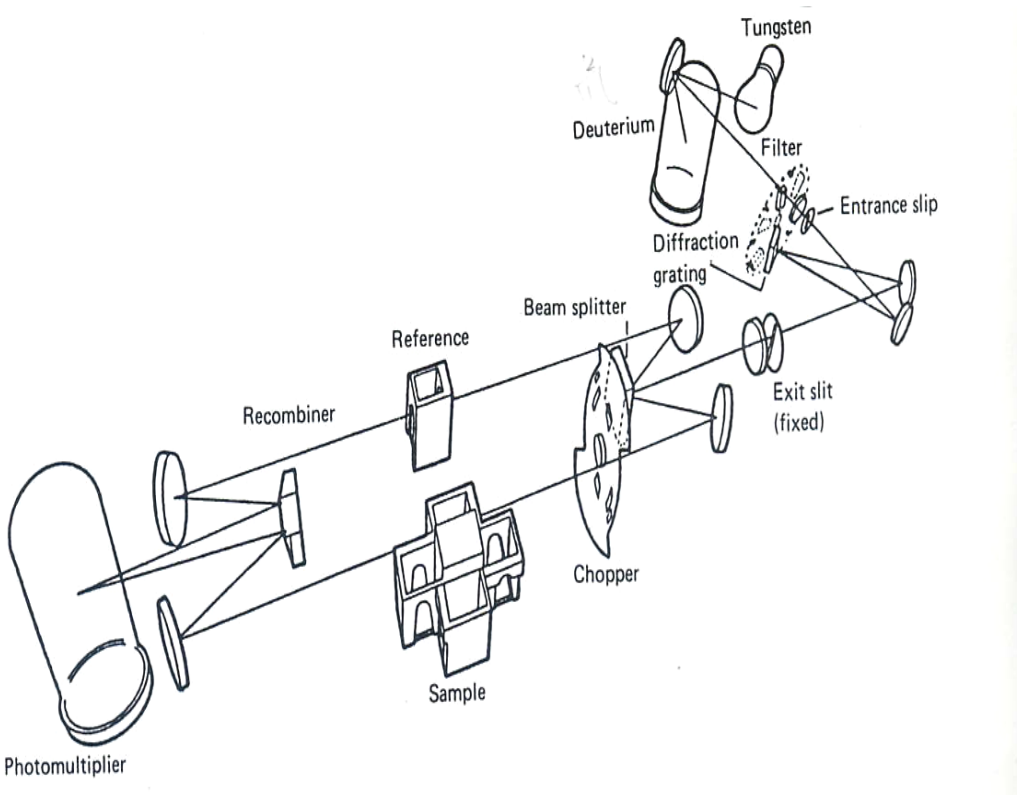


Figure A.19 Typical arrangement of a double beam spectrophotometer Reproduced from Clark [28].

A.3.3 Microspectrophotometry

Microspectrophotometry is a combination of two techniques, optical microscopy and spectrophotometry [29] and as such a microspectrophotometer is a combination of an optical microscope and a spectrophotometer. The first microspectrophotometers were built in the 1950s mainly by individual laboratories for the purpose of measuring the absorption spectra of coloured fibres [29]. Depending on the type of the electromagnetic radiation or light energy used, The MSP can be used can be involve the use of infra-red (IR) and/or ultra violet and visible (UV/Vis) microspectrophotometry.

The IR microspectrophotometry is mainly used to gather information about the composition and structure within a material while UV/Vis microspectrophotometry is mainly used to study the colour of very small materials for examples a single fibre, paint flakes or ink lines. Of the two microspectrophotometry techniques, the latter is of interest in this study.

A.3.3 Basic Operation of a Microspectrophotometer

A microspectrophotometer can operate in at least three different modes which are transmittance, reflectance, and emission.

In the transmittance mode, light is focussed onto the sample by the microscope condenser and the transmitted light (from the sample) is collected by the microscope objective. In the reflectance mode, light is focussed onto sample by the objective and the reflected light (from the sample) is collected by the same objective. While the transmittance and reflectance modes measure light transmitted and reflected from the sample, the emission mode measures light emitted from the sample after the sample has been 'stimulated' in some manner where the emitted light is collected by the microscope objective. An example of this is the fluorescence phenomenon.

In each case, a holographic grating in the spectrophotometer separates the light into its component wavelength where it is then focussed onto the detector. The detector measures the intensity of each individual wavelength and the outcomes are displayed as a spectrum. Figure A.20 – A.22 depict the transmittance, reflectance and emission mode respectively.

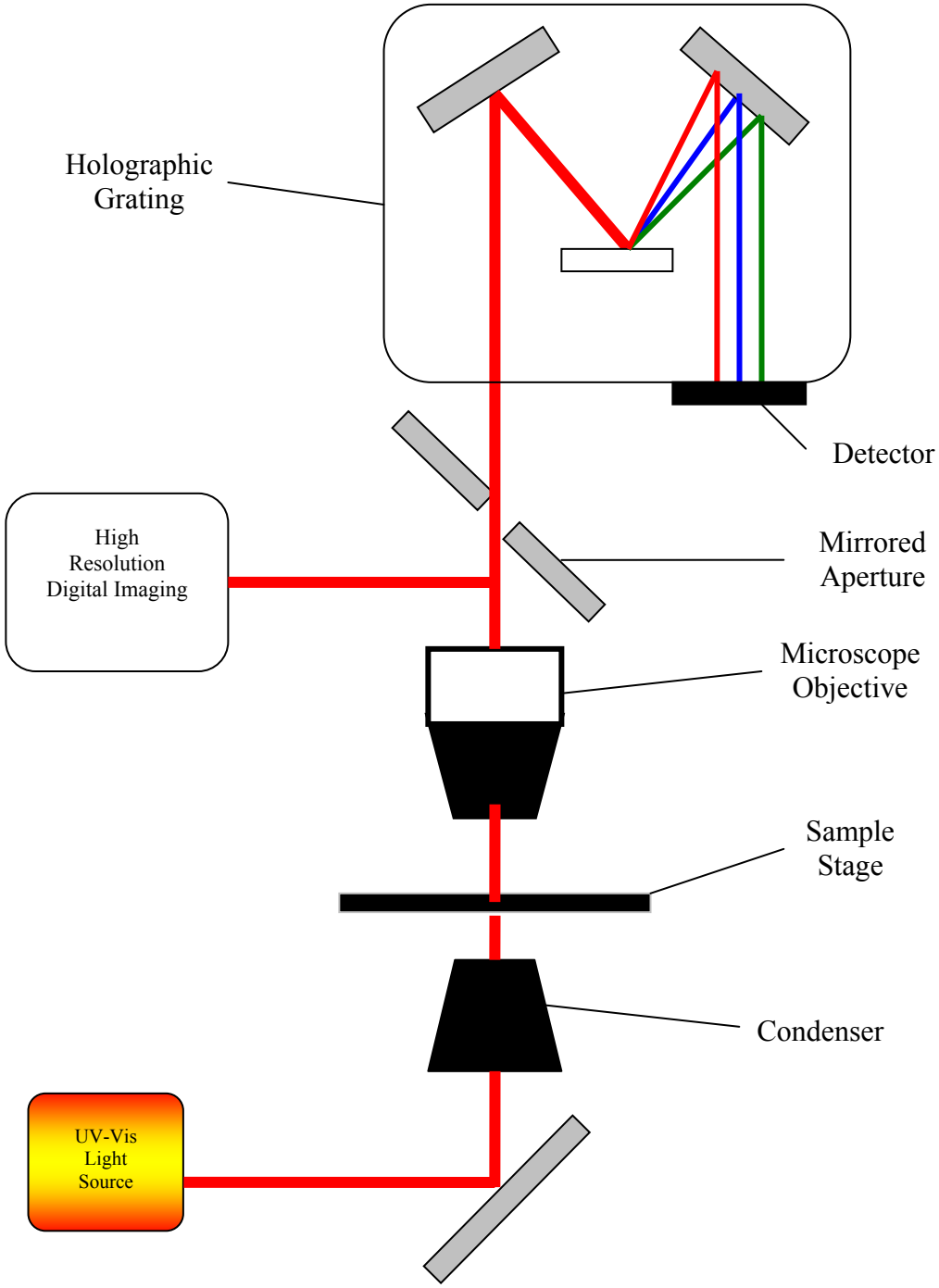


Figure A.20 Microspectrophotometry transmission modes

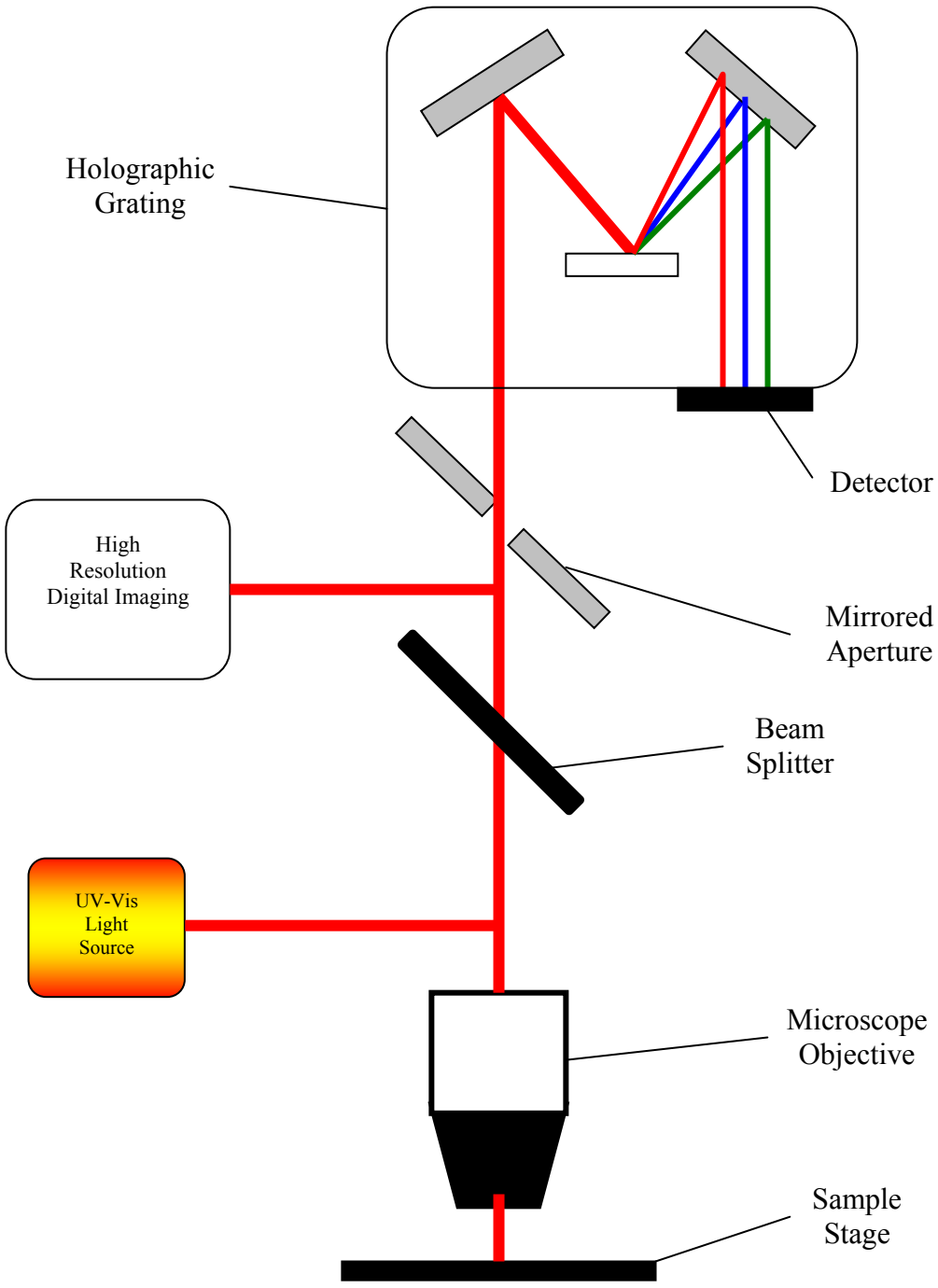


Figure A.21 Microspectrophotometry reflectance modes

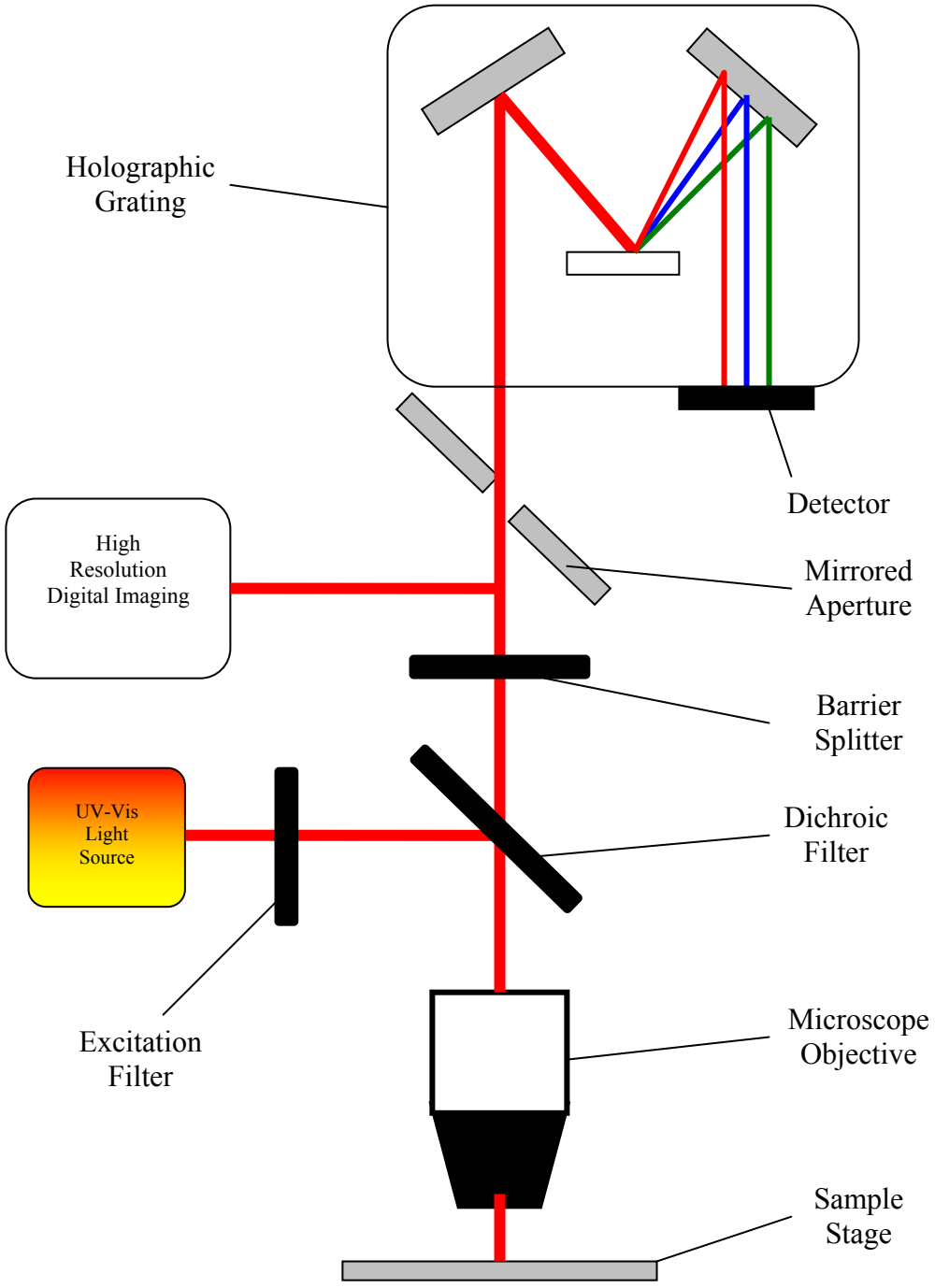


Figure A.22 Microspectrophotometry emission modes

A.4 REFERENCES

1. Braithwaite, A. and Smith, F.J., *Chromatographic Methods*. 1990, London: Chapman and Hall.

2. Stock, R. and Rice, C.B.F., *Chromatographic Methods*. 1974, London: Chapman and Hall.

3. Engelhardt, H., *One Century of Liquid Chromatography From Tswett's Columns to Modern High Speed and High Performance Separations.* Journal of Chromatography B, 2004. **800**: p. 3 - 6.

4. Hill, D.W., *The Application of Gas Chromatography to Forensic Science.* Journal of Forensic Science Society, 1961. **2**(1): p. 32 - 39.

5. McNair, H.M. and Miller, J.M., *Basic Gas Chromatography*. 1997, New York: John Wiley and Sons Incorporated.

6. Cole, M.D. and Caddy, B., *The Analysis of Drugs of Abuse an Instruction Manual*. 1995, Hertfordshire: Ellis Horwood Limited.

7. Blackledge, R.D. and Iwan, M., *Differentiation between Inks of the Same Brand by Infrared Luminescence Photography of their Thin Layer Chromatograms.* Forensic Science International, 1983. **21**: p. 165 - 173.

8. Jasuja, O.P., Singla, A.K. and Seema, B.L., *Thin Layer Chromatographic Analysis of Indian Stamp Pad Inks.* Forensic Science International, 1989. **42**(255 - 262).

9. Beattie, I.B., Roberts, H.L. and Dudley, R.J., *Thin Layer Chromatography of Dyes Extracted from Polyester, Nylon and Polyacrylonitrile Fibres.* Forensic Science International, 1981. **17**: p. 57 - 69.

10. Laing, D.K., Hartshorne, A.W. and Bennett, D.C., *Thin Layer Chromatography of Azoic Dyes Extracted from Cotton Fibres.* Journal of Forensic Science Society, 1990. **30**: p. 309 - 315.

11. Chiarotti, M., Carnevale, A. and Giovanni, N.D., *Capillary Gas Chromatographic Analysis of Illicit Diamorphine Preparations.* Forensic Science International, 1983. **21**: p. 245 - 251.

12. Karkkainen, M., Sippola, E., Pikkarainen, A.L., Rautio, T. and Himberg, K., *Automated Gas Chromatographic Amphetamine Profiling.* Forensic Science International, 1994. **69**.

13. Bland, H.H., *Petrol, Paraffin and Arson.* Journal of Forensic Science Society, 1979. **19**: p. 81 - 86.

14. Rella, R., Sturaro, A., Parvoli, G. and Ferrara, D., *A Brush Fire Forensic Case.* Science and Justice, 2005. **45**(1): p. 29 - 34.

15. Elliott, S.P. and Hale, K.A., *Analysis of Etorphine in Post Mortem Samples by HPLC with Diode-Array Detection.* Forensic Science International, 1999. **101**: p. 9 - 16.

16. Yonemitsu, K., Koreeda, A., Kibayashi, K., Ng'walali, P., Mbonde, M., Kitinya, J. and Tsunenari, S., *HPLC Analysis of Anti-Malaria Agent, Chloroquine in Blood and Tissue from Forensic Autopsy Cases in Tanzania.* Legal Medicine, 2005. **7**: p. 113 - 116.

17. Law, B., Goddard, C.P., Japp, M. and Humphreys, I.J., *The Characterisation of Illicit Heroin by the Analysis of Impurities Using High Performance Liquid Chromatography.* Journal of Forensic Science Society, 1984. **24**: p. 561 - 567.

18. *Gas Chromatography. http://uga.edu/srel/AACES/GCtutorial/page1.html. Last accessed 21/05/2004.*

19. Wall, P.E., *Thin Layer Chromatography: A Modern Practical Approach*. 2005, Cambridge: Royal Society of Chemistry.

20. Meritt, W. and Settle, D., *Instrumental Methods of Analysis*. 6th Ed. 1981, New York: Van Nostrand Company.

21. Simpson, C.F., *Practical High Performance Liquid Chromatography*. 1976, London: Hayden and Son Limited.

22. Bobbitt, J.M., *Thin Layer Chromatography*. 1963, New York: Reinhold Book Corporation.

23. Baugh, P.J., *Gas Chromatography*. 1993, Oxford: Oxford University Press.

24. *Spectroscopy. http://www.technology.niagarac.on.ca/courses/phtn1333. Last accessed 23/12/2005.*

25. Thomas, M., *Ultraviolet and Visible Spectroscopy*. 1996, New York: John Wiley and Sons.

26. Skoog, D.A., West, D.M. and Holler, F.J., *Fundamental of Analytical Chemistry*. 7th Ed. 1996, New York: Saunders College Publishing.

27. Christian, G.D., *Analytical Chemistry*. 6th Ed. 2004, New Jersey: John Wiley and Sons Incorporated.

28. Clark, B.J., Frost, T. and Rusell, M.A., *UV Spectroscopy, Techniques, Instrumentation, Data Handling*. 1993, London: Chapman and Hall.

29. Palus, J.Z., in *Encyclopaedia of Analytical Chemistry: Applications, Theory and Instrumentation*, Meyers, R.A, Editor. 2000, New Jersey: John Wiley and Sons Ltd.

APPENDIX B

APPENDIX B: VALIDATION STUDIES

B.1 INTRODUCTION

This appendix describes the analysis of the wax based products investigated in this work. This included a variety of lipsticks, lip balms and shoe polishes. Examination was undertaken using conventional techniques which included visual inspection, TLC, MSP, UV/Vis and GC-FID.

B.2 Sample Collection and Preparation

In total sixty three (63) samples consisting of 21 lip balms, 17 shoe polishes and 25 lipsticks were analysed. These were obtained from various retail outlets in Glasgow, Scotland. All of the samples investigated as part of this study were given individual code and are listed in Table B.1 a-c.

Table B.1a List of lip balm (LB) samples analysed in the study.

Sample Name	Type	Colour	Code
Body Shop Born Lippy Strawberry	LB	Red	LB1
Body Shop Born Lippy Watermelon	LB	Light Red/Pink	LB2
Body Shop Born Lippy Passionberry	LB	Purple	LB3
Body Shop Born Lippy Mango Peach	LB	Yellow	LB4
Body Shop Born Lippy Raspberry	LB	Pink	LB5
Chap Stick Flava Craze Fruit Craze	LB	Light Red/Pink	LB6
Chap Stick Flava Craze Blue Crazeberry	LB	Light Blue	LB7
Chap Stick Flava Craze Grape Craze	LB	Purple	LB8
Superdrug Strawberry Lip Balm	LB	Red	LB9
Superdrug Vanilla Lip Balm	LB	Colourless	LB10
Lipmate Original Lip Balm	LB	Colourless	LB11
Lypsyl Original Lip Balm	LB	Colourless	LB12
ASDA Original Lip Balm	LB	Colourless	LB13
TESCO Medicated Lip Balm	LB	Colourless	LB14
NIVEA Lip Care Essential Lip Balm	LB	Colourless	LB15
Metholatum Soft Lips French Vanilla	LB	Colourless	LB16
Lypsyl Strawberry	LB	Light Pink	LB17
Vaseline Original	LB	Colourless	LB18
Vaseline Rosy Lips	LB	Red	LB19
Vaseline Aloe Vera	LB	Colourless	LB20
Vaseline Sun Protection	LB	Colourless	LB21

Table B.1b List of shoe polish samples (SP) analysed in the study.

Sample Name	Type	Colour	Code
Kiwi Shoe Polish Dark Tan	SP	Tan	SP1
Kiwi Shoe Polish Black	SP	Black	SP2
Kiwi Shoe Polish Blue	SP	Blue	SP3
Kiwi Shoe Polish Red	SP	Red	SP4
Granger's Shoe Polish Brown	SP	Brown	SP5
Granger's Shoe Polish Black	SP	Black	SP6
Cherry Blossom Shoe Polish Tan	SP	Tan	SP7
Safeway Shoe Polish Dark Tan	SP	Tan	SP8
ASDA Shoe Polish Brown	SP	Brown	SP9
ASDA Shoe Polish Neutral	SP	Colourless	SP10
Kiwi Shoe Polish Neutral	SP	Colourless	SP11
Granger's Shoe Polish Neutral	SP	Colourless	SP12
Punch Taupe Shoe Cream	SP	Light Black	SP13
Meltonian Shoe Cream Black	SP	Black	SP14
Cherry Blossom Iris Blue	SP	Blue	SP15
Meltonian Shoe Cream Navy Blue	SP	Blue	SP16
Tuxan Shoe Care	SP	Brown	SP17

Table B.1c List of lipstick (L) samples analysed in the study

Sample Name	Type	Colour	Code
Star Gazer Lipstick 133	L	Purple	L1
Star Gazer Lipstick 105	L	Blue	L2
Star Gazer Lipstick 110	L	Black	L3
Clinique Super Spice	L	Brown	L4
Jane Seymour Pineapple Pink	L	Orange	L5
Body Collection Crushed Rose	L	Brown	L6
RIMMEL London Rich Raisin	L	Brown	L7
Clinique Raspberry Glace	L	Brown	L8
Collection 2000 Advance Colour Cream	L	Brown	L9
Boots Mulberry	L	Light Brown	L10
Estee Lauder French Fig	L	Brown	L11
Amway Debut Peach	L	Orange	L12
Clinique Ginger Flower	L	Red	L13
Estee Lauder All Day Carol Melon	L	Red	L14
Dare Dolly	L	Orange	L15
Apricot Crush	L	Orange	L16
Nutmeg	L	Brown	L17
Red Carnation	L	Red	L18
Marie France Touch of Spice	L	Brown	L19
Revlon Moondrops	L	Orange	L20
TESCO Shade 12	L	Red	L21
TESCO Shade 3	L	Brown	L22
TESCO Shade 13	L	Dark Brown	L23
Max Factor Sunset Rose	L	Red	L24
Body Collection Mango Mood	L	Orange	L25

B.3 Materials and Methods

In each case, smears of the sample were prepared by rubbing the sample onto a clean white cotton cloth swatch (1 cm x 1 cm). The colour of each sample smear was first observed by eye. This was followed by extraction of the sample smear by solvent using an optimised procedure. Analysis was performed by TLC initially and each developed spot was then analysed by MSP using reflectance mode. The sample extracts were also analysed by UV/Vis and GC-FID. The repeatability of the extraction and analysis was established and verified in each case.

In order to establish and optimise the extraction and analytical systems, one sample from each sample set separated by type was selected. These were TESCO Shade 12 lipstick (lipstick L21), Vaseline Rosy Lips lip balm (lip balm LB19) and Kiwi Shoe Polish Black (shoe polish SP2). Lipstick L21 was chosen because of its intense red colour. The lip balm LB19 sample was selected as it was also red in colour and provided a good contrasting sample to the lipstick L21 sample. Shoe polish SP2 was selected because it was a common brand of shoe polish. In each case a smear of the material was prepared using even pressure to transfer the product onto a clean white cotton swatch.

Once the extraction and analytical procedures were optimised all of the remaining samples were analysed. Following this, a set of selected samples were weathered both naturally and artificially and subjected to the same extraction and analytical regime.

B.3.1 Optimisation of Sample Extraction

A number of solvents and extraction techniques, suggested in the literature, were evaluated in order to determine the optimum conditions for the efficient extraction of the samples from the white cotton matrix and are presented in Table B.2. Solvents used were of reagent grade unless stated otherwise. Reagents and materials were purchased from commercial suppliers. Acetone, dichloromethane, and tetrahydrofuran, were purchased from Sigma Aldrich, Germany. Chloroform was purchased from WVR

International. The sonicator was manufactured by Decon Ultrasonic, England and vortex from Fisons Whirlmixer was utilised.

Table B.2 Extraction solvents and techniques.

Parameter	Solvents/Method
Extraction Solvent	1. Acetone 2. Chloroform 3. Dicholoromethane (DCM) 4. Tetrahydrofuran (THF) 5. Acetone: Chloroform (50: 50 v/v)
Extraction Technique	1. 24 hour standing at room temperature 2. Sonicate for 5 minutes 3. Vortex for 5 minutes

Using a pair of clean tweezers, the white cotton cloth bearing the sample smear was placed into a 5 mL clean glass vial containing 2 mL of the extraction solvent. For each extracting solvent the vial was either left to stand for 24 hours at room temperature, sonicated for 5 minutes or vortexed for 5 minutes prior to spotting on TLC plate. Once the extraction process was completed, the white cotton cloth was removed from the vial using the tweezers and the sample extract was then spotted onto a TLC plate using a pulled capillary tube. Six extractions were preformed of each sample type using each solvent and extraction technique combination.

B.3.2 Optimisation of the TLC system

Three solvent systems, taken from the literature [1-6] and presented in Table B.3 were evaluated in order to determine the optimum conditions for effective TLC analysis for the samples under study. Methanol and hydrochloric acid were purchased from Sigma Aldrich, Germany. Ammonium hydroxide and n-butanol were purchased from BDH Chemical, Poole, England.

Table B.3 TLC solvent systems

Parameter	Solvent system
Solvent System	1. chloroform: methanol: distilled water (5: 1.5: 0.2 v/v/v) 2. ethyl acetate: methanol: ammonium hydroxide (5: 1: 1 v/v/v) 3. n-butanol: 2N hydrochloric acid (HCl) (11: 1 v/v)

TLC analyses were carried out on Silica Gel 60 plates (Merck). Sample extracts were spotted on to the TLC plate using a pulled capillary tube supplied by S. Murray and Co., Surrey, England.

A double development technique was used to facilitate the best separation of the samples. In the first development, the solvent front was allowed to traverse 1 cm from the origin before the plate was removed from the TLC chamber and dried. In the second development, the solvent front was allowed to traverse 2 cm from the origin before the plate was once again removed from the TLC chamber. In each development, the colour of spot (s) and its retardation factor (R_f) were recorded.

The time taken for elution to complete, the colour of any developed spots and their R_f value for each solvent system were recorded. All extracts were spotted in triplicate onto the plate and the extracting solvent was allowed to evaporate between applications for each sample. Six TLC plates were run for each solvent system. In order to rule out any interference from the background material, a blank extract of the white cotton only, was also spotted onto the TLC plate along with the various samples under test. Each of the test samples were subjected to each of the extraction techniques as previously described and developed using the double development technique using each of the three TLC solvent systems. This facilitated the optimisation of the TLC solvent system.

B.3.3 TLC Analysis - Variation Studies

B.3.3.1 TLC Development Variation

The influence of the variation of the TLC development performed on the same and different TLC plates where the samples derived from the same extracted wax based product were evaluated. Six samples of the same extracted material were spotted onto the same TLC plate and onto six separate TLC plates using the method described in section B.3.2 and illustrated in Figure B.1.

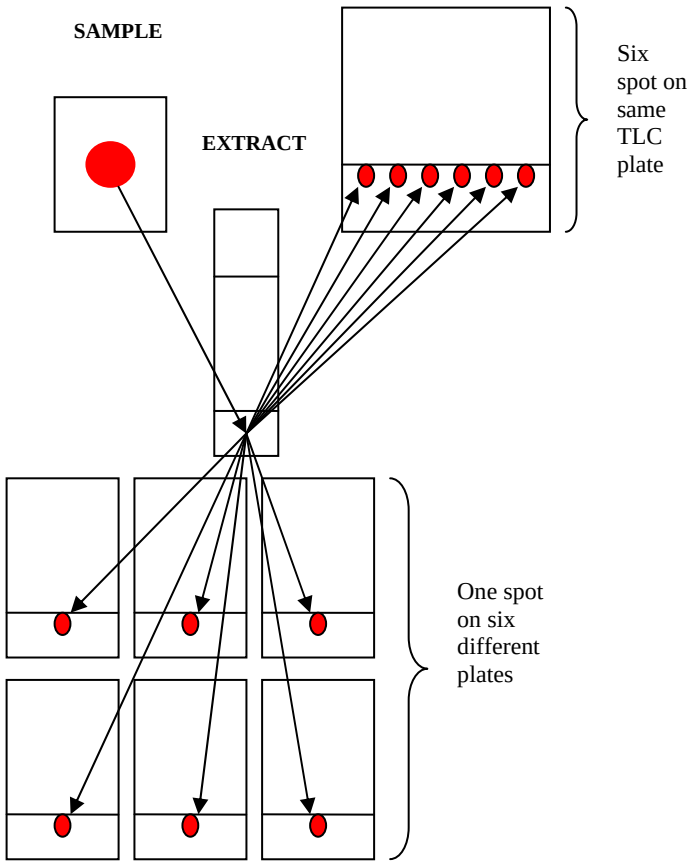


Figure B.1 Within TLC development variations on the same and different TLC plates

B.3.3.2 Within and Between Sample Variation

Six different areas of the same smear were extracted and each extract was spotted onto the same and onto six different TLC plates as described in section B.3.2. Wax based products are manufactured materials and as such little variation would be expected from different samples of the same brand of material. In order to assess the influence of between sample variations, six different samples of same brand were also examined. Smears of each of the six samples were extracted and analysed on the same and different TLC plates as previously described. The within and between sampling and analysis are illustrated in Figure B.2.

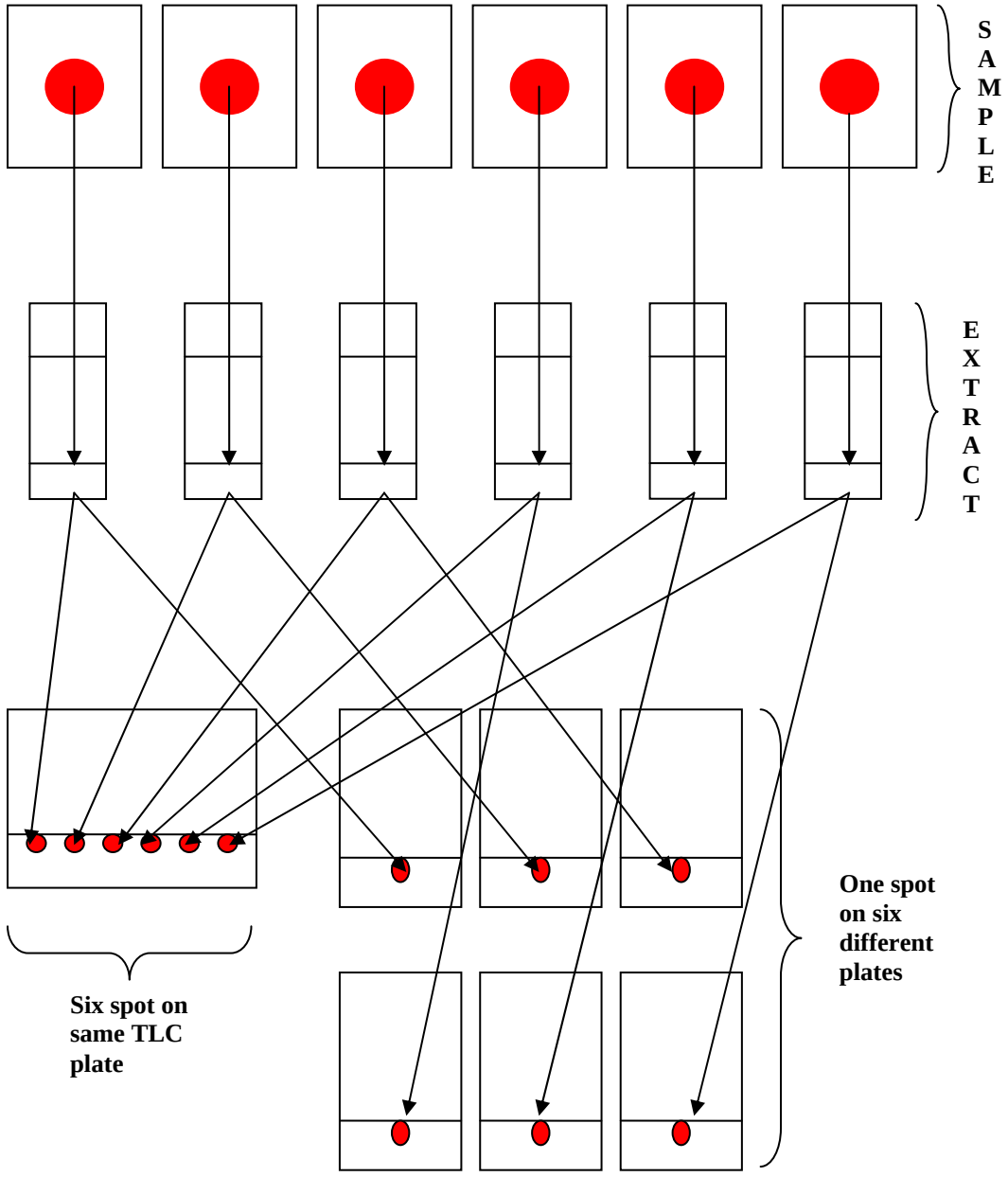


Figure B.2 Between-sample variations on the same and six different TLC plates. Note: For the within sample variation, samples are taken from a single sample but from different locations. For the between sample variations, samples are taken from six different samples of same brand.

B.3.3.3 TLC Analysis of the Full Sample Set

Once the optimum extraction and TLC solvent system had been selected the remaining wax based products were analysed using the selected system and the method described in section B.3.2.

B.3.4 Microspectrophotometry (MSP)

Microspectrophotometric measurements were made directly from spots observed on the TLC plates using reflectance mode. The reflectance spectra of the spots were obtained using a Cavendish Instruments microspectrophotometer equipped with an Olympus BX41 microscope which was linked to a computer and visualised using Onyx Version 1.9.0.0 software. Reflectance spectra acquisition settings were in the range of 362 – 780 nm with 1500 integration time. The developed TLC plate was placed directly onto the stage of the microscope and the reflectance beam focused on to the spot of interest on the plate using the ocular system of the MSP under X10 objective.

B.3.4.1 Instrumental Variation

The instrumental variation of the MSP was evaluated by recording six repeated measurements of the reflectance spectra on the same location of the same spot on a TLC plate as illustrated in Figure B.3.

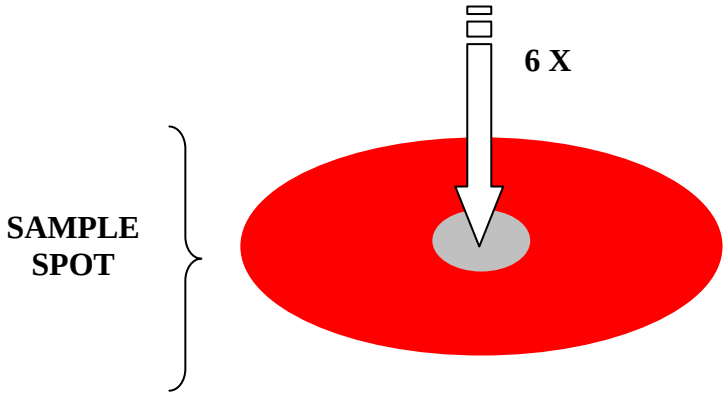


Figure B.3 Sample analyses for determining instrumental variation

B.3.4.2 Within and Between Spot Variation

Within spot variation was carried out by performing the MSP measurements at six different locations of same spot. MSP measurements were also obtained for each of six different spots of same sample on the same and on six different TLC plates. Figure B.4 illustrates the process.

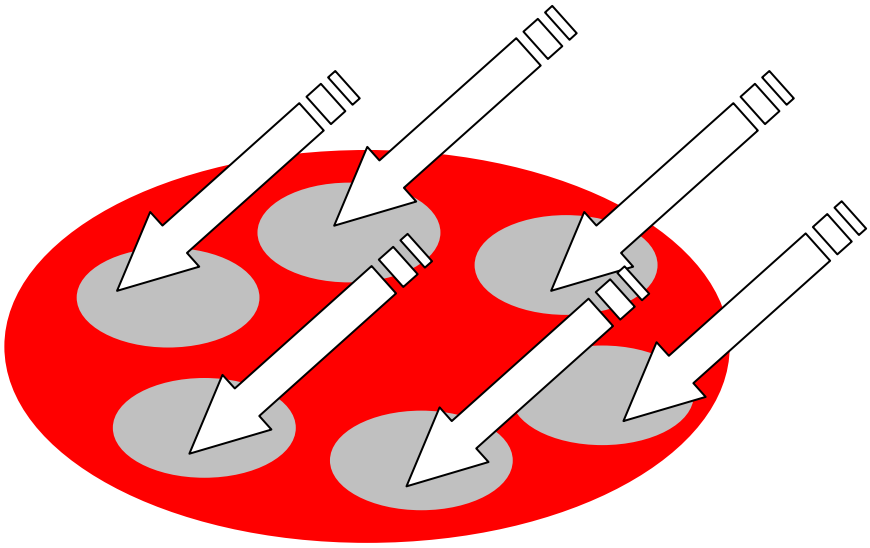


Figure B.4 Within spot variation for microspectrophotometric analysis

B.3.4.3 Within and Between Sample Variation

MSP measurements were obtained for each of six separate extractions of the same sample analysed on the same and different TLC plates. Measurements were also obtained for each of six separate extractions of the different samples of the same brand which were analysed on the same and different TLC plates.

B.3.5 UV/Vis Spectroscopy

Investigation of the UV/Vis spectroscopic characteristics of each extracted sample was undertaken. Reagents and materials were purchased from commercial suppliers as previously detailed. Quartz cuvette (10 mm), purchased from VWR International Limited were used for all measurements. UV/Vis spectra of all samples were obtained from the sample extracts using a UNICAM UV/Vis Spectrophotometer linked to a personal computer (PC) with Vision Scan V3.40 software. The scan range was set to be from 300 – 700 nm.

B.3.5.1 Instrumental Variation

The instrumental variation was assessed by undertaking six repetitive analysis of one extract from each of the test samples.

B.3.5.2 Within and Between Sample Variation

The variation in analysis of the extracted samples was assessed by performing the analysis on each of six separate extractions of the same sample and the analysis of six extracts from different samples of the same brand.

B.3.6 Gas Chromatography – Flame Ionisation Detection (GC-FID)

All GC analyses were performed using a Varian 3400 Gas Chromatograph with a flame ionisation detector (FID). The column was a ZEBRON ZB-1 100% dimethylpolysiloxane column supplied by J & W (30 m length x 0.32 mm inner diameter, 25 μm film thickness). A temperature program (modified from published methods [3-5]) was used as follows: 60^{o}C for 5 minutes, 7°C/min to 300^{o}C, hold at 300°C for 10 minutes. The injector and detector temperatures were set at 260°C and 310°C, respectively.

Nitrogen (BOC, Manchester) was used as the carrier gas at a constant column flow rate of 2 mL/min. 1 μL of the extract was injected in the split mode with a split ratio of 50:1. A D7500 integrator from Merck, Hitachi, Japan was used for data acquisition and

integration of the gas chromatograms. Reagents and materials were purchased from commercial suppliers as previously detailed.

B.3.6.1 Assessment of Column Performance

A mixture of acids, bases, alcohols, hydrocarbons and neutral compounds was first suggested by Grob *et al*. in 1978 as a single test mixture for the evaluation of capillary columns [7, 8]. The purpose of any capillary column test mix is to determine the ability of a given chromatographic system to effect separation of the analytes within the test mixture. Grob mixtures can also be used to monitor the performance and deterioration of a column during use. A modified Grob test mixture was prepared for this work using the components presented in Table B.4. All components were purchased from Sigma, Aldrich, Germany.

Table B.4 Hydrocarbons used in the Grob mixture used for this work.

Hydrocarbon	Name	Formula
C_{10}	Decane	$CH_3(CH_2)_8CH_3$
C_{10}-Fatty Acid Methyl Ester (FAME)	Methyl Decanoate	$C_{11}H_{22}O_2$
C_{12}	Dodecane	$CH_3(CH_2)_{10}CH_3$
C_{14}	Tetradecane	$CH_3(CH_2)_{12}CH_3$
C_{14}-Fatty Acid Methyl Ester (FAME)	Butyl Caproate	$C_{14}H_{28}O_2$
C_{16}	Hexadecane	$CH_3(CH_2)_{14}CH_3$
C_{18}	Octadecane	$CH_3(CH_2)_{16}CH_3$
C_{20}	Eicosane	$CH_3(CH_2)_{18}CH_3$
C_{22}	Docosane	$CH_3(CH_2)_{20}CH_3$
C_{24}	Tetracosane	$CH_3(CH_2)_{22}CH_3$

Each compound (20 mg) was weighed into a separate clean glass vial. Dichloromethane (10 mL) was added to each vial using a calibrated pipette. Aliquots (1 mL) of each sample were then combined together to give a Grob mixture of all ten components, each at a concentration of 0.2 mg/mL. 1 µL of the mixture was introduced to the GC during each injection.

B.3.6.2 Instrumental Precision

Instrumental precision was assessed based on the response for the internal standard decane (C_{10}) peak for six injections of one extract (0.2mg/mL). The relative standard deviation of the peak across these six injections was calculated. Similar to the assessment of instrumental precision, the repeatability of the chromatography was assessed by six replicate injections of the Grob mixture.

B.3.6.3 Peak Symmetry

Peak symmetry can be used as an indication of column efficiency. The asymmetry factor, *A*, can be calculated according to the following equation:

$$A = b/a \qquad \text{.....} \quad \textbf{equation B.1}$$

where *a* and *b* are the left and right halves of the peak width at 10% peak height. Measurements were made by hand for peak symmetry of the Grob mixture components.

B.3.6.4 Within and Between Sample Variation

The variation as a result of the extraction of the sample was evaluated by analysing each of six extractions of the same sample carried out according to the method detailed previously.

The variation between samples of the same brand was evaluated by analysing each of six extractions of different samples of the same brand.

B.4 Descriptive Statistics

The average, standard deviation (s) and percentage relative standard deviation (%RSD) were calculated using Equations B.2, B.3 and B.4 respectively.

$$Average = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{..... equation B.2}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - average)}{n-1}} \quad \text{..... equation B.3}$$

$$\%RSD = \frac{s}{Average} x100 \quad \text{..... equation B.4}$$

Where *n* is the number of elements or observations.

B.5 Effect of Ageing

The effect of weathering of the sample smears was studied by ageing selected samples according to condition employed by Cole [4] at room temperature under normal lighting conditions and at 56^{o}C in an oven.

Each sample was exposed to the conditions for periods of 14 and 28 days. The smears were then analysed by visual observation, TLC, microspectrophotometry, UV/Vis and GC-FID after each ageing period using the methods previously detailed.

B.6 Results and Discussion

B.6.1 Optimisation of Sample Extraction and TLC Solvent System

Colour in wax based products is imparted by a combination of dyes and pigments which are mixed together during the manufacturing process to produce the desired colour or

shade. While pigments are mostly made up of inorganic compounds, dyes are mostly organic compounds making them suitable to be extracted and analysed by thin layer chromatography (TLC). Like any other chromatographic technique, the objective of TLC is to 'achieve' separation of a mixture of compounds into individual analytes. TLC is a basic and easy to perform chromatographic technique and is suitable for rapid screening purposes. Griffin [9] suggested that if the sample smear was light or faint, it may be necessary to extract the whole area using an extraction solvent. They also suggested scraping heavy smears using a scalpel blade and dissolving the scrapings in an appropriate extraction solvent

One example of each of the types of samples under test was examined during the optimisation experiments. These were lipstick L21 (TESCO Shade 12), shoe polish SP2 (Kiwi Shoe Polish Black) and lip balm LB19 (Vaseline Rosy Lips). In each case the relevant sample was smeared onto a clean white cotton support and a portion of the stained cotton excised from this (1cm x 1cm). The excised portion of the cotton was placed in a clean glass vial and the relevant solvent (2 mL) and extraction method carried out.

Various solvents and solvent mixtures have been suggested for use in the extraction of wax based products [3- 9]. The abilities of these solvents together with dichloromethane (DCM), to extract the test samples, were assessed. Although chloroform and the mixture of acetone: chloroform worked very well, it was decided to choose DCM because the former extraction solvents are quite harsh on the stationary phase of the GC column compared to DCM. The sample was immersed in the relevant extracting solvent and three different agitation methods (standing for 24 hours, agitation using sonication and agitation using a vortex) were evaluated. TLC solvent systems previously reported in the work of Andrasko [1] and Jasuja [2] (solvent system 1 and 2), Cole [3, 4] and Byrne [5] (solvent system 1) and Russell [6] (solvent system 3) were all evaluated for each extraction solvent and technique.

The results of the various combinations of extraction solvent, extraction technique and TLC solvent system are presented in Tables B.5 to B.13. The following notation is used in all tables; R_f = retardation factor, L = lipstick, LB = lip balm, SP = shoe polish. The numbers in parentheses are the distances travelled by the solvent front. Six repetitive analyses were undertaken throughout and solvent and substrate blanks were negative in all cases.

Table B.5 TLC outcomes for 24 hours stand at room temperature with chloroform: methanol: distilled water (5: 1.5: 0.2 v/v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
24 Hours	DCM	L	-	-	-	-
		LB	-	-	-	-
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	-	-	-	-
		LB	-	-	-	-
		SP	0.40	purple	0.40	purple
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	-	-	-	-
		LB	-	-	-	-
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	-	-	-	-
		LB	-	-	-	-
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	-	-	-	-
		LB	-	-	-	-
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-

Table B.6 TLC outcomes for sonication for 5 minutes with chloroform: methanol: distilled water (5: 1.5: 0.2 v/v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
Sonicate	DCM	L	-	-	0.15	pink
		LB	0.81	pink	0.84	pink
		SP	0.50	purple	0.65	purple
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	-	-	-	-
		LB	-	-	-	-
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	-	-	0.10	pink
		LB	0.72	pink	0.74	pink
		SP	0.30	purple	0.55	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	-	-	-	-
		LB	-	-	-	-
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	-	-	0.15	pink
		LB	0.70	pink	0.72	pink
		SP	0.40	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-

Table B.7 TLC outcomes for vortex for 5 minutes with chloroform: methanol: distilled water (5: 1.5: 0.2 v/v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
Vortex	DCM	L	-	-	0.10	pink
		LB	0.80	pink	0.82	pink
		SP	0.40	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	-	-	-	-
		LB	-	-	-	-
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	-	-	0.15	pink
		LB	0.80	pink	0.80	pink
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	-	-	-	-
		LB	-	-	-	-
		SP	0.40	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	-	-	0.15	pink
		LB	0.78	pink	0.78	pink
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-

Table B.8 TLC outcomes for 24 hours stand at room temperature with ethyl acetate: methanol: ammonium hydroxide (11: 1: 1 v/v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
24 Hours	DCM	L	-	-	-	-
		LB	-	-	-	-
		SP	1.00	purple	1.00	purple
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	-	-	-	-
		LB	-	-	-	-
		SP	1.00	purple	1.00	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	-	-	-	-
		LB	-	-	-	-
		SP	1.00	purple	1.00	purple
		B	-	-	-	-
		BS	-	-	-	-

Table B.9 TLC outcomes for sonication for 5 minutes with ethyl acetate: methanol: ammonium hydroxide (11: 1: 1 v/v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
Sonicate	DCM	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	-	-	-	-
		LB	-	-	-	-
		SP	1.00	purple	1.00	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-

Table B.10 TLC outcomes for vortex for 5 minutes with ethyl acetate: methanol: ammonium hydroxide (11: 1: 1 v/v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
Vortex	DCM	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	-	-	-	-
		LB	-	-	-	-
		SP	-	-	-	-
		B	-	-	-	-
		BS	-	-	-	-

Table B.11 TLC outcomes for 24 hour stand at room temperature with n-Butanol: 2N hydrochloric acid (11: 1 v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
24 Hours	DCM	L	0.50	pink	0.50	pink
		LB	0.70	pink	0.70	pink
		SP	0.30	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	0.50	pink	0.50	pink
		LB	0.70	pink	0.72	pink
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	0.50	pink	0.45	pink
		LB	0.74	pink	0.76	pink
		SP	0.50	purple	0.45	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	-	-	-	-
		LB	-	-	-	-
		SP	0.60	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	0.60	pink	0.50	pink
		LB	0.80	pink	0.80	pink
		SP	0.40	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-

Table B.12 TLC outcomes for sonication for 5 minutes with n-Butanol: 2N hydrochloric acid (11: 1 v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
Sonicate	DCM	L	0.60	pink	0.65	pink
		LB	0.80	pink	0.85	pink
		SP	0.70	purple	0.70	purple
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	0.50	pink	0.50	pink
		LB	0.75	pink	0.75	pink
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	0.50	pink	0.60	pink
		LB	0.75	pink	0.75	pink
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	0.70	pink	0.75	pink
		LB	0.90	pink	0.90	pink
		SP	0.80	purple	0.80	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	0.50	pink	0.50	pink
		LB	0.75	pink	0.75	pink
		SP	0.50	pink	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-

Table B.13 TLC outcomes for vortex for 5 minutes with n-Butanol: 2N hydrochloric acid (11: 1 v/v) as the solvent system.

EXTRACTION TECHNIQUE	EXTRACTION SOLVENT	SAMPLE	R_F (1.0)	SPOT COLOUR	R_F (2.0)	SPOT COLOUR
Vortex	DCM	L	0.50	pink	0.50	pink
		LB	0.70	pink	0.70	pink
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	THF	L	0.60	pink	0.60	pink
		LB	0.75	pink	0.75	pink
		SP	0.50	purple	0.60	purple
		B	-	-	-	-
		BS	-	-	-	-
	Chloroform	L	0.70	pink	0.70	pink
		LB	0.90	pink	0.70	pink
		SP	0.80	purple	0.80	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone	L	-	-	-	-
		LB	-	-	-	-
		SP	0.50	purple	0.50	purple
		B	-	-	-	-
		BS	-	-	-	-
	Acetone/Chloroform	L	0.50	pink	0.50	pink
		LB	0.80	pink	0.85	pink
		SP	0.70	purple	0.70	purple
		B	-	-	-	-
		BS	-	-	-	-

The most effective separation mechanism involved sonication of the samples followed by TLC using the n-Butanol: 2N hydrochloric acid (11: 1 v/v) solvent system. However, this development system was found to be very slow and required 10 - 15 minutes to develop the samples to 1 cm. For this reason, it was decided to choose chloroform: methanol: distilled water (5: 1.5: 0.2 v/v/v) as the solvent system for running the rest of the samples as this was considerably faster in terms of sample elution time (4 – 5 minutes).

For ease of use and versatility of sample analysis the final extraction method chosen for the samples was:

Extraction with DCM using 5 minutes sonication and elution using chloroform: methanol: distilled water (5: 1.5: 0.2 v/v/v).

B.6.2 TLC Repeatability Studies

B.6.2.1 Variation of the TLC system

The variation between TLC plates for analysis of the same sample extract was evaluated. This represented a means to assess both the variation in the TLC process and the variation in the extraction process for a given sample. The results are presented in Tables B.14 to B.16 for samples run on the same TLC plate and Tables B.17 to B.19 for samples run on six different TLC plates.

Table B.14 TLC system variations for the lipstick (L21) sample analysed on the same TLC plate.

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
L21 (extract 1, spot 1)	-	-	0.13	pink
L21 (extract 1, spot 2)	-	-	0.14	pink
L21 (extract 1, spot 3)	-	-	0.14	pink
L21 (extract 1, spot 4)	-	-	0.14	pink
L21 (extract 1, spot 5)	-	-	0.13	pink
L21 (extract 1, spot 6)	-	-	0.13	pink
Average	**-**		**0.14**	
SD	**-**		**0.01**	
% RSD	**-**		**4.06**	

Table B.15 TLC system variation for the lip balm (LB19) sample analysed on the same TLC plate.

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
LB19 (extract 1, spot 1)	0.85	pink	0.83	pink
LB19 (extract 2, spot 2)	0.80	pink	0.80	pink
LB19 (extract 3, spot 3)	0.85	pink	0.83	pink
LB19 (extract 4, spot 4)	0.85	pink	0.86	pink
LB19 (extract 5, spot 5)	0.85	pink	0.83	pink
LB19 (extract 6, spot 6)	0.85	pink	0.88	pink
Average	**0.84**		**0.84**	
SD	**0.02**		**0.03**	
% RSD	**2.38**		**3.57**	

Table B.16 TLC system variation for the shoe polish (SP2) sample analysed on the same TLC plate.

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
SP2 (extract 1, spot 1)	0.45	purple	0.50	purple
SP2 (extract 2, spot 2)	0.45	purple	0.55	purple
SP2 (extract 3, spot 3)	0.45	purple	0.55	purple
SP2 (extract 4, spot 4)	0.50	purple	0.55	purple
SP2 (extract 5, spot 5)	0.45	purple	0.55	purple
SP2 (extract 6, spot 6)	0.45	purple	0.55	purple
Average	**0.46**		**0.54**	
SD	**0.02**		**0.02**	
% RSD	**4.35**		**3.70**	

Table B.17 TLC system variation for the lipstick sample (L21) analysed on six different TLC plates.

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
L21 (extract 1, Plate 1)	-		0.13	pink
L21 (extract 1, Plate 2)	-		0.13	pink
L21 (extract 1, Plate 3)	-		0.12	pink
L21 (extract 1, Plate 4)	-		0.12	pink
L21 (extract 1, Plate 5)	-		0.12	pink
L21 (extract 1, Plate 6)	-		0.13	pink
Average	**-**		**0.13**	
SD	**-**		**0.01**	
% RSD	**-**		**4.38**	

Table B.18 TLC system variation for the lip balm (LB19) sample analysed on six different TLC plates.

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	4.00	Spot Colour
LB19 (extract 1, Plate 1)	0.80	pink	0.85	pink
LB19 (extract 1, Plate 2)	0.75	pink	0.83	pink
LB19 (extract 1, Plate 3)	0.75	pink	0.80	pink
LB19 (extract 1, Plate 4)	0.85	pink	0.85	pink
LB19 (extract 1, Plate 5)	0.80	pink	0.85	pink
LB19 (extract 1, Plate 6)	0.85	pink	0.88	pink
Average	**0.80**		**0.84**	
SD	**0.04**		**0.03**	
% RSD	**5.59**		**3.15**	

Table B.19 TLC system variation for the shoe polish (SP2) sample analysed on six different TLC plates

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
SP2 (extract 1, Plate 1)	0.50	purple	0.55	purple
SP2 (extract 1, Plate 2)	0.50	purple	0.60	purple
SP2 (extract 1, Plate 3)	0.50	purple	0.60	purple
SP2 (extract 1, Plate 4)	0.45	purple	0.50	purple
SP2 (extract 1, Plate 5)	0.50	purple	0.55	purple
SP2 (extract 1, Plate 6)	0.55	purple	0.55	purple
Average	**0.50**		**0.55**	
SD	**0.03**		**0.03**	
% RSD	**6.32**		**6.74**	

The variation in the TLC system and extraction process where samples were analysed on the same TLC plate were less than 5% RSD, while the variation using different TLC plates were only slightly higher and less than 7% RSD. Both of the RSD values obtained were considered acceptable for the examinations undertaken based on the guideline that an RSD value for precise analysis is below 10% RSD [10].

B.6.2.2 Within and Between Sample Variation

Six separate extracts from the same sample were analysed using the optimised extraction and TLC systems as a set of six samples on one TLC plate and as one sample on six separate TLC plates. The results are presented in Tables B.20 to B.25.

Table B.20 Within sample variation for the lipstick (L21) sample analysed on the same TLC plate

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
L21 (extract 1)	-	-	0.20	pink
L21 (extract 2)	-	-	0.20	pink
L21 (extract 3)	-	-	0.18	pink
L21 (extract 4)	-	-	0.20	pink
L21 (extract 5)	-	-	0.18	pink
L21 (extract 6)	-	-	0.18	pink
Average	**-**		**0.19**	
SD	**-**		**0.01**	
% RSD	**-**		**5.76**	

Table B.21 Within sample variation for the lip balm (LB19) sample analysed on same TLC plate

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
LB19 (extract 1)	0.90	pink	0.90	pink
LB19 (extract 2)	0.90	pink	0.95	pink
LB19 (extract 3)	0.85	pink	0.90	pink
LB19 (extract 4)	0.90	pink	0.90	pink
LB19 (extract 5)	0.85	pink	0.95	pink
LB19 (extract 6)	0.85	pink	0.90	pink
Average	**0.88**		**0.92**	
SD	**0.03**		**0.03**	
% RSD	**3.13**		**2.82**	

Table B.22 Within sample variation for the shoe polish (SP2) sample analysed on same TLC plate

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
SP2 (extract 1)	0.60	purple	0.60	purple
SP2 (extract 2)	0.60	purple	0.60	purple
SP2 (extract 3)	0.65	purple	0.65	purple
SP2 (extract 4)	0.60	purple	0.65	purple
SP2 (extract 5)	0.60	purple	0.60	purple
SP2 (extract 6)	0.70	purple	0.70	purple
Average	**0.63**		**0.54**	
SD	**0.04**		**0.04**	
% RSD	**6.69**		**6.45**	

Table B.23 Within sample variation for the lipstick (L21) sample analysed on six different TLC plates

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
L21 (extract 1, Plate 1)	-	-	0.20	pink
L21 (extract 2, Plate 2)	-	-	0.18	pink
L21 (extract 3, Plate 3)	-	-	0.18	pink
L21 (extract 4, Plate 4)	-	-	0.20	pink
L21 (extract 5, Plate 5)	-	-	0.20	pink
L21 (extract 6, Plate 6)	-	-	0.18	pink
Average	**-**		**0.19**	
SD	**-**		**0.01**	
% RSD	**-**		**5.77**	

Table B.24 Within sample variation for the lip balm (LB19) sample analysed on six different TLC plates

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
LB19 (extract 1, Plate 1)	0.80	pink	0.90	pink
LB19 (extract 2, Plate 2)	0.85	pink	0.90	pink
LB19 (extract 3, Plate 3)	0.85	pink	0.85	pink
LB19 (extract 4, Plate 4)	0.80	pink	0.85	pink
LB19 (extract 5, Plate 5)	0.80	pink	0.90	pink
LB19 (extract 6, Plate 6)	0.85	pink	0.90	pink
Average	**0.83**		**0.88**	
SD	**0.03**		**0.03**	
% RSD	**3.32**		**2.92**	

Table B.25 Within sample variation for the shoe polish (SP2) sample analysed on six different TLC plates

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
SP2 (extract 1, Plate 1)	0.60	purple	0.55	purple
SP2 (extract 2, Plate 2)	0.55	purple	0.55	purple
SP2 (extract 3, Plate 3)	0.60	purple	0.60	purple
SP2 (extract 4, Plate 4)	0.55	purple	0.55	purple
SP2 (extract 5, Plate 5)	0.55	purple	0.50	purple
SP2 (extract 6, Plate 6)	0.50	purple	0.50	purple
Average	**0.56**		**0.54**	
SD	**0.04**		**0.04**	
% RSD	**6.74**		**6.94**	

The within sample variation has been revealed as being within acceptable RSD values for repeat analysis whether analysed on the same or different TLC plates.

Six separate extracts from six samples of the same brand of product for each sample type were analysed using the optimised extraction and TLC systems. The six extracts were analysed on the same and separate TLC plates and the results compared. The results are presented in tables B.26 to B.31.

Table B.26 Between sample variation for the lipstick (L21) sample analysed on same TLC plate

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
L21 (sample 1)	-	-	0.20	pink
L21 (sample 2)	-	-	0.20	pink
L21 (sample 3)	-	-	0.18	pink
L21 (sample 4)	-	-	0.20	pink
L21 (sample 5)	-	-	0.20	pink
L21 (sample 6)	-	-	0.18	pink
Average	**-**		**0.19**	
SD	**-**		**0.10**	
% RSD	**-**		**5.34**	

Table B.27 Between sample variation for the lip balm (LB19) sample analysed on same TLC plate

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
LB19 (sample 1)	0.80	pink	0.80	pink
LB19 (sample 2)	0.80	pink	0.80	pink
LB19 (sample 3)	0.80	pink	0.85	pink
LB19 (sample 4)	0.85	pink	0.80	pink
LB19 (sample 5)	0.85	pink	0.80	pink
LB19 (sample 6)	0.85	pink	0.85	pink
Average	**0.83**		**0.82**	
SD	**0.03**		**0.03**	
% RSD	**3.32**		**3.16**	

Table B.28 Between sample variation for the shoe polish (SP2) sample analysed on same TLC plate

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
SP2 (sample 1)	0.50	purple	0.55	purple
SP2 (sample 2)	0.50	purple	0.55	purple
SP2 (sample 3)	0.55	purple	0.60	purple
SP2 (sample 4)	0.55	purple	0.55	purple
SP2 (sample 5)	0.55	purple	0.55	purple
SP2 (sample 6)	0.50	purple	0.50	purple
Average	**0.53**		**0.55**	
SD	**0.03**		**0.03**	
% RSD	**5.23**		**5.75**	

Table B.29 Between sample variation for the lipstick (L21) sample analysed on six different TLC plates

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
L21 (sample 1,Plate 1)	-	-	0.20	pink
L21 (sample 2,Plate 2)	-	-	0.20	pink
L21 (sample 3,Plate 3)	-	-	0.20	pink
L21 (sample 4,Plate 4)	-	-	0.20	pink
L21 (sample 5,Plate 5)	-	-	0.18	pink
L21 (sample 6,Plate 6)	-	-	0.18	pink
Average	**-**		**0.19**	
SD	**-**		**0.01**	
% RSD	**-**		**5.34**	

Table B.30 Between sample variation for the lip balm (LB19) sample analysed on six different TLC plates

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
LB19 (sample 1, Plate 1)	0.80	pink	0.80	pink
LB19 (sample 2, Plate 2)	0.85	pink	0.85	pink
LB19 (sample 3, Plate 3)	0.85	pink	0.85	pink
LB19 (sample 4, Plate 4)	0.80	pink	0.80	pink
LB19 (sample 5, Plate 5)	0.80	pink	0.85	pink
LB19 (sample 6, Plate 6)	0.80	pink	0.80	pink
Average	**0.82**		**0.83**	
SD	**0.03**		**0.03**	
% RSD	**3.16**		**3.32**	

Table B.31 Between sample variation for the shoe polish (SP2) sample analysed on six different TLC plates

SAMPLE	R_F			
	1st Dev. (1.00)	Spot Colour	2nd Dev. (2.00)	Spot Colour
SP2 (sample 1, Plate 1)	0.50	purple	0.50	purple
SP2 (sample 2, Plate 2)	0.55	purple	0.60	purple
SP2 (sample 3, Plate 3)	0.50	purple	0.55	purple
SP2 (sample 4, Plate 4)	0.55	purple	0.55	purple
SP2 (sample 5, Plate 5)	0.55	purple	0.55	purple
SP2 (sample 6, Plate 6)	0.50	purple	0.55	purple
Average	**0.53**		**0.55**	
SD	**0.03**		**0.03**	
% RSD	**5.22**		**5.75**	

The between sample variation has acceptable RSD values for repeat analysis using the extraction system and TLC solvent system presented. This is irrespective of whether the samples are analysed on the same or different TLC plates.

B.6.2.3 TLC Analysis of the Sample Set

All of the test samples presented in Table B.1a – c were analysed in triplicate using the optimised extraction method and TLC system and the results of the analysis are presented in Tables B.32 to B.34, revealing their spot colour(s), R_f value(s), average, SD and % RSD.

Table B.32 TLC analysis outcomes for the lipstick samples under study. ND = not detected

SAMPLE	SPOT COLOUR	AVERAGE		SD		RSD%	
		1st Dev.	2nd Dev.	1st Dev.	2nd Dev	1st Dev.	2nd Dev.
L1	ND	-	-	-	-	-	-
L2	ND	-	-	-	-	-	-
L3	ND	-	-	-	-	-	-
L4	red	-	0.12	-	0.01	-	8.08
L5	orange	0.85	0.87	0.03	0.02	3.72	1.78
L6	red	-	0.13	-	0.01	-	8.43
L7	red	-	0.15	-	0.01	-	4.96
L8	red	0.10	0.13	0.00	0.01	0.00	8.43
L9	pink	-	0.16	-	0.01	-	7.38
L10	pink	-	0.17	-	0.01	-	7.74
L11	red	-	0.16	-	0.01	-	7.38
L12	red	-	0.16	-	0.01	-	8.36
L13	red	0.10	0.13	0.00	0.01	0.00	7.47
L14	red	-	0.11	-	0.01	-	9.68
L15	orange	0.24	0.32	0.01	0.02	3.36	7.72
L16	orange	0.83	0.85	0.04	0.02	4.90	1.88
L17	orange	0.83	0.84	0.04	0.02	4.90	2.43
L18	orange orange	0.24 0.45	0.31 0.48	0.01 0.03	0.03 0.02	4.24 7.03	9.39 3.81
L19	red	-	0.14	-	0.01	-	9.07
L20	orange orange	0.25 0.90	0.33 0.85	0.02 0.03	0.02 0.01	6.45 2.87	6.89 1.29
L21	pink	-	0.14	-	0.01	-	4.06
L22	pink	-	0.12	-	0.01	-	9.61
L23	pink	-	0.14	-	0.01	-	7.56
L24	orange	0.28	0.37	0.03	0.02	9.96	4.19
L25	red	-	0.12	-	0.01	-	9.61

Table B.33 TLC analysis outcomes for the lip balm samples under study. ND = not detected

SAMPLE	SPOT COLOUR	AVERAGE		SD		RSD%	
		1st Dev.	2nd Dev.	1st Dev.	2nd Dev.	1st Dev.	2nd Dev.
LB1	ND	-	-	-	-	-	-
LB2	ND	-	-	-	-	-	-
LB3	ND	-	-	-	-	-	-
LB4	ND	-	-	-	-	-	-
LB5	ND	-	-	-	-	-	-
LB6	ND	-	-	-	-	-	-
LB7	ND	-	-	-	-	-	-
LB8	ND	-	-	-	-	-	-
LB9	ND	-	-	-	-	-	-
LB10	ND	-	-	-	-	-	-
LB11	ND	-	-	-	-	-	-
LB12	ND	-	-	-	-	-	-
LB13	ND	-	-	-	-	-	-
LB14	ND	-	-	-	-	-	-
LB15	ND	-	-	-	-	-	-
LB16	ND	-	-	-	-	-	-
LB17	ND	-	-	-	-	-	-
LB18	ND	-	-	-	-	-	-
LB19	pink	0.84	0.84	0.02	0.03	2.38	3.57
LB20	ND	-	-	-	-	-	-
LB21	ND	-	-	-	-	-	-

Table B.34 TLC analysis outcomes for the shoe polish samples under study. ND = not detected

SAMPLE	SPOT COLOUR	AVERAGE		SD		RSD%	
		1st Dev.	2nd Dev.	1st Dev.	2nd Dev.	1st Dev.	2nd Dev.
SP1	pink	0.48	0.41	0.03	0.03	5.77	7.81
SP2	black	0.45	0.54	0.02	0.02	4.35	3.70
SP3	blue	0.55 0.88	0.56 0.87	0.05 0.03	0.01 0.02	9.96 3.13	0.99 1.90
SP4	red	0.87	0.86	0.03	0.03	3.66	3.20
SP5	black	0.91	0.88	0.05	0.02	5.41	2.78
SP6	black	0.50	0.53	0.04	0.03	8.94	5.59
SP7	ND	-	-	-	-	-	-
SP8	orange	0.66	0.85	0.04	0.02	5.72	2.88
SP9	orange	0.88	0.87	0.04	0.02	4.78	2.85
SP10	ND	-	-	-	-	-	-
SP11	ND	-	-	-	-	-	-
SP12	ND	-	-	-	-	-	-
SP13	orange	0.88	0.87	0.05	0.02	5.85	2.46
SP14	black	0.54	0.30	0.04	0.03	6.95	9.21
SP15	ND	-	-	-	-	-	-
SP16	pink	0.85	0.86	0.03	0.02	3.72	1.80
SP17	ND	-	-	-	-	-	-

The colour observed for the spots developed using TLC could be subjectively categorised into five groups. Twenty one of the twenty four lipsticks and eleven of the seventeen shoe polish samples produced coloured spots, however only one of the twenty one lip balm samples produced a coloured spot using TLC. The various coloured spots are listed by sample in Table B.35

Table B.35 Spot colour(s) observed in the samples under study

Spot Colour	Observed in Sample
Pink	LB19, L21, L22, L22 and SP1
Red	SP4, L4, L6, L7, L8, L11, L12, L13, L14, L19 and L25
Orange	L5, L15, L16, L17, L18, L20, L24, SP8, SP9 and SP13
Blue	SP3
Black	SP2, SP5, SP6 and SP14

B.6.3 Microspectrophotometry

Microspectrophotometric (MSP) data was acquired directly from the sample spot on the TLC plate in each case once the TLC analysis was completed. As such MSP data was only available for those samples which produced coloured spots on analysis by TLC. Variations within the acquisition method and derived data were assessed using one sample of each of the products under study as previously detailed.

Spectra were recorded in the range of 360 – 780 nm wavelengths. The average, standard deviation and percentage relative standard deviation for six replicates of the percentage reflectance (normalised to the total response) for each test sample was determined. Examples of the MSP spectra of the test samples L21, LB19 and SP2 are presented in Figures B.5 to B.7 respectively.

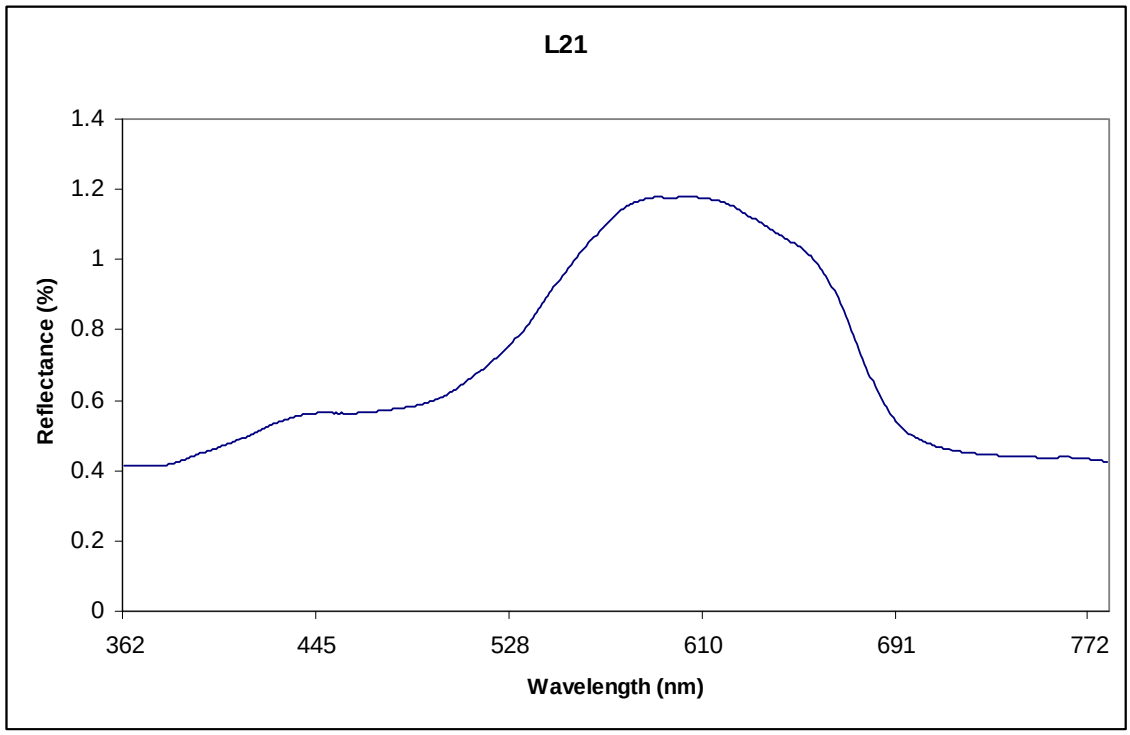


Figure B.5 Spectrum of lipstick (L21) sample

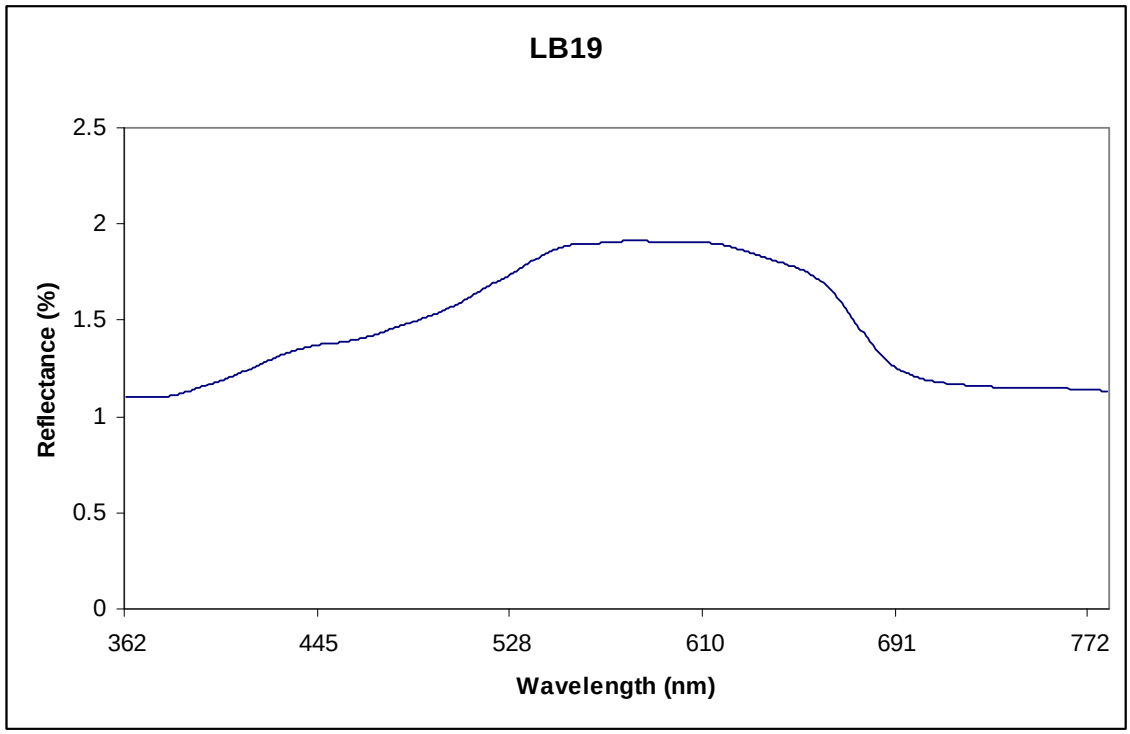


Figure B.6 Spectrum of lip balm (LB19) sample

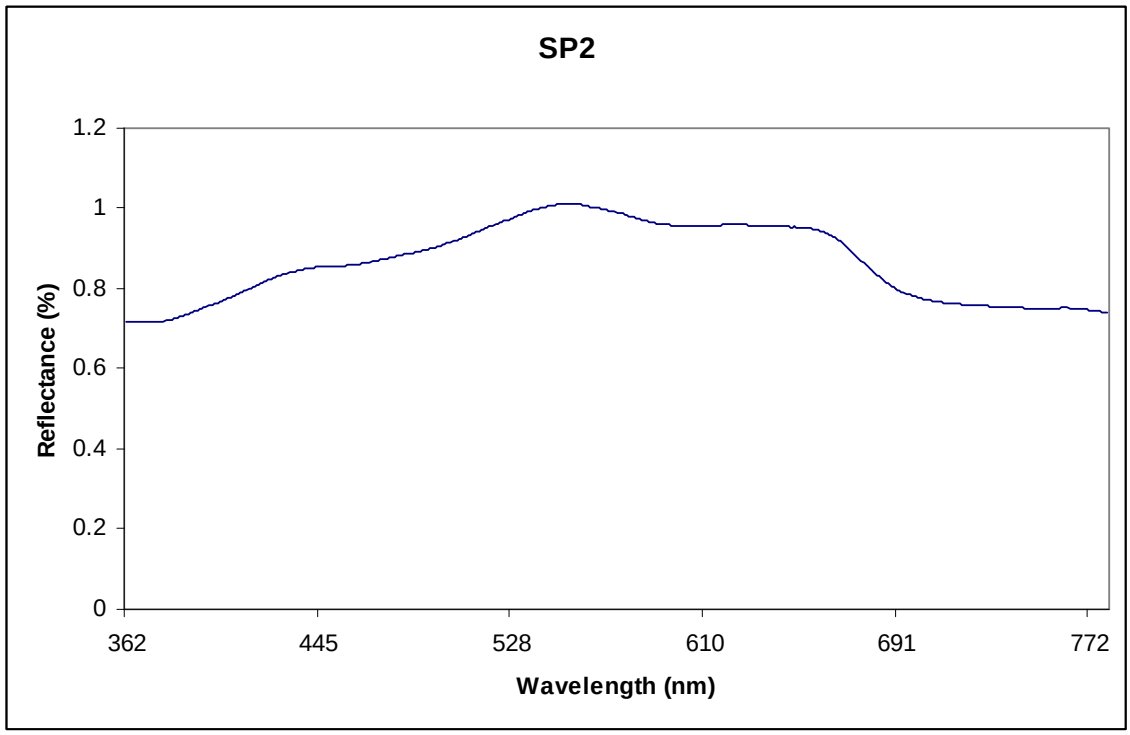


Figure B.7 Spectrum of shoe polish (SP2) sample

B.6.3.1 Instrumental Variation

Instrumental variation was assessed by six repetitive measurements of the same sample spot and the data is presented in Table B.36 for lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample respectively.

Table B.36 The outcomes for the assessment of the instrumental variation

SAMPLE	AVERAGE (% Reflectance)	SD	%RSD
L21	1.063	0.009	0.928
LB19	1.195	0.011	1.017
SP2	0.749	0.010	1.405

The %RSD for all the representative samples indicated good instrumental repeatability.

B.6.3.2 Within and Between Spot Variation

MSP data was recorded for six locations on the same spot on the TLC plate for each of the three representative samples in order to make an assessment of the within spot repeatability. The results are presented in Table B.37 indicating excellent repeatability within a TLC spot.

Table B.37 The outcomes for the within spot variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample.

SAMPLE	AVERAGE (% Reflectance)	SD	%RSD
L21	0.974	0.016	1.771
LB19	1.385	0.019	1.493
SP2	0.703	0.018	2.296

The repeatability of the MSP analysis between replicate spots of the same sample extract developed on the same and separate TLC plates are presented in Tables B.38 and B.39 respectively and again excellent repeatability was observed.

Table B.38 The outcomes for the between spot variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample on same TLC plates.

SAMPLE	AVERAGE (% Reflectance)	SD	%RSD
L21	1.333	0.016	1.273
LB19	0.375	0.013	2.137
SP2	0.759	0.021	2.849

Table B.39 The outcomes for the between spot variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample on six different TLC plates.

SAMPLE	AVERAGE (% Reflectance)	SD	%RSD
L21	0.564	0.012	2.748
LB19	1.304	0.023	1.863
SP2	0.032	0.021	3.430

B.6.3.3 Within and Between Sample Variation

The repeatability of the MSP analysis between different extracts of the same sample developed on the same and separate TLC plates are presented in Tables B.40 and B.41 and excellent repeatability was observed.

Table B.40 The outcomes for the within sample between spot variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample on same TLC plate

SAMPLE	AVERAGE (% Reflectance)	SD	%RSD
L21	1.037	0.024	2.454
LB19	1.238	0.022	1.831
SP2	0.820	0.034	3.869

Table B.41 The outcomes for the within sample between spot variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample on six different TLC plates.

SAMPLE	AVERAGE (% Reflectance)	SD	%RSD
L21	0.699	0.021	3.333
LB19	1.460	0.027	1.863
SP2	0.840	0.027	3.182

Finally the repeatability of the MSP analysis between extracts of different samples of the same brand of product developed on the same TLC plate and on separate TLC plates are presented in Tables B.42 and B.43 and again good repeatability was observed with only one %RSD value being slightly greater than 5%.

Table B.42 The outcomes for the between sample variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample on same TLC plate.

SAMPLE	AVERAGE (% Reflectance)	SD	%RSD
L21	0.468	0.015	4.622
LB19	0.499	0.014	4.081
SP2	0.635	0.019	3.063

Table B.43 The outcomes for the between sample variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample on six different TLC plates.

SAMPLE	AVERAGE (% Reflectance)	SD	%RSD
L21	0.631	0.018	3.358
LB19	0.416	0.009	5.562
SP2	0.695	0.029	4.299

It is clear from the MSP variation studies that the instrument, extraction and method chosen for the recording of the MSP data (i.e. directly from the TLC plate) gave excellent results in terms of repeatability.

B.6.4 UV/Vis Analysis Variation Studies

B.6.4.1 Instrumental Variation

Instrumental variation was assessed by analysing six repetitive measurements of an extraction of the same sample. The spectra from each of the representative samples, lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) are presented in Figures B.8 to B.10 and the variation in maximum absorbance of the samples are presented in Table B.44 as an average of the six samples.

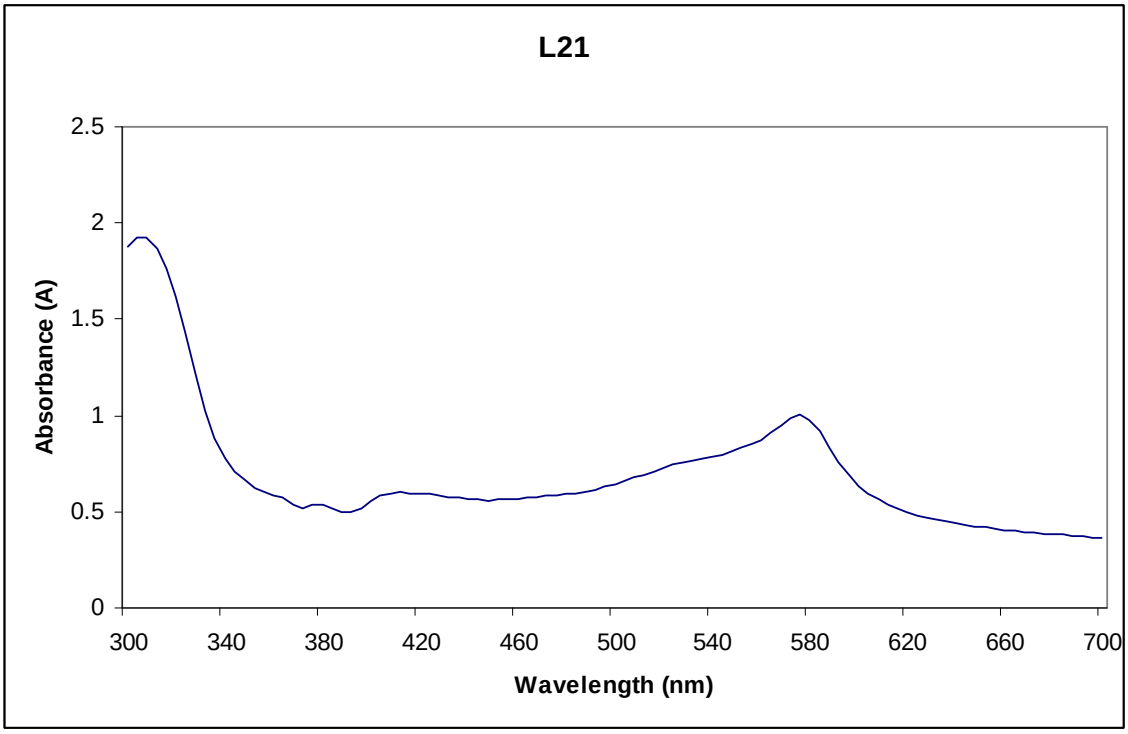


Figure B.8 Spectrum of lipstick (L21) sample

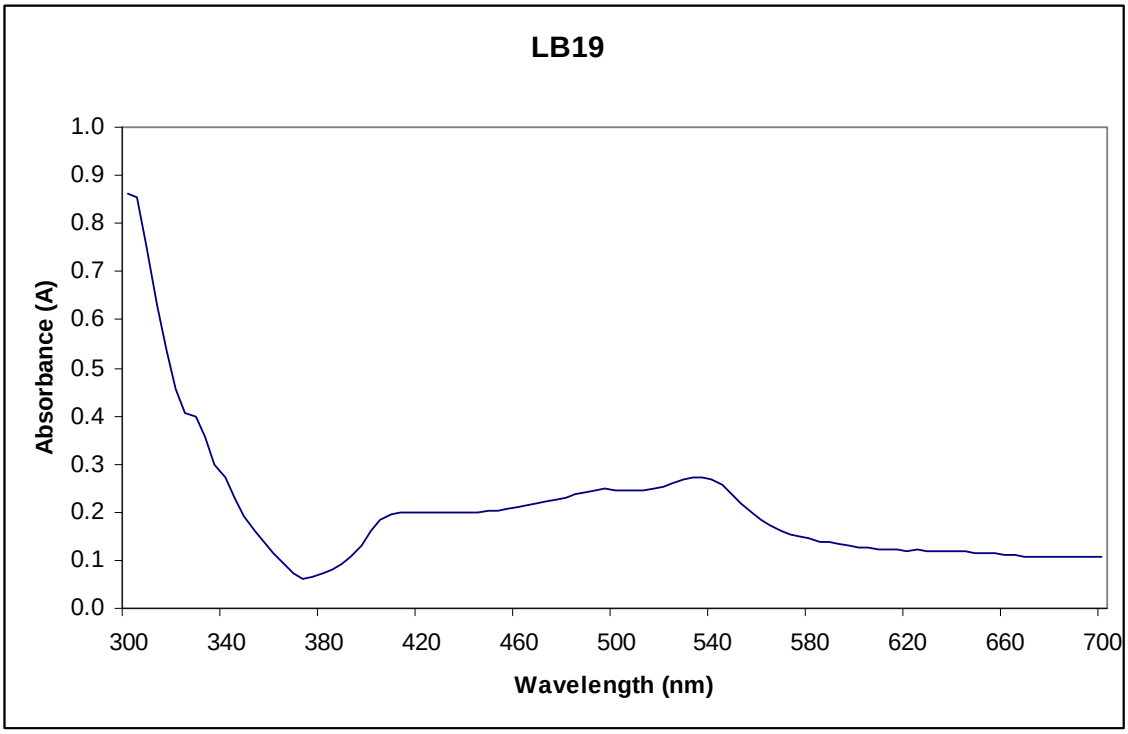
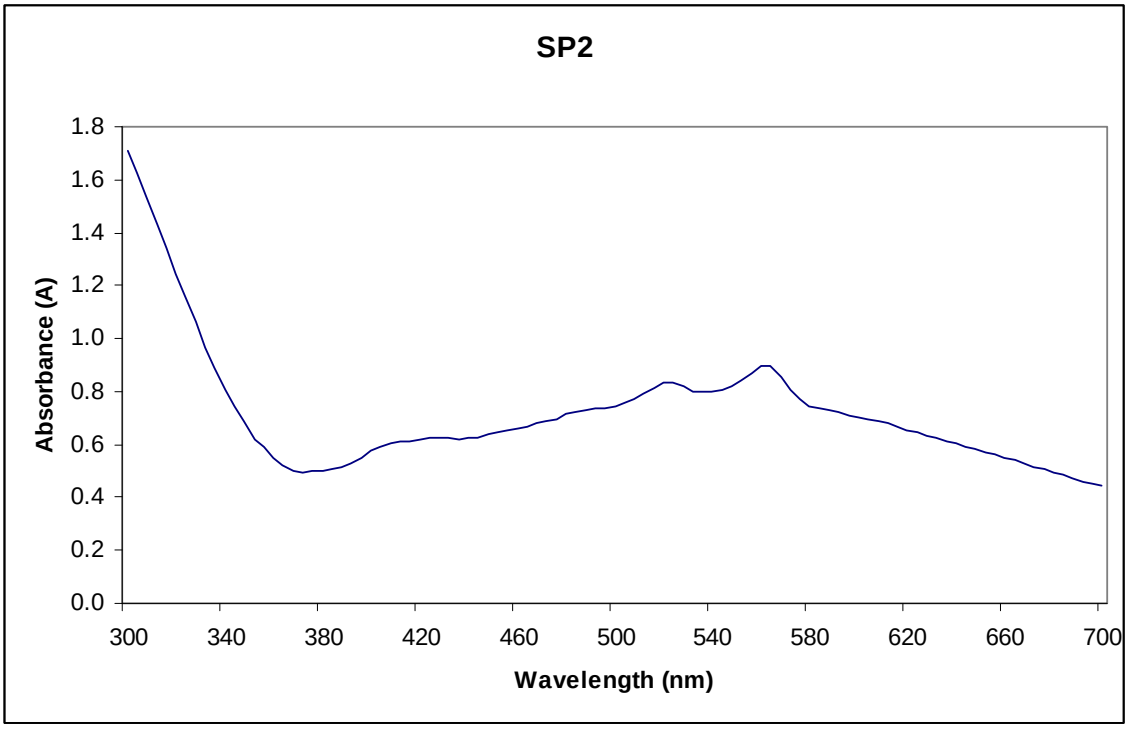


Figure B.9 Spectrum of lip balm (LB19) sample



FigureB.10 Spectrum of shoe polish (SP2) sample

Table B.44 The outcomes for the variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample.

SAMPLE	AVERAGE (Absorbance)	SD	%RSD
L21	0.691	0.009	1.456
LB19	0.330	0.002	0.833
SP2	0.497	0.001	0.317

As it can be seen from the table, the range of %RSD calculated for the representative samples all indicated excellent repeatability.

B.6.4.2 Within and Between Sample Variation

Six separate extractions of the same and different samples of the representative samples were analysed and the data are presented in Table B.45 and B.46 respectively. The %RSD recorded across the samples indicated that there was good repeatability between different extractions of the same sample.

Table B.45 The outcomes for the within sample variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample.

SAMPLE	AVERAGE (Absorbance)	SD	%RSD
L21	0.665	0.019	3.164
LB19	0.205	0.007	4.063
SP2	0.722	0.010	1.403

Table B.46 The outcomes for the between sample variation studies for the lipstick (L21), the lip balm (LB19) and the shoe polish (SP2) sample.

SAMPLE	AVERAGE (Absorbance)	SD	%RSD
L21	0.825	0.044	5.603
LB19	0.228	0.006	2.713
SP2	0.749	0.044	6.176

B.6.5 Gas Chromatographic Analysis

B.6.5.1 Instrumental Variation

The GC instrumental precision was confirmed using a prepared Grob mixture and the data is presented in Table B.47 and B.48 as un-normalised and normalised peak area data. This facilitated the confirmation of the GC system in term of its repeatability and the repeatability of the manual injections used. Figure B.11 shows the GC profile of the Grob mixture.

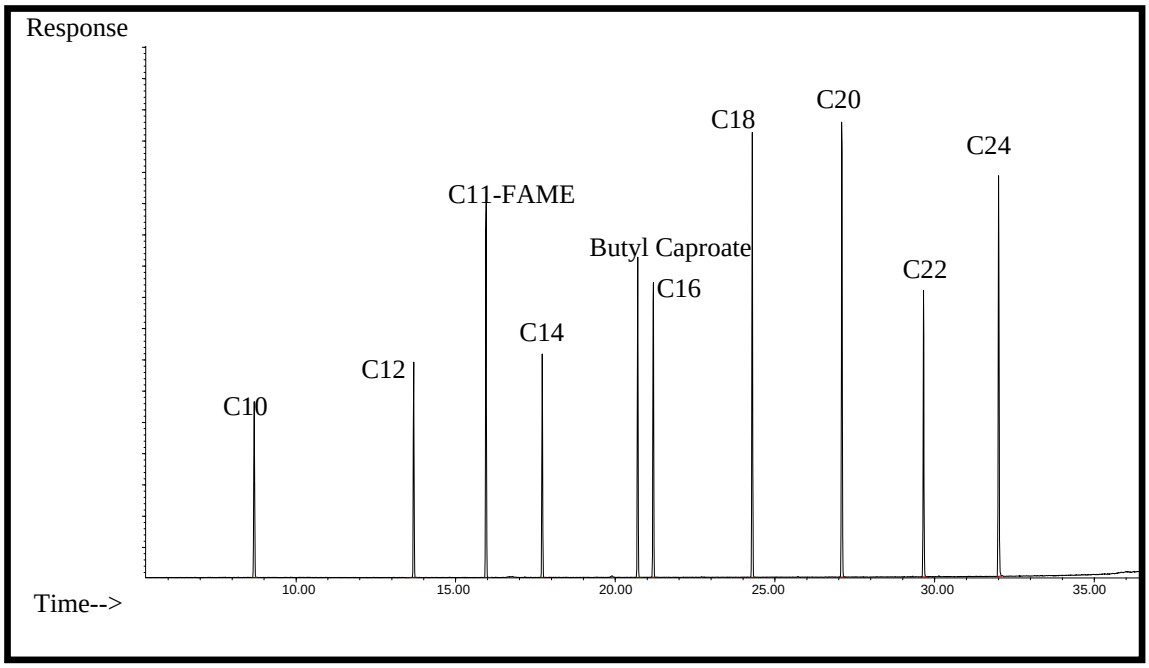


Figure B.11 GC profile of the Grob mixture

The Grob mixture was formulated to reflect the main compounds found within wax based products and as such contained alkanes, methyl decanoate, (C_{11}-FAME, $C_{11}H_{22}O_2$) and butyl caproate ($C_{14}H_{28}O_2$) were included in the mixture because these compounds (or similar compounds of these types) are commonly used as fragrances in cosmetic products especially in lipsticks and lip balms.

Table B.47 Average (peak area), SD and % RSD for six injections of the Grob mixture

COMPOUND	AVERAGE (Peak Area)	SD	%RSD
C10 (Decane)	1534161	6957	0.454
C12 (Dodecane)	1447704	5131	0.355
C11-FAME	1044183	5634	0.540
C14 (Tetradecane)	1487382	12502	0.841
Butyl Caproate	982223	10824	1.102
C16 (Hexadecane)	1234943	14538	1.177
C18 (Octadecane)	1054784	5840	0.544
C20 (Eicosane)	970413	9543	0.983
C22 (Docosane)	795795	14588	1.833
C24 (Tetracosane)	978205	16229	1.659

Table B.48 Average, SD and %RSD for the Grob mixture normalised against C10 (Decane) i.e. the internal standard (ISTD).

COMPOUND	AVERAGE (Normalised Peak Area)	SD	% RSD
C10 (Decane) (ISTD)	6.0000	0.0000	0.0000
C12 (Dodecane)	5.6233	0.0152	0.2707
C11-FAME	3.9937	0.0070	0.1744
C14 (Tetradecane)	5.7804	0.0310	0.5365
Butyl Caproate	3.7938	0.0374	0.9869
C16 (Hexadecane)	4.8378	0.0427	0.8831
C18 (Octadecane)	4.3614	0.0540	1.2379
C20 (Eicosane)	3.8875	0.0356	0.9185
C22 (Docosane)	3.1877	0.0377	1.1832
C24 (Tetracosane)	2.1837	0.0556	2.5440

The %RSD(s) recorded for all components in the Grob mixture were less than 5% indicating the repeatability of both the GC instrument and the injection mechanism were within an acceptable range.

B.6.5.2 Peak Asymmetry Factors

Peak symmetry of the Grob mixture was accessed by calculation of the asymmetry factor for each peak as outlined in Section B.3.6.3. The results of the calculations are as displayed in Table B.49.

Table B.49 Peak asymmetry factor of the Grob mixture

PEAK	COMPOUND	ASYMMETRY FACTOR
1	C10 (Decane)	1.67
2	C12 (Dodecane)	1.38
3	C11-FAME	1.00
4	C14 (Tetradecane)	0.89
5	Butyl Caproate	1.00
6	C16 (Hexadecane)	1.05
7	C18 (Octadecane)	1.36
8	C20 (Eicosane)	0.85
9	C22 (Docosane)	1.00
10	C24 (Tetracosane)	1.00

The peak asymmetry factor is a measure of peak tailing. Compounds having a peak asymmetry factor greater than 1.00 indicates that they are absorbed to the column thus resulted in peak tailing. This can be quite common for compounds containing the –OH functional group. For neutral compounds such as long chain hydrocarbons an asymmetry factor of 1.00 is expected and it was observed that there was some deviation from this value in the Grob samples analysed. This may have been due, in part, to the fact that the asymmetry measurements were made by hand.

B.6.5.3 GC Analysis Variation Studies

Sample variation studies were conducted using the representative samples (L21, LB19 and SP2). An example of the chromatogram of each of the representative samples are presented in Figures B.12 to B.14.

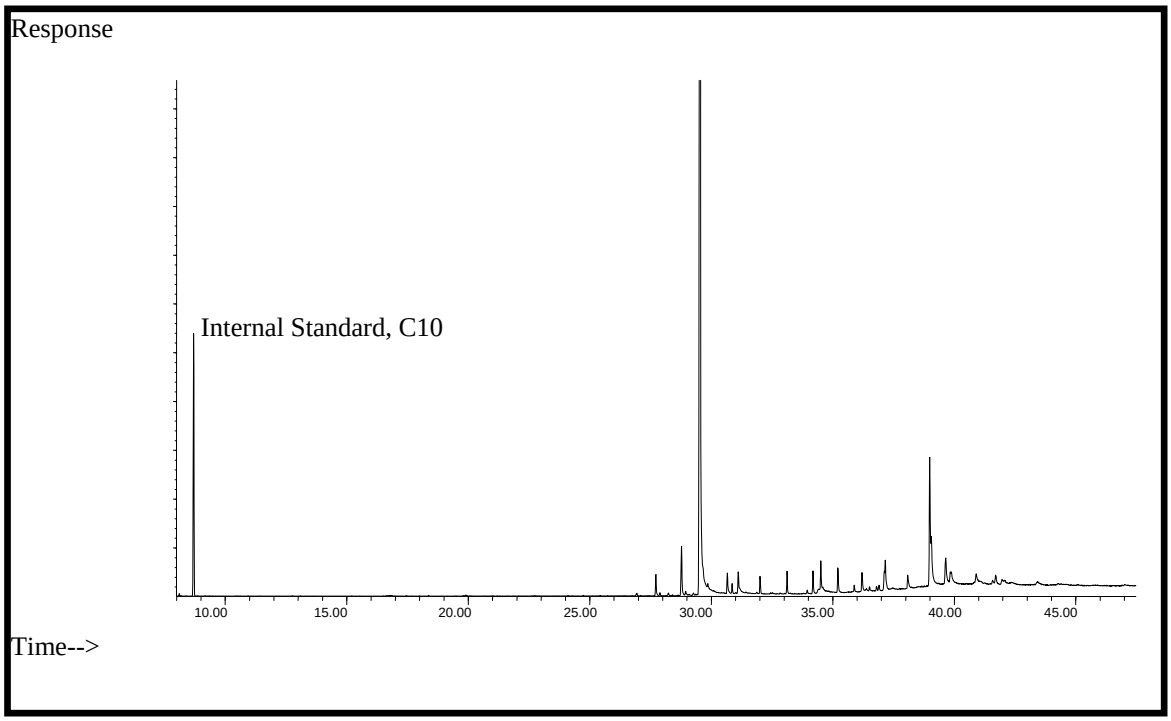


Figure B.12 GC profile of the lipstick (L21) sample

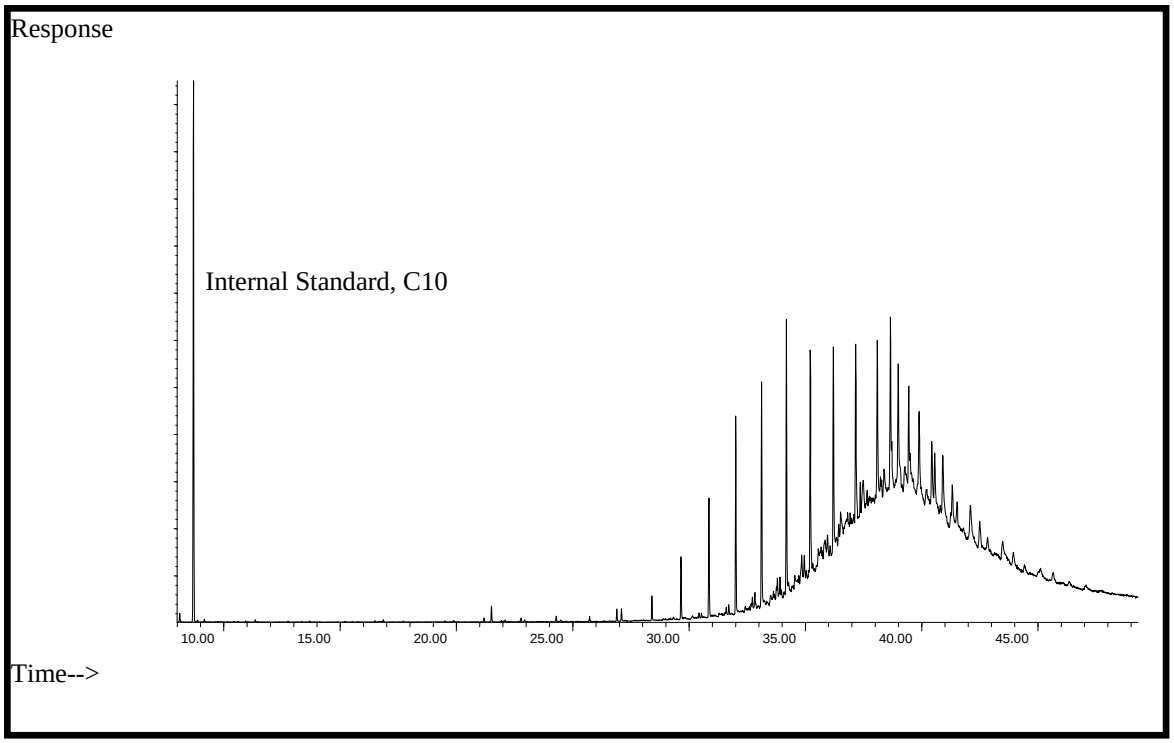


Figure B.13 GC profile of the lip balm (LB19) sample.

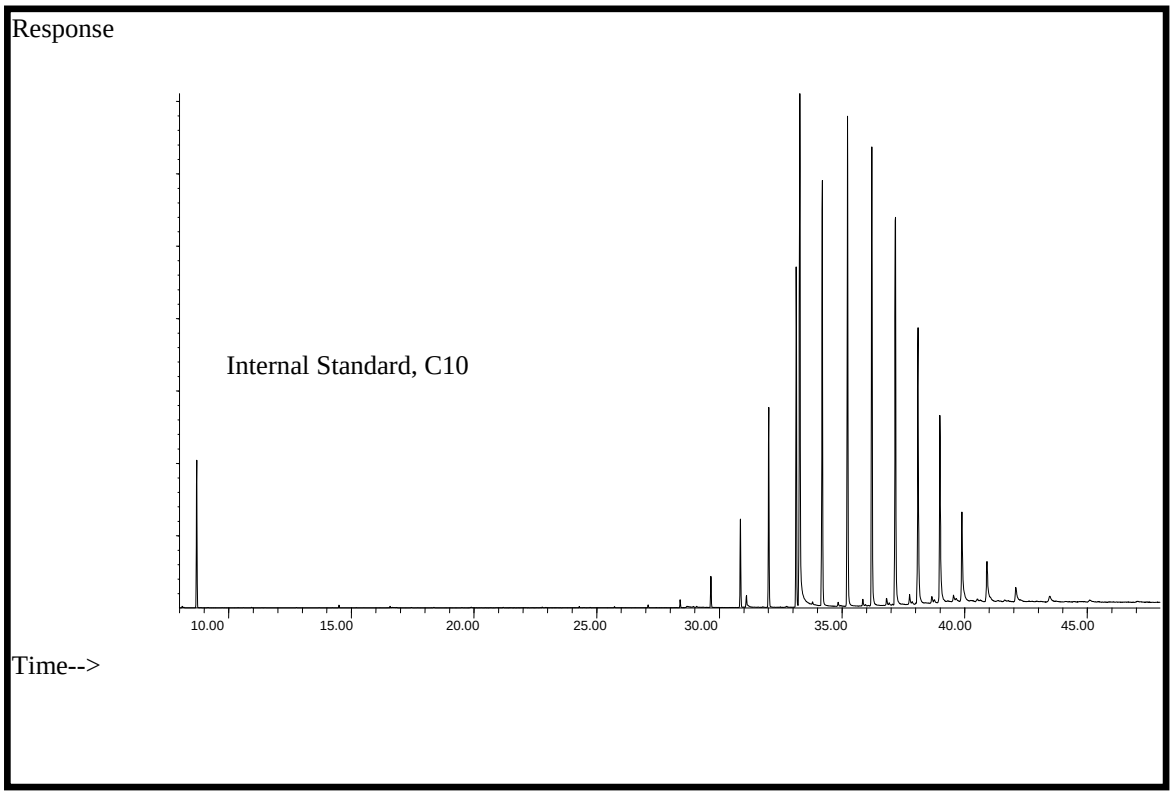


Figure B.14 GC profile of the shoe polish (SP2) sample

B.6.5.4 Variation Within the Same Extraction

Six injections of the same extraction of each of the representative samples were performed and the resultant data were normalised to the peak area of the internal standard. In each case the important factor was the presence or absence of peaks rather than determining the identity of the responsible components. As such, the various peaks were not identified other than by retention time. In each case a different set of peaks were recorded and only peaks which produced a response with a %RSD value below 15 % were selected for inclusion in the study The threshold of 15% RSD was chosen as an arbitrary value to represent an acceptable variation in the chromatographic response of analysis of the same sample. The average, SD and %RSD of each of the representative samples are presented in Tables B.50 to B.52.

Table B.50 Average, SD and % RSD for twenty five peaks chosen in lipstick (L21) samples in within-run variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD	% RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.31)	0.2048	0.0181	8.8266
2 (23.37)	0.1890	0.0153	8.0867
3 (25.99)	0.1471	0.0179	12.1889
4 (27.26)	0.0202	0.0019	9.4109
5 (28.36)	0.1037	0.0083	8.0020
6 (29.98)	0.0724	0.0065	8.9522
7 (30.24)	0.0351	0.0032	9.2137
8 (30.55)	0.0774	0.0075	9.7308
9 (30.82)	4.4454	0.2308	5.1914
10 (32.07)	0.1224	0.0080	6.5136
11 (32.22)	0.0218	0.0013	5.8386
12 (32.49)	0.0496	0.0047	9.5040
13 (33.46)	0.0749	0.0077	10.2266
14 (34.31)	0.0485	0.0051	10.5298
15 (34.66)	0.0910	0.0067	7.3307
16 (35.84)	0.0697	0.0065	9.3526
17 (37.06)	0.0685	0.0071	10.3489
18 (38.34)	0.0758	0.0073	9.6507
19 (38.63)	0.0844	0.0084	9.9267
20 (39.73)	0.1412	0.0144	10.1782
21 (41.27)	0.1524	0.0146	9.5602
22 (43.01)	0.5044	0.0308	6.0988
23 (43.80)	0.0801	0.0077	9.6533
24 (44.60)	0.1705	0.0131	7.6975
25 (44.96)	0.0748	0.0076	10.1449
Average	0.3125	0.0177	8.5445

Table B.51 Average, SD and % RSD for eighteen peaks chosen in lip-balm (LB19) sample in within-run variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD.	%RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.03)	0.2827	0.0207	7.3278
2 (23.10)	0.2935	0.0285	9.6989
3 (24.86)	0.0858	0.0104	12.0751
4 (25.73)	0.1895	0.0110	5.7901
5 (27.87)	0.0412	0.0044	10.6361
6 (28.09)	0.1128	0.0142	12.6086
7 (29.29)	0.0275	0.0023	8.2546
8 (30.25)	0.0752	0.0080	10.6437
9 (30.64)	0.0866	0.0085	9.7626
10 (31.93)	0.0947	0.0095	9.9792
11 (33.18)	0.1707	0.0144	8.4553
12 (34.37)	0.2052	0.0273	13.2842
13 (35.54)	0.2799	0.0382	13.6459
14 (36.74)	0.2873	0.0328	11.4012
15 (38.00)	0.2595	0.0242	9.3277
16 (39.36)	0.2657	0.0228	8.5651
17 (40.86)	0.2848	0.0334	11.7260
18 (44.45)	0.1842	0.0217	11.7664
Average	0.2225	0.0175	9.7341

Table B.52 Average, SD and %RSD for eighteen peaks chosen in shoe polish (SP2) sample in within-run variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD.	%RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.13)	0.3555	0.0408	11.4910
2 (23.18)	0.3429	0.0314	9.1589
3 (24.93)	0.0431	0.0052	11.9981
4 (25.80)	0.2328	0.0244	10.4820
5 (27.95)	0.0576	0.0067	11.6537
6 (28.15)	0.1395	0.0155	11.1152
7 (29.62)	0.2348	0.0189	8.0285
8 (32.01)	0.1169	0.0119	10.2128
9 (32.29)	0.0480	0.0067	13.8869
10 (33.26)	0.2911	0.0421	14.4535
11 (34.45)	0.4802	0.0204	4.2484
12 (35.62)	0.7105	0.0315	4.4407
13 (36.83)	1.0367	0.0729	7.0357
14 (38.10)	1.1130	0.1061	9.5358
15 (39.46)	1.1448	0.0977	8.5326
16 (40.97)	1.0142	0.0236	2.3306
17 (42.67)	0.9392	0.0469	4.9919
18 (44.58)	0.7193	0.0447	6.2082
Average	0.5274	0.0341	8.4108

B.6.5.5 Within Sample Variation

Six injections of different extractions of the same sample of each of the representative samples were performed and the resultant data were normalised to the peak area of the internal standard. Data is presented for the same peaks within each sample chosen in the previous section. The average, SD and %RSD of each of the representative samples are presented in Tables B.53 to B.55.

Table B.53 Average, SD and % RSD for twenty five peaks chosen in lipstick (L21) sample in within sample variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD.	%RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.31)	0.1676	0.0202	12.0717
2 (23.37)	0.1997	0.0311	15.5816
3 (25.99)	0.1646	0.0182	11.0887
4 (27.26)	0.0254	0.0041	16.0691
5 (28.36)	0.0898	0.0050	5.5402
6 (29.98)	0.0801	0.0091	11.3504
7 (30.24)	0.0388	0.0066	16.8888
8 (30.55)	0.0667	0.0044	6.6611
9 (30.82)	4.8168	0.4922	10.2189
10 (32.07)	0.1285	0.0104	8.0617
11 (32.22)	0.0246	0.0037	14.8124
12 (32.49)	0.0387	0.0059	15.3535
13 (33.46)	0.0451	0.0083	18.3855
14 (34.31)	0.0415	0.0083	19.9818
15 (34.66)	0.0804	0.0137	17.0029
16 (35.84)	0.0733	0.0115	15.7377
17 (37.06)	0.0769	0.0141	18.3181
18 (38.34)	0.0662	0.0140	21.1333
19 (38.63)	0.0713	0.0135	19.0140
20 (39.73)	0.1064	0.0173	16.2200
21 (41.27)	0.0583	0.0079	13.4947
22 (43.01)	0.4658	0.0671	14.4026
23 (43.80)	0.0728	0.0126	17.2529
24 (44.60)	0.1548	0.0202	13.0556
25 (44.96)	0.0653	0.0121	18.4490
Average	0.3161	0.0320	14.0825

Table B.54 Average, SD and % RSD for eighteen peaks chosen in lip-balm (LB19) sample in within sample variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD.	%RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.03)	0.0905	0.0073	8.0561
2 (23.10)	0.3313	0.0285	8.5983
3 (24.86)	0.0971	0.0115	11.8396
4 (25.73)	0.0890	0.0070	7.8123
5 (27.87)	0.1342	0.0100	7.4548
6 (28.09)	0.1247	0.0143	11.4402
7 (29.29)	0.1296	0.0134	10.3350
8 (30.25)	0.0703	0.0097	13.8245
9 (30.64)	0.2528	0.0214	8.4763
10 (31.93)	0.3927	0.0319	8.1231
11 (33.18)	0.3398	0.0374	11.0199
12 (34.37)	0.3220	0.0365	11.3435
13 (35.54)	0.3713	0.0266	7.1526
14 (36.74)	0.3861	0.0288	7.4484
15 (38.00)	0.3796	0.0301	7.9302
16 (39.36)	0.2894	0.0306	10.5843
17 (40.86)	0.2368	0.0197	8.3384
18 (44.45)	0.3845	0.0275	7.1552
Average	0.2854	0.0206	8.7859

Table B.55 Average, SD and % RSD for eighteen peaks chosen in shoe polish (SP2) sample in within sample variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD.	%RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.13)	0.6670	0.0527	7.8951
2 (23.18)	0.6531	0.0563	8.6270
3 (24.93)	0.0279	0.0057	20.4427
4 (25.80)	0.4964	0.0659	13.2682
5 (27.95)	0.0501	0.0070	14.0305
6 (28.15)	0.2132	0.0320	15.0278
7 (29.62)	0.2670	0.0259	9.6918
8 (32.01)	0.1302	0.0137	10.5278
9 (32.29)	0.0635	0.0108	17.0488
10 (33.26)	0.2956	0.0553	18.6999
11 (34.45)	0.5803	0.0439	7.5607
12 (35.62)	0.8782	0.0517	5.8914
13 (36.83)	1.2535	0.0926	7.3914
14 (38.10)	1.3166	0.1358	10.3108
15 (39.46)	1.5150	0.1762	11.6280
16 (40.97)	1.2810	0.1712	13.3603
17 (42.67)	1.1269	0.1283	11.3845
18 (44.58)	0.8533	0.1354	15.8669
Average	0.6668	0.0663	11.5081

B.6.5.6 Between Sample Variation

Six injections of single extractions of six different batches of the same sample of each representative sample were performed and the resultant data were normalised to the peak area of the internal standard. Again the same peaks for each sample were compared. The average, SD and %RSD of each of the representative samples are presented in Tables B.56 to B.58.

Table B.56 Average, SD and % RSD for twenty five peaks chosen in lipstick (L21) sample in between sample variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD.	%RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.31)	0.1761	0.0373	21.1837
2 (23.37)	0.1980	0.0380	19.2197
3 (25.99)	0.1527	0.0205	13.4349
4 (27.26)	0.0185	0.0050	27.1340
5 (28.36)	0.1026	0.0191	18.6246
6 (29.98)	0.0886	0.0205	23.1707
7 (30.24)	0.0339	0.0052	15.2497
8 (30.55)	0.0764	0.0189	24.6772
9 (30.82)	3.4585	0.5784	16.7229
10 (32.07)	0.1624	0.0296	18.2041
11 (32.22)	0.0385	0.0059	15.4253
12 (32.49)	0.0434	0.0092	21.2170
13 (33.46)	0.0346	0.0057	16.5336
14 (34.31)	0.0434	0.0091	20.9063
15 (34.66)	0.0871	0.0184	21.1168
16 (35.84)	0.0602	0.0144	23.8786
17 (37.06)	0.1302	0.0237	18.1832
18 (38.34)	0.0933	0.0131	14.0172
19 (38.63)	0.1736	0.0377	21.6936
20 (39.73)	0.2189	0.0505	23.0881
21 (41.27)	0.1028	0.0260	25.2794
22 (43.01)	0.3953	0.0886	22.4038
23 (43.80)	0.1044	0.0183	17.5525
24 (44.60)	0.1559	0.0394	25.2770
25 (44.96)	0.0964	0.0210	21.8174
Average	0.2787	0.0444	19.4620

Table B.57 Average, SD and % RSD for eighteen peaks chosen in lip-balm (LB19) sample in between sample variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD.	%RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.03)	0.2656	0.0558	21.0129
2 (23.10)	0.2812	0.0429	15.2677
3 (24.86)	0.1011	0.0171	16.9396
4 (25.73)	0.1920	0.0417	21.6920
5 (27.87)	0.0263	0.0061	23.2058
6 (28.09)	0.1166	0.0196	16.7845
7 (29.29)	0.0352	0.0084	23.7257
8 (30.25)	0.0872	0.0210	24.0757
9 (30.64)	0.1326	0.0222	16.7601
10 (31.93)	0.2136	0.0338	15.8169
11 (33.18)	0.3133	0.0517	16.5122
12 (34.37)	0.2176	0.0546	25.0756
13 (35.54)	0.2479	0.0644	25.9822
14 (36.74)	0.2656	0.0631	23.7662
15 (38.00)	0.2420	0.0503	20.7953
16 (39.36)	0.3291	0.0722	21.9242
17 (40.86)	0.3618	0.0904	24.9966
18 (44.45)	0.2494	0.0365	14.6234
Average	0.2462	0.0396	19.4188

Table B.58 Average, SD and % RSD for eighteen peaks chosen in shoe polish (SP2) sample in between sample variation studies

Peak No. (Retention Time)	Average (Normalised Area)	SD.	%RSD
C10 (9.46)	1.0000	0.0000	0.0000
1 (20.13)	0.3808	0.0443	11.6353
2 (23.18)	0.4822	0.0886	18.3662
3 (24.93)	0.0398	0.0086	21.6876
4 (25.80)	0.2343	0.0245	10.4781
5 (27.95)	0.0515	0.0099	19.2382
6 (28.15)	0.1832	0.0419	22.8688
7 (29.62)	0.2491	0.0441	17.7157
8 (32.01)	0.1072	0.0269	25.0900
9 (32.29)	0.0550	0.0095	17.3074
10 (33.26)	0.2500	0.0643	25.7154
11 (34.45)	0.5226	0.1224	23.4236
12 (35.62)	0.7398	0.1384	18.7022
13 (36.83)	1.2018	0.1825	15.1857
14 (38.10)	1.3371	0.1421	10.6293
15 (39.46)	1.3088	0.2752	21.0269
16 (40.97)	1.0897	0.2078	19.0660
17 (42.67)	0.9652	0.1468	15.2105
18 (44.58)	1.0132	0.2690	26.5462
Average	0.5901	0.0972	17.8891

Higher average %RSD values were observed for the between sample and within sample extractions when compared to repetitive analysis of the same extracted sample. This is in line with expectations as both the sample to sample and the extraction variability is being included. Notwithstanding this, these values are still acceptable considering that other studies [11, 12] had also reported high %RSD for similar analysis i.e. up to 25.6% [11] and 30% [12].

B.7 Conclusions

The percentage relative standard deviations (%RSD) reported during the validation studies for the extracted wax based products indicated good repeatability of analysis across all of the techniques employed in this work. Gas chromatographic analysis demonstrated higher variations than the other analytical techniques between sample extractions. This is to be expected to some degree given that the analysis was preformed using manual injections. Not withstanding this, the variation from sample to sample was considered acceptable when compared with other studies.

B.8 REFERENCES

1. Andrasko, J., *Forensic Analysis of Lipsticks.* Forensic Science International, 1981. **17**: p. 235 - 251.

2. Jasuja, O.P. and R.Singh, *Thin Layer Chromatographic Analysis of Liquid Lipsticks.* Journal of Forensic Identification, 2005. **55**(1): p. 28 - 35.

3. Cole, M.D. and Thorpe, J.W., *The Analysis of Black Shoe Polish Marks on Clothing.* Journal of Forensic Science Society, 1972. **12**: p. 449 - 451.

4. Cole, M.D., Milligan, F. and Thorpe, J.W., *The Examination of Black Wax Shoe Polish Stains After Ageing and Weathering.* Journal of Forensic Science Society, 1994. **34**: p. 23 - 27.

5. Byrne, L.M., Cole, M.D., Milligan, F. and Thorpe, J.W., *Shoe Polish Stains on Fabric: A Comparison of Different Shoe Polish Types.* Journal of Forensic Science Society, 1994. **34**: p. 354 - 362.

6. Russell, L.W., *Analysis of Lipsticks.* Forensic Science International, 1984. **25**: p. 105 - 116.

7. Grob, K. Jr., Grob, G. and Grob, K., *Comprehensive Standardised Quality Test for Glass Capillary Columns.* Journal of Chromatography A, 1978. **156**(1): p. 1 - 20.

8. Grob, K. Jr. and Grob, K., *Evaluation of Capillary Columns by Separation Number or Plate Number.* Journal of Chromatography A, 1981. **207**(3): p. 291 - 297.

9. Griffin, G.M.E., Doolan, K., Campbell, M., Hamill, J. and Kee, T.G., *Analysis of Wax Based Products by Capillary Gas Chromatogrpahy-Mass Spectrometry.* Science and Justice, 1996. **36**.

10. Ambros, A., Fuzesi, I., Lantos, J., Korsos, I., Szathmary, M. and Hatfaludi, T., *Application of TLC for Confirmation and Screening of Pesticides Residues in Fruits, Vegetables and Cereal Grains: Part 2. Repeatability and Reproducibility of Rf and MDQ Values.* Journal of Environmental Science and Health B, 2005. **40**: p. 485 - 511.

11. Ginemo, P., Besacier, F., Chaudron-Thozet, H., Girard, J. and Lamonette, A., *A Contribution to the Chemical Profiling of 3, 4 Methylenedioxymethamphetamine (MDMA) Tablets.* Forensic Science International, 2002. **127**: p. 1 - 44.

12. Soini, H.A., Bruce, K.E., Klouckova, I., Brereton, R.G., Penn, D.J. and Novothy, M.V., *In Situ Surface Sampling of Biological Objects and Preconcentration of Their Volatiles for Chromatographic Analysis.* Analytical Chemistry, 2006. **78**: p. 7161 - 7168.

APPENDIX C

APPENDIX C - LIST OF DATASETS

1. TLC data
2. MSP data
3. UV/Vis data
4. GC data
5. GC Normalised data
6. TLC data + MSP data
7. TLC data + UV/Vis data
8. TLC data + GC Normalised data
9. TLC data + GC data
10. MSP data + UV/Vis data
11. MSP data + GC Normalised data
12. MSP data + GC data
13. MSP Square Root
14. UV/Vis Square Root
15. GC Normalised Square Root
16. GC Square Root
17. TLC + MSP Square Root
18. TLC + UV/Vis Square Root
19. MSP + UV/Vis Square Root
20. UV/Vis + MSP Square Root
21. TLC + GC Normalised Square Root
22. TLC + GC Square Root
23. MSP + GC Normalised Square Root
24. MSP + GC Square Root
25. UV/Vis + GC Normalised Square Root
26. UV/Vis + GC Square Root
27. MSP Square Root + UV/Vis Square Root
28. MSP Square Root + GC Normalised Square Root

29. MSP Square Root + GC Square Root
30. UV/Vis Square Root + GC Normalised Square Root
31. UV/Vis Square Root + GC Square Root
32. MSP Fourth Root
33. UV/Vis Fourth Root
34. GC Normalised Fourth Root
35. GC Fourth Root
36. TLC + MSP Fourth Root
37. TLC + UV/Vis Fourth Root
38. MSP + UV/Vis Fourth Root
39. TLC + GC Normalised Fourth Root
40. TLC + GC Fourth Root
41. MSP + GC Normalised Fourth Root
42. MSP + GC Fourth Root
43. UV/Vis + GC Normalised Fourth Root
44. UV/Vis + GC Fourth Root
45. MSP Square Root + UV/Vis Fourth Root
46. MSP Square Root + GC Normalised Fourth Root
47. MSP Square Root + GC Fourth Root
48. UV/Vis Square Root + GC Normalised Fourth Root
49. UV/Vis Square Root + GC Fourth Root
50. UV/Vis Square Root + MSP Fourth Root
51. UV/Vis Fourth Root + GC Data
52. UV/Vis Fourth Root + MSP Fourth Root
53. UV/Vis Fourth Root + GC Normalised Fourth Root
54. UV/Vis Fourth Root + GC Fourth Root

APPENDIX D

Anal. Chem. **2010,** *82,* 6395–6400

Application of Unsupervised Chemometric Analysis and Self-organizing Feature Map (SOFM) for the Classification of Lighter Fuels

Wan N. S. Mat Desa,† Niamh Nic Daéid,* Dzulkiflee Ismail, and Kathleen Savage

Centre for Forensic Science, Department of Pure and Applied Chemistry, University of Strathclyde, 204 George Street, Glasgow G1 1WX

A variety of lighter fuel samples from different manufacturers (both unevaporated and evaporated) were analyzed using conventional gas chromatography−mass spectrometry (GC-MS) analysis. In total 51 characteristic peaks were selected as variables and subjected to data preprocessing prior to subsequent analysis using unsupervised chemometric analysis (PCA and HCA) and a SOFM artificial neural network. The results obtained revealed that SOFM acted as a powerful means of evaluating and linking degraded ignitable liquid sample data to their parent unevaporated liquids.

Ignitable liquids are commonly used as accelerants to intensify the rate of fire development in cases of deliberate fire setting. Identification and discrimination of ignitable liquid or ignitable liquid residues within recovered samples are, therefore, of interest and the characterization, identification, and linkage of such ignitable liquids is highly desirable. At present, gas chromatography-flame ionization detector (GC-FID) and gas chromatography-mass spectrometry (GC-MS) are widely accepted as effective methods for the analysis and identification of ignitable liquids.[1−3] Interpretation of the instrumental data and sample classification processes are based mainly on visual comparison of the sample to an ignitable liquid database.[4−6] Other approaches involve selective ion monitoring of the resultant chromatographic data producing a target compound chromatogram (TCC) of the sample which can be compared to TCC of the ignitable liquid from a database.[7−10]

These methods however can be difficult, time-consuming, highly subjective and rely heavily on the skill and experience of the analyst. Petrol for example, consists of a wide range of over 500 hydrocarbon compounds together with other additives.[11] Exposure to heat or aging can result in substantial changes in the liquid's composition, which in turn greatly affects their chromatographic profile. Another common complication encountered in fire debris analysis is the appearance of hydrocarbon byproduct from the combustion and pyrolysis of background matrices.[12−15]

The applications of multivariate pattern recognition to discriminate and classify ignitable liquid samples are suggested as a means of facilitating the pattern matching process and rendering it less subjective. Pattern recognition techniques using chemometric approaches have been used to establish underlying relationships among variables within complex data sets and can be used to differentiate groups within a given data set. Principle component analysis (PCA) and hierarchical cluster analysis (HCA), are both regarded as unsupervised-learning methods which do not require a training set of known categories to derive the classification model. Instead they use the given data to self-establish grouping structures.[16,17]

PCA and HCA have been employed for the classification of ignitable liquid analysis.[18−22]Typically, numerical data sets of selected compounds within the sample are processed by PCA, canonical variate analysis (CVA), and orthogonal canonical variate analysis (OCVA) coupled with linear discriminate analysis (LDA) for discrimination. Petraco et al. demonstrated the use of such methods for the discrimination of 20 gasoline samples, however some misclassification occurred.[22] The authors emphasized the

* To whom correspondence should be addressed. Phone: + 44-141-5484700. Fax: +44-141-5482532. E-mail: n.nicdaeid@strath.ac.uk.

† Permanent address: School of Health Sciences, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan, Malaysia.

(1) American Society for Testing and Materials Method E1618-01. In *Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography-Mass Spectrometry*; American Society for Testing and Materials: West Conshohocken, PA, 2001.
(2) American Society for Testing and Materials Method E1387-01. In *Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography*; American Society for Testing and Materials: West Conshohocken, PA, 2001.
(3) DeHaan, J. D., *Kirk's Fire Investigation*, 6th ed.; Brady: New York, 2007.
(4) Almirall, J. R.; Furton, K. G., *Analysis and Interpretation of Fire Scene Evidence.*; CRC Press: Boca Raton, 2004.
(5) Newman, R.; Gilbert, M.; Lothridge, K. *GC-MS Guide to Ignitable Liquids*; CRC Press: Boca Raton, 1998.
(6) Vella, A. J. *J. Forensic Sci. Soc.* **1992**, *32*, 131–142.
(7) Nowicki, J. *J. Forensic Sci.* **1990**, *35*, 1064–1086.
(8) Nowicki, J. *J Forensic Sci.* **1991**, *36*, 1536–1550.
(9) Lennard, C. J.; Rochaix, V. T.; Margot, P. *Sci. Justice.* **1995**, *35*, 19–30.
(10) Keto, R.; Wineman, P. L. *Anal. Chem.* **1991**, *63*, 1964–1971.
(11) Wang, Z.; Fingas, M. *J. Chromatogr., A* **1995**, *712*, 321–343.
(12) Bertsch, W. *J. Chromatogr., A* **1994**, *674*, 329–333.
(13) Dehaan, J.; Bonarius, K. *J. Forensic Sci. Soc.* **1988**, *28*, 299–309.
(14) Lentini, J. J., *Scientific Protocols for Fire Investigation*; Taylor & Francis Gp, CRC Press: Boca Raton, 2006.
(15) Lentini, J. J. *J. Forensic Sci.* **1998**, *43* (1), 97–113.
(16) Everitt, B. S.; Dunn, G., *Applied Multivariate Data Analysis.*; Arnold: London, 2001.
(17) Berrueta, L. A. e. a. *J. Chromatog. A* **2001**, *1158*, 196–214.
(18) Tan, B.; Hardy, J. K.; R.E., S. *Anal. Chim. Acta* **2000**, *422*, 37–46.
(19) Bodle, E. S.; J. K., H. *Anal. Chim. Acta* **2007**, *589*, 247–254.
(20) Doble, P.; al., e., 2003, *132*, 26−39.
(21) Lu, Y.; Harrington, P. B. *Anal. Chem.* **2007**, *79*, 6752−6759.
(22) Petraco, N. D. K.; et al. *J. Forensic Sci.* **2008**, *53*, 1092−1101.

10.1021/ac100381a
Published on Web 07/02/2010

importance of statistical methods of pattern recognition for fire debris casework rather than relying on visual comparison alone.

The application of artificial neural networks (ANNs) in forensic data analysis on the other hand, is relatively unusual. ANN has been suggested as an effective data analysis mechanism for various data sets relating to cocaine analysis,[23] toolmark comparison[24] and gasoline identification.[20] Doble et al. demonstrated the application of supervised ANNs using a multilayer perceptron (MLP) arrangement with a back-propagation (BP) algorithm to discriminate between regular and premium gasoline and reported a 97% correct classification-prediction rate.[20] To date no other application of ANN in ignitable liquid analysis has been reported and there is no reported application of self organizing feature map (SOFM) (or Kohonen mapping).[25,26]

SOFM is typically arranged in a two-layer form consisting of an input and output layer. The output neurons of the network are normally arranged as a two-dimensional grid where the multidimensional input vectors, (the variables describing the data), are mapped. The input neurons are fully connected to every neuron in the output layer. Associated with every connection is a weighting, which influences the final output of the network. The network training in SOFM is a competitive process where all the neurons compete to be stimulated by the input vectors. The similarities between the input and output vectors are computed using the Euclidean distance. The neuron with weight vectors which is most similar to the input vector is chosen as the best matching neuron (BMU). The weights of the BMU's and those of neighboring neurons are adjusted so that they will become closer to the input vectors in the next iteration of the algorithm. As the training process continues, the SOFM organizes into a state whereby similar input vectors are mapped onto similar neurons on the output layer. By the end of the training process, the output neuron is labeled according to the input (or data) that has stimulated or mapped onto it in order to reveal if clustering has emerged from the training.

In ignitable liquid analysis, the relevant chromatographic data are isolated by targeting compounds identified as discriminating for that particular sample.[22,27−29] The selected data may also undergo further transformations or data pretreatment (such as normalization, logarithmic transformation and root transformation), which are commonly used to minimize the effect of large peaks or eliminate signal noise in order to make the data distribution more symmetrical and facilitate multivariate analysis.[25,30−33] In this work chromatographic data derived from various brands of unevaporated and partially evaporated lighter fuel was subjected to chemometric analysis using unsupervised methods (PCA, HCA) and an SOFM artificial neural network. The goal was to reveal the most appropriate preprocessing methods for this type of data and to determine the feasibility of using unsupervised classification techniques, particularly SOFM for pattern recognition of chromatographic data with a view to assignation of partially evaporated materials with their corresponding unevaporated source.

EXPERIMENTAL SECTION

Chemicals and Materials. Fifteen refill lighter fluid samples from five different brands (Zippo, Swan, Ronsonol, Perma, and Dunhill) were purchased from commercial outlets. For each lighter fluid brand, a set of partially evaporated samples was generated. This was achieved by gently heating the unevaporated lighter fluid (100 mL) and removing a sub sample at specific intervals when the original liquid had been reduced by 10, 25, 50, 75, 90, and 95 mL in volume. This produced a set of partially evaporated samples for each lighter fluid liquid at approximately 10, 25, 50, 75, 90, and 95% evaporation. Each sample was diluted to 2% in pentane (HPLC grade, WVR International) with 0.5 mg/mL Tetrachloroethylene (Sigma Aldrich, >99%) as internal standard. All samples were stored in screw capped vials and darkness at room temperature until analyzed. An aluminum foil shield was placed inside the screw cap to prevent evaporation of the sample. The prepared samples were at the approximate percentages of evaporation of interest and as such, any additional evaporation which may have occurred during storage was considered to be of little effect to the overall results obtained.

According to ASTM E1618 guidelines, pentane is listed as one of the recommended nonpolar solvents for fire debris analysis and is as efficient as other solvents but relatively less toxic, safer and easier to handle.[1,34] The solvent delay on the MSD was set in order to allow the pentane peak to elute undetected and this had a minimal effect on the overall patterns of interest in the test samples, particularly in relation to the evaporated samples.

The application and efficiency of using volatile chlorinated compound as the internal standard for fire debris analysis, such as Tetrachloroethylene has been reported.[35,36] A chlorinated compound was used mainly because it is not commonly encountered in ignitable liquid samples and is cost-effective and easily identifiable.

Instrumentation. Gas chromatographic analysis was performed using a Hewlett-Packard (HP) 6890/5973 gas chromatograph with a mass selective detector (GC-MSD). Data acquisitions were performed by MS ChemStation (version B.00.01, Hewlett-Packard, Agilent Technologies). Chromatographic separation was achieved using a DB1-MS fused silica capillary column (25.0 m × 0.20 mm i.d. × 0.33 μm film thickness). The injection port temperature set at 250 °C. The oven temperature was set at 40 °C for 5 min and ramped at a rate of 15 °C/min to 280 °C, and

(23) Casale, J. F.; Watterson, J. W. *J. Forensic Sci.* **1993**, *38*, 292–301.
(24) Kinston, C. *J. Forensic Sci.* **1992**, *37*, 252–264.
(25) Brereton, R. G., *Chemometrics for Pattern Recognition.*; John Wiley and Sons: West Sussex, U.K., 2009.
(26) Kohonen, T. *Self-Organising Maps*; Springer-Verlag: Heidelberg, Germany, 1995.
(27) Borusiewicz, R.; Zadora, G.; J., Z. b.-P. *J. Chromatogr. (Supplement)* **2004**, *60*, S133–S142.
(28) Sandercock, P. M. L.; Pasquier, E. D. *Forensic Sci. Int.* **2003**, *134*, 1–10.
(29) Sandercock, P. M. L.; Pasquier, E. D. *Forensic Sci. Int.* **2004**, *140*, 43–59.
(30) Wold, S.; Sjostrom, M.; Erikksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
(31) Rietjens, M. *Anal. Chim. Acta* **1995**, *316*, 205–215.
(32) Andersson, K. e. a. *Forensic Sci. Int.* **2007**, *169*, 86–99.
(33) Stauffer, E.; Dolan, J. A.; Newman, R. *Fire Debris Analysis*; Academic Press: New York, 2008.
(34) Stauffer, E. *Sci. Justice* **2003**, *43* (1), 29–40.
(35) Locke, A. K.; Basara, G. J.; Sandercock, P. M. L. *Journal Forensic Science* **2009**, *54* (2), 320–327.
(36) *NIST/EPA/NIH Mass Spectral Library (NIST 08) and NIST Mass Spectral Search Program (Version 2.0f)*: U.S. Department of Commerce, National Institute of Standards and Technology (NIST), Standard Reference Data Program, Gaithersburg, MD 20899: 2008.

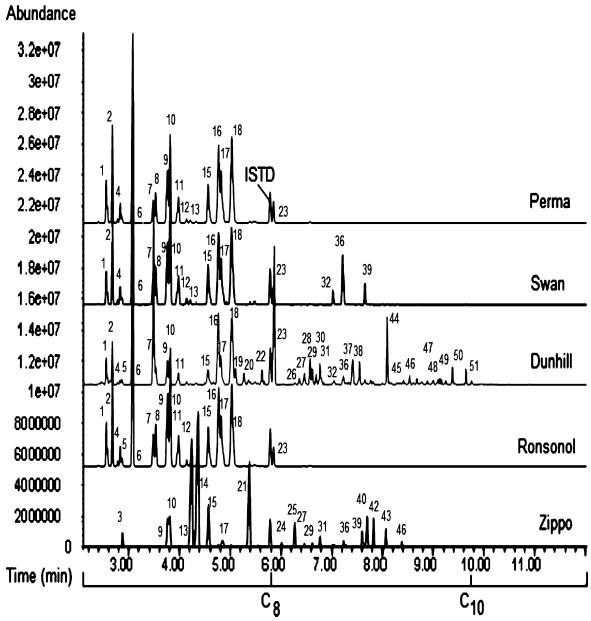


Figure 1. Chromatograms of pure lighter fluid samples (2% in pentane with 0.5 mg/mL Tetracholoroethylene as ISTD).

maintained at this temperature for 2 min. Helium was used as the carrier gas and was maintained at a constant flow rate of 1.0 mL/min. The temperatures of the ion source and the quadrupole were set at 150 and 280 °C, respectively. The MS analyses were performed at full scan mode (from 30 to 300 amu) with two minutes solvent delay. Injections were carried out using a 7673A Hewlett-Packard automatic liquid sampler. Each sample was analyzed in triplicate and the injection volume for each sample was 1 μL with a 20:1 split ratio.

Data Collection and Preprocessing. The chromatograms were visually compared and components of similar retention time and relative standards deviation of less than 5% on triplicate analysis were selected. Peak response data was acquired as comma separated values (CSV) files and converted into an Excel (version 10) spread-sheet for ease of use. In total, 51 components were selected as variables from the pure and evaporated samples. Missing components in any individual sample were given a zero value. The peak area response were normalized against the internal standard. Further data pre processing was undertaken by applying a square root transformation and a fourth root transformation to the normalized data set. The four resultant data sets (i.e., raw and processed) were inputted to MATLAB 2008a (version 7.6, Mathwork Inc.) where principal component analysis (PCA) and hierarchical cluster analysis (HCA) were performed. SOFM-artificial neural network analysis was performed using ViscoverySOMine (version 5.0.2, Viscovery Software GmbH).

RESULTS AND DISCUSSION

Chromatographic Analysis. Initial examination of the chromatographic pattern of each lighter fluid sample from each of the five brands revealed compositional differences for Swan, Dunhill, and Zippo samples. By contrast Perma and Ronsonol revealed chromatographic patterns which were very similar to each other as shown in Figure 1. Identification of the components common in the various lighter fluid samples is presented in Table 1. Identification is made based on NIST mass spectral database using AMDIS software (version 2.0).[37]

Table 1. List of Components Identified in Lighter Fuel Samples

no.	RT	peak identification
1	2.55	2-methylhexane
2	2.68	3-methylhexane
3	2.84	3-ethylpentane
4	2.89	2,2,4-trimethylpentane
5	3.08	heptane
6	3.47	1,2-dimethylcyclopentane
7	3.48	cyclomethylhexane
8	3.53	2,2-dimethyl-3-hexene
9	3.77	2,5-dimethylhexane
10	3.81	2,4-dimethylhexane
11	3.99	3,3-dimethylhexane
12	4.15	1,2,3-trimethylcyclopentane
13	4.23	2,3,4-trimethylpentane
14	4.37	2,3,3-trimethylpentane
15	4.56	2,3-dimethylhexane
16	4.77	2-methylheptane
17	4.82	4-methylheptane
18	5.03	3-methylheptane
19	5.11	1,4-dicyclohexane
20	5.28	1,1-dicyclohexane
21	5.37	2,2,4-trimethylhexane
22	5.64	1,2-Dicyclohexane
23	5.85	octane
24	6.01	2,4,4-trimethylhexane
25	6.28	2,3,5-trimethylhexane
26	6.36	2,2-dimethylheptane
27	6.47	2,4-dimethylheptane
28	6.58	ethylcyclohexane
29	6.63	2,6-dimethylheptane
30	6.70	1,1,3-trimethylcyclohexane
31	6.78	2,5-dimethylheptane
32	7.02	2,3,4-trimethylhexane
33	7.04	3-methylheptane
34	7.06	1,3,5-trimethylcyclohexane
35	7.21	*m*-xylene
36	7.23	2,3-dimethylheptane
37	7.43	2-Methyl octane
38	7.56	3-methyl octane
39	7.66	*p*-xylene
40	7.71	2,3,6-trimethylheptane
41	7.78	1-ethyl-4-methylcyclohexane
42	7.83	2,2,4-trimethylheptane
43	8.07	3,3-dimethyloctane
44	8.09	nonane
45	8.30	3,4-dimethyloctane
46	8.41	2,4-dimethyl-3-ethylpentane
47	8.54	cis-1,1,3,5-tetramethylcyclohexane
48	8.69	2,6-dimethyloctane
49	9.12	1-ethyl-1,3-dimethylcyclohexane
50	9.15	toluene
51	9.65	decane

The effect of evaporation of an ignitable liquid is 2-fold - lower boiling compounds diminish or are lost completely while higher boiling compounds increase in their abundance relative to neighboring compounds, and as such the chromatographic profiles of the evaporated samples are different to those of pure samples,[5,38]. The exemplar of this is illustrated in Figure 2 which clearly demonstrates substantial changes in peaks abundance.

PCA Classification. PCA was performed on each data set (raw data, normalized data only, normalized square root and normalized

(37) Newman, R. *Modern Laboratory Techniques Involved in the Analysis of Fire Debris Samples*; CRC Press: Boca Raton, 2004.

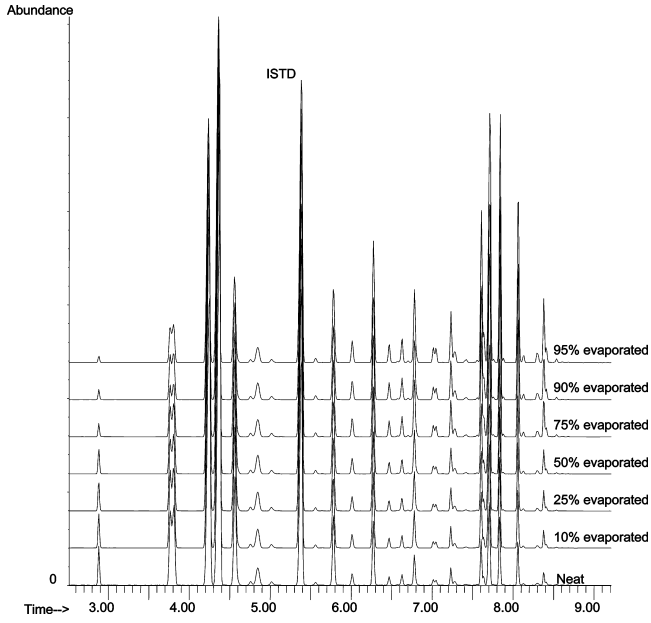


Figure 2. Chromatograms of the pure Zippo lighter fluid and evaporated Zippo lighter fluid to demonstrate gradual changes in major peaks abundance as the sample evaporated.

fourth root) to assess the effectiveness of the method to distinguish between the various lighter fluid brands and whether it was possible to establish a link between the pure and evaporated samples of a specific lighter fluid brand. The score plots obtained when all data sets were subjected to PCA analysis are given in Figure 3. Each score plot obtains a series of principal components (PC) where these values represent variances within the data sets.

For all of data sets analyzed only two brands of lighter fluid were successfully resolved irrespective of the data pretreatment method used.

HCA Classification. HCA was performed using the same four data sets used in PCA. Similarities between the samples were measured using complete linkage and the results are shown in Figure 4.

Like PCA, HCA did not correctly classify the samples by brand when using the raw data, normalized or normalized square root data sets. However, in contrast to PCA, the normalized fourth root data set produced a HCA classification which was capable of separating all of the pure and evaporated samples by brand and is shown in Figure 4 (D). The HCA classification suggests similarities between Swan and Perma brand samples as well as similarities between both of these samples and those of the Dunhill and Zippo brands, however this was not reflected in the visual comparison of the chromatographic profiles of these samples where clear differences were in evidence. The Ronsonol sample set was clustered away from all other sample sets including

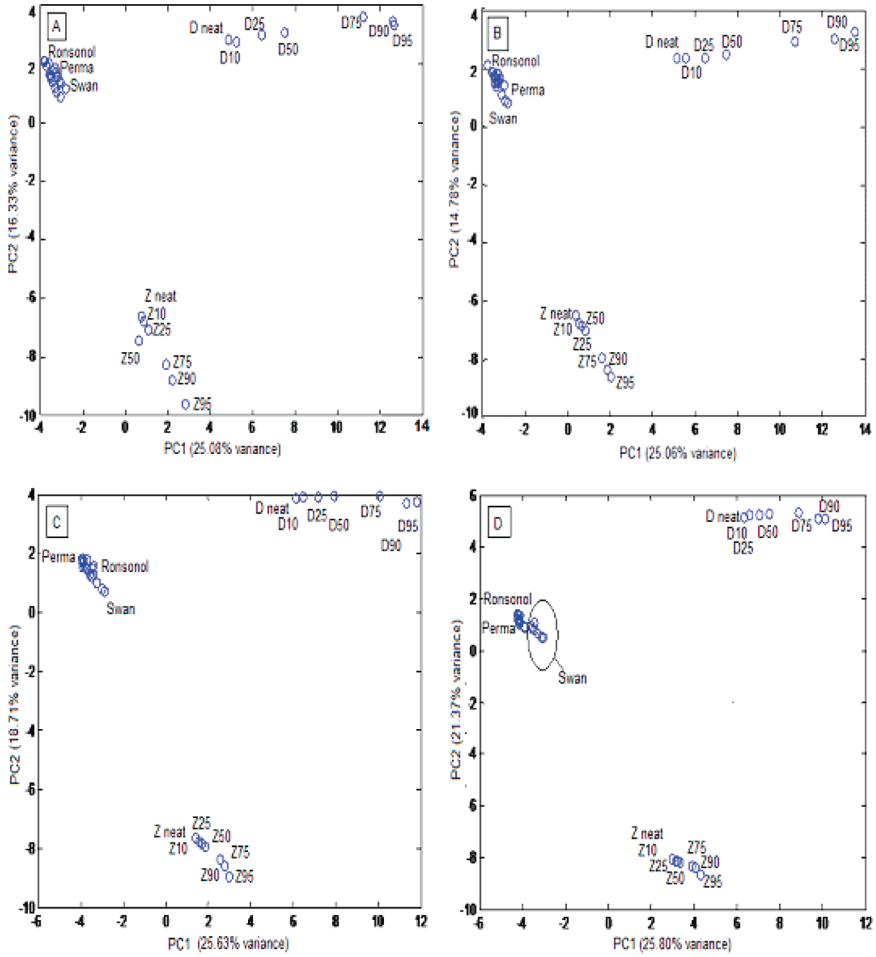


Figure 3. Principal component score plots of pure and evaporated samples. A, B, C, and D represent plots of raw data, normalized, normalized square root and normalized fourth root transformation data sets respectively. (*D* = Dunhill, *P* = Perma, *R* = Ronsonol, *S* = Swan, and *Z* = Zippo. Numbers represent degree of evaporation).

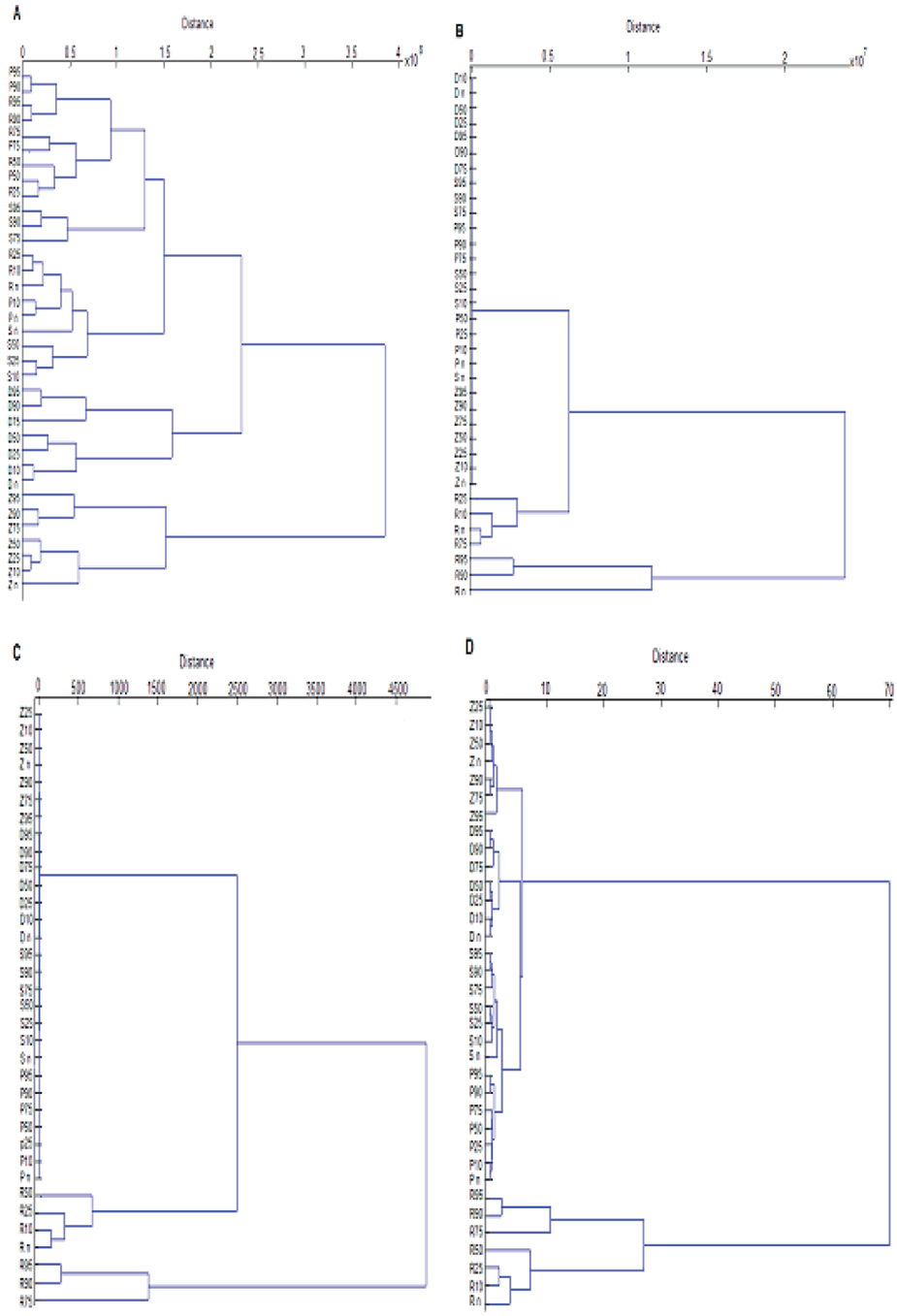


Figure 4. Hierarchical clustering of pure and evaporated samples. A, B, C, and D represent dendrograms of raw data, normalized, normalized square root and normalized fourth root transformation data sets respectively. (*D* = Dunhill, *P* = Perma, *R* = Ronsonol, *S* = Swan, and *Z* = Zippo. Numbers represent degree of evaporation).

the Perma sample set even though their chromatographic profiles were visually similar.

SOFM Classification. The ability of SOFM to cluster related samples using the four data sets is shown in Figure 5. Data visualization in SOFM can be accomplished by a number of techniques which can be a hit histogram, component planes and the U-Matrix display.[25] The one that is shown in this paper follows the U-Matrix display where the groupings and the degree of dissimilarity between groupings are shown by the more intensely colored boundary lines, where darker boundary lines reflect a greater distance between the adjoining samples.

As a result of using the U-Matrix visualization method, groupings within and between sample brands and the cluster presentation was straightforward and easy to understand. The effectiveness of SOFM over both PCA and HCA is demonstrated by its ability to correctly classify samples using the normalized and preprocessed data sets, although as with HCA the best result was observed with normalized forth root data. In addition SOFM clearly illustrates samples that have similar chromatographic profiles (such as Ronsonol and Parma) by placing these sample clusters in close proximity with each other. This was not achievable using HCA.

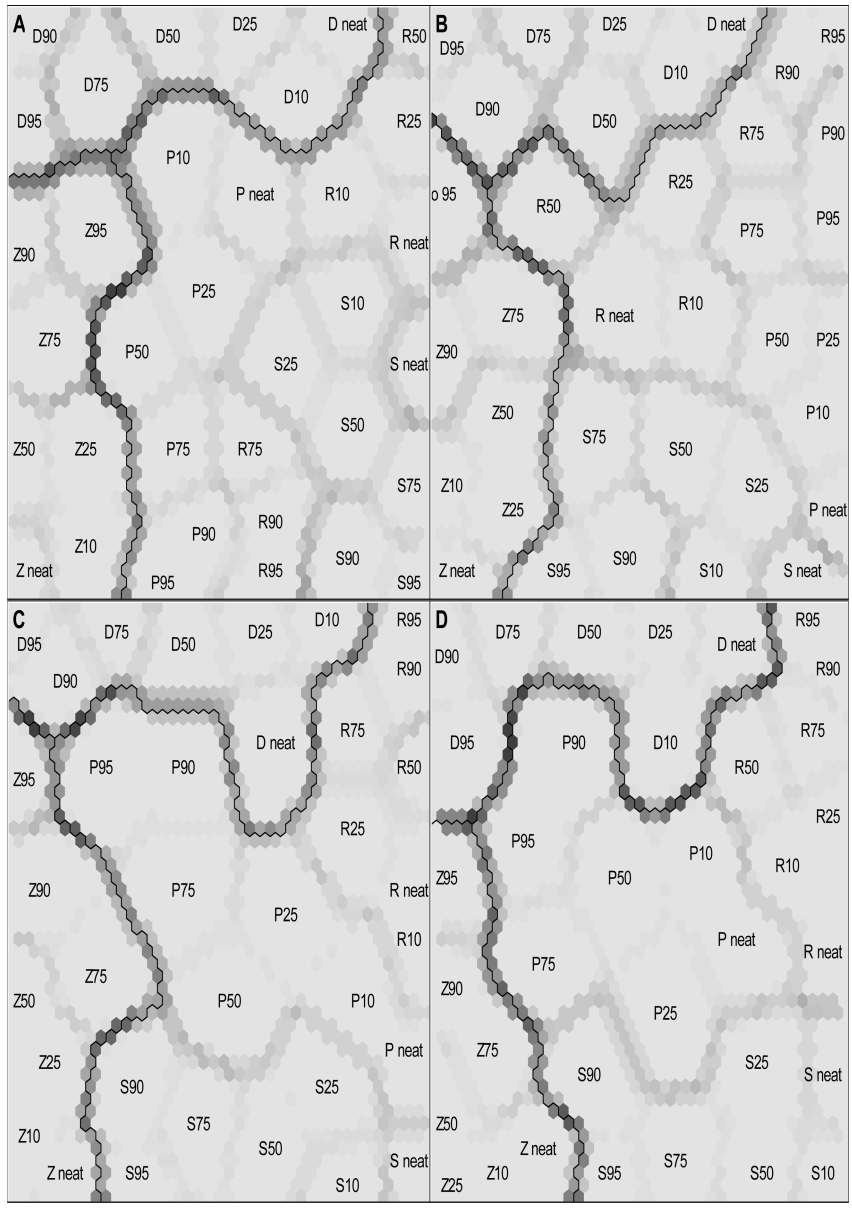


Figure 5. SOFM topographic maps of pure and evaporated samples. A, B, C, and D represent the maps of raw data, normalized, normalized square root and normalized fourth root transformation data set respectively. (*D* = Dunhill, *P* = Perma, *R* = Ronsonol, *S* = Swan, and *Z* = Zippo. Numbers represent degree of evaporation).

CONCLUSIONS

The results of this study demonstrate that pretreated chromatographic data from analyzed lighter fuel samples which have been evaporated to varying degrees can be interrogated using chemometric methods and successfully linked back to their parent pure ignitable liquid sample. Data pretreatment was essential in order to obtain accurate classifications. Three methods of data pretreatment were applied and the best overall discrimination within and between samples was achieved with a normalized fourth root transformation data. It was determined that successful sample classification (by brand) was only achieved using HCA and SOFM where SOFM proved to have a more robust sample linkage capacity and confirmed visual similarities and differences between the samples in evidence in the chromatograms. This has demonstrated a potential means whereby pure and evaporated ignitable liquid samples can be linked successfully by brand and as such presents a powerful new means of interpreting chromatographic data retrieved from fire debris samples. Further studies in this area are ongoing within our laboratory.

ACKNOWLEDGMENT

We acknowledge the Ministry of Higher Education of Malaysia (SLAB Grant) which provided studentship to W.M.-D. and D.I.

Received for review February 10, 2010. Accepted June 23, 2010.

AC100381A

APPENDIX E

Comparison of Smears of Wax Based Products using Thin Layer Chromatography (TLC) and Microspectrophotometric (MSP) Detection

Dzulkiflee Ismail[1, 2] and Niamh Nic Daeid[1,*]

1. Centre for Forensic Science, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow G1 1XW, UK.
2. Fakulti Sains Kesihatan, Jabatan Sains Forensik, Universiti Sains Malaysia (USM), 16150, Malaysia.

Author for correspondence:

Niamh Nic Daeid
Centre for Forensic Science,
Department of Pure and Applied Chemistry,
University of Strathclyde,
Glasgow G1 1XW,
UK.
n.nicdaeid@strath.ac.uk

Abstract:

This work introduces a rapid and effective technique for the discrimination of smears of coloured wax based products (such as lipstick and shoe polish) on fabrics. Forty two samples of commonly available wax based products were analysed. The analytical techniques used were a combination of thin layer chromatography (TLC) and direct microspectrophotometry (MSP) of the subsequent TLC plate. The resultant data was analysed using Self Organising Feature Mapping, an Artificial Neural Network system. The combination of both TLC and MSP facilitated the discrimination of all samples and the SOFM system provided an easy to understand visual representation of the sample discrimination by type.

Introduction:

In forensic examinations of garments, wax-based products are sometimes found as small stains or smears [1, 2]. The diversity of wax-based products available to the consumer at any given time often makes the determination of even class characteristics ambiguous and the development of an analytical methodology which reduces this ambiguity and produces a quick sample turnaround time would be advantageous.

Wax based products, as their name suggests are commercially available materials which contain various types of waxes as the major constitutes of the product. In consumer products such as lipstick and shoe polishes, the waxes are mostly of natural origin (paraffin wax, carnauba wax, candelilla wax and beeswax). Such products also contain other materials such as dyes, pigments, perfumes and solvents which are added in varying proportion to the waxes [3-5].

Previous studies in this area are very limited and have tended to concentrate on analytical techniques such as gas chromatography mass spectroscopy (GCMS) to characterise the samples. GCMS requires the sample to be extracted, often with reagents such as dichloromethane (DCM), chloroform or tetrahydrofuran (THF) which can promote deterioration of the GC column. The nature of the samples, often containing

long chain alkanes can also adhere to the GC column, shortening column life and functionality. Furthermore, GCMS can be time consuming and expensive [1, 2, 6].

This work, by contrast, has made use of a simple low cost method for the discrimination of wax based products which involves the examination of the samples using thin layer chromatography (TLC) followed by microspectrophotometry (MSP) of the eluted TLC plate. The combination of basic data with powerful chemometric data analysis has provided a means to definitively discriminate all of the wax based samples from each other.

The Self Organising Feature Maps (SOFM) neural network was first introduced by Kohonen in 1982 [7-9]. SOFM is an unsupervised pattern recognition technique and does not require any prior knowledge of the output patterns. The technique uses inputted data to determine, for example class membership where the membership of each class or variable is defined by the users and is an input pattern that shares common identifiable features [10]. The SOFM neural network is normally arranged in a two-layer system which consists of an input and an output layer. The output layer is normally arranged in a two-dimensional grid consisting of a number of units or neurons where the input patterns describing the objects are mapped. The input neurons in the input layer are all fully connected to each neuron in the output layer.

Associated with the output map are component maps whose number corresponds to the number of variables used in the training of the SOFM. The component map is commonly used to study the relationship of each variable such that if groupings exist after training of the SOFM, the responsible variables can be identified [11]. The arrangement of the component maps in SOFM can be visualised as multi-tier planes as illustrated in Figure 1.

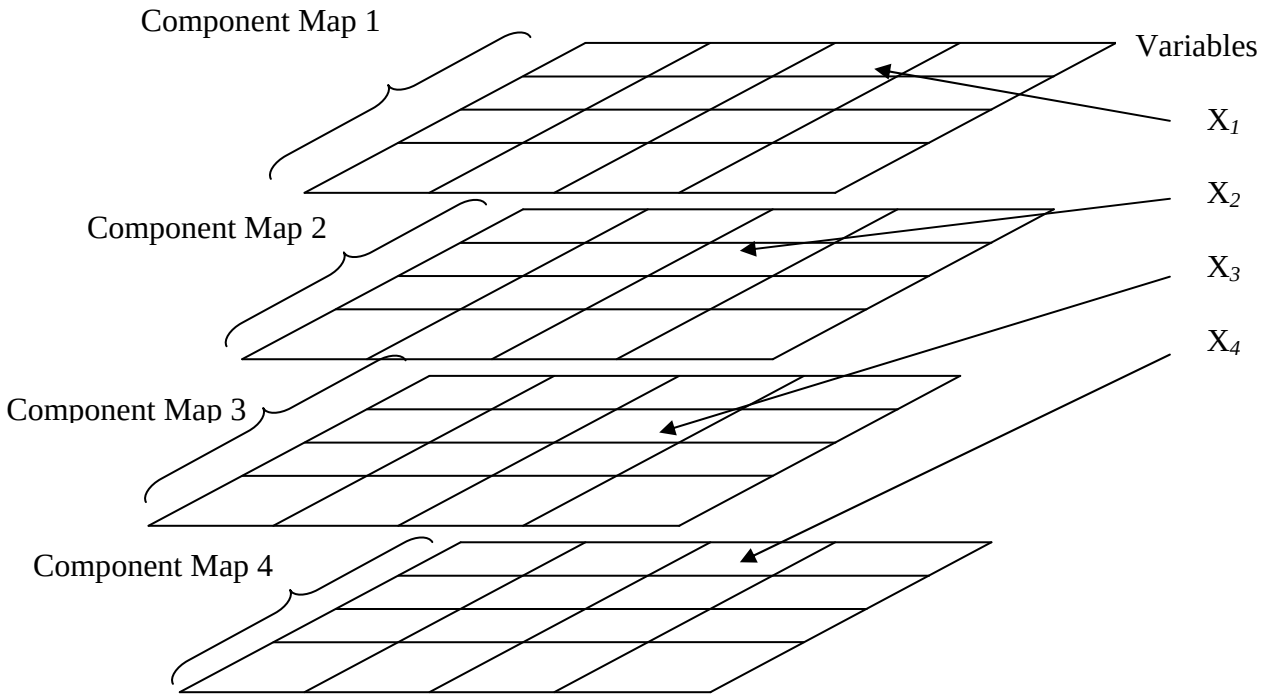


Figure 1 Arrangement of component maps in SOFM

Materials and Methods:

In total 17 shoe polish and 25 lipstick samples were examined. In the initial phase of the study 6 samples (three lipstick and three shoe polish samples) were selected in order to optimise the extraction and analytical techniques. All 42 samples (presented in Tables 1 and 2) were then analysed using the optimised system. The generated data was analysed using self organising feature map (SOFM) using Viscovery SOMine® 4.0.2 (Eudaptics) software.

Table 1 List of lipstick (L) samples analysed in the study

Sample Name	Type	Colour	Code
Star Gazer Lipstick 133	L	Purple	L1
Star Gazer Lipstick 105	L	Blue	L2
Star Gazer Lipstick 110	L	Black	L3
Clinique Super Spice	L	Brown	L4
Jane Seymour Pineapple Pink	L	Orange	L5
Body Collection Crushed Rose	L	Brown	L6
RIMMEL London Rich Raisin	L	Brown	L7
Clinique Raspberry Glace	L	Brown	L8
Collection 2000 Advance Colour Cream	L	Brown	L9
Boots Mulberry	L	Light Brown	L10
Estee Lauder French Fig	L	Brown	L11
Amway Debut Peach	L	Orange	L12
Clinique Ginger Flower	L	Red	L13
Estee Lauder All Day Carol Melon	L	Red	L14
Dare Dolly	L	Orange	L15
Apricot Crush	L	Orange	L16
Nutmeg	L	Brown	L17
Red Carnation	L	Red	L18
Marie France Touch of Spice	L	Brown	L19
Revlon Moondrops	L	Orange	L20
TESCO Shade 12	L	Red	L21
TESCO Shade 3	L	Brown	L22
TESCO Shade 13	L	Dark Brown	L23
Max Factor Sunset Rose	L	Red	L24
Body Collection Mango Mood	L	Orange	L25

Table 2 List of shoe polish samples (SP) analysed in the study.

Sample Name	Type	Colour	Code
Kiwi Shoe Polish Dark Tan	SP	Tan	SP1
Kiwi Shoe Polish Black	SP	Black	SP2
Kiwi Shoe Polish Blue	SP	Blue	SP3
Kiwi Shoe Polish Red	SP	Red	SP4
Granger's Shoe Polish Brown	SP	Brown	SP5
Granger's Shoe Polish Black	SP	Black	SP6
Cherry Blossom Shoe Polish Tan	SP	Tan	SP7
Safeway Shoe Polish Dark Tan	SP	Tan	SP8
ASDA Shoe Polish Brown	SP	Brown	SP9
ASDA Shoe Polish Neutral	SP	Colourless	SP10
Kiwi Shoe Polish Neutral	SP	Colourless	SP11
Granger's Shoe Polish Neutral	SP	Colourless	SP12
Punch Taupe Shoe Cream	SP	Light Black	SP13
Meltonian Shoe Cream Black	SP	Black	SP14
Cherry Blossom Iris Blue	SP	Blue	SP15
Meltonian Shoe Cream Navy Blue	SP	Blue	SP16
Tuxan Shoe Care	SP	Brown	SP17

Optimisation of the sample Extraction and TLC Technique:

Three lipstick samples (red, brown and black) and three shoe polish samples (red, brown and black) were chosen in order to optimise the sample extraction and TLC systems. The samples were chosen because they were products of the same colour and could not be distinguished from each other on the basis of visual examination alone.

The extraction method and solvent system were optimised using a series of experiments where the solvent, method and time of agitation of the sample within the solvent and the choice of TLC mobile phase were systematically altered. After testing various extraction systems the optimized conditions for extraction were determined to be the emersion of a smear of the sample prepared on clean white cotton (1cm x 1cm) in 1mL of DCM and sonicated for 5minutes.

The optimized TLC mobile phase was determined to be a mixture of chloroform: methanol: water (5:1.5:0.2 v/v/v) and a double development technique was applied with 1 cm and 2 cm as the first and second development distance respectively.

TLC Analysis:

Smears of each of the 42 samples were placed onto clean white cotton swatches. A small sample of the cotton swatch (1cm x 1cm) containing the smear was removed using a scalpel and transferred into a glass vial containing 1 mL of DCM and sonicated for 5 minutes to extract any dyes present in the smear. After sonication, the extract was spotted onto a TLC plate (5 cm x 5 cm) using fine capillary tubes. The prepared TLC plates were eluted and the retardation factor (R_f) of any developed spots were calculated for the first and second development. Each sample was analysed six times to assure good reproducibility.

MSP Analysis:

The coloured spots revealed upon development of the TLC plates were examined using a Cavendish Instruments microspectrophotometer attached to an Olympus BX41

microscope (at X10 magnification). The MSP was linked to a PC and the spectra captured using Onyx software (version 1.9.0.0). The reflectance spectrum of each spot was recorded by placing the TLC plate on the microscope's sample stage and fastened with double sided tape to prevent movement of the plate. Reflectance spectra acquisition settings were set to be in the range of 362 – 780 nm with 1500 integration time.

Results and Discussions:

TLC analysis:

The TLC results of the samples are presented in Table 3 and 4. In each case an average value of the six sample repeats have been presented.

Table 3 TLC results for lipstick samples (each sample was analysed 6 times and the average result is presented)

SAMPLE	Colour	SPOT COLOUR	AVERAGE		SD		RSD%	
			1st Dev.	2nd Dev.	1st Dev.	2nd Dev	1st Dev.	2nd Dev.
L1	Purple	ND	-	-	-	-	-	-
L2	Blue	ND	-	-	-	-	-	-
L3	Black	ND	-	-	-	-	-	-
L4	Brown	red	-	0.12	-	0.01	-	8.08
L5	Orange	orange	0.85	0.87	0.03	0.02	3.72	1.78
L6	Brown	red	-	0.13	-	0.01	-	8.43
L7	Brown	red	-	0.15	-	0.01	-	4.96
L8	Brown	red	0.10	0.13	0.00	0.01	0.00	8.43
L9	Brown	pink	-	0.16	-	0.01	-	7.38
L10	Light Brown	pink	-	0.17	-	0.01	-	7.74
L11	Brown	red	-	0.16	-	0.01	-	7.38
L12	Orange	red	-	0.16	-	0.01	-	8.36
L13	Red	red	0.10	0.13	0.00	0.01	0.00	7.47
L14	Red	red	-	0.11	-	0.01	-	9.68
L15	Orange	orange	0.24	0.32	0.01	0.02	3.36	7.72
L16	Orange	orange	0.83	0.85	0.04	0.02	4.90	1.88
L17	Brown	orange	0.83	0.84	0.04	0.02	4.90	2.43
L18	Red	orange orange	0.24 0.45	0.31 0.48	0.01 0.03	0.03 0.02	4.24 7.03	9.39 3.81
L19	Brown	red	-	0.14	-	0.01	-	9.07
L20	Orange	orange orange	0.25 0.90	0.33 0.85	0.02 0.03	0.02 0.01	6.45 2.87	6.89 1.29
L21	Red	pink	-	0.14	-	0.01	-	4.06
L22	Brown	pink	-	0.12	-	0.01	-	9.61
L23	Dark Brown	pink	-	0.14	-	0.01	-	7.56
L24	Red	orange	0.28	0.37	0.03	0.02	9.96	4.19
L25	Orange	red	-	0.12	-	0.01	-	9.61

ND = not detected

Table 4 TLC results for Shoe Polish samples

SAMPLE	Colour	SPOT COLOUR	AVERAGE		SD		RSD%	
			1st Dev.	2nd Dev.	1st Dev.	2nd Dev.	1st Dev.	2nd Dev.
SP1	Tan	pink	0.48	0.41	0.03	0.03	5.77	7.81
SP2	Black	black	0.45	0.54	0.02	0.02	4.35	3.70
SP3	Blue	blue	0.55 0.88	0.56 0.87	0.05 0.03	0.01 0.02	9.96 3.13	0.99 1.90
SP4	Red	red	0.87	0.86	0.03	0.03	3.66	3.20
SP5	Brown	black	0.91	0.88	0.05	0.02	5.41	2.78
SP6	Black	black	0.50	0.53	0.04	0.03	8.94	5.59
SP7	Tan	ND	-	-	-	-	-	-
SP8	Tan	orange	0.66	0.85	0.04	0.02	5.72	2.88
SP9	Brown	orange	0.88	0.87	0.04	0.02	4.78	2.85
SP10	Colourless	ND	-	-	-	-	-	-
SP11	Colourless	ND	-	-	-	-	-	-
SP12	Colourless	ND	-	-	-	-	-	-
SP13	Light Black	orange	0.88	0.87	0.05	0.02	5.85	2.46
SP14	Black	black	0.54	0.30	0.04	0.03	6.95	9.21
SP15	Blue	ND	-	-	-	-	-	-
SP16	Blue	pink	0.85	0.86	0.03	0.02	3.72	1.80
SP17	Brown	ND	-	-	-	-	-	-

ND = not detected

Twenty two out of the twenty five lipstick samples and eleven out of the seventeen shoe polish samples produced coloured spots when analysed using TLC.

The data illustrates that some of the samples analysed could be distinguished on the basis of TLC alone however this is not universally the case.

MSP analysis:

The thirty three samples which produced coloured spots on TLC were analysed using MSP. The MSP analysis was carried out directly on the spots developed on the TLC plates. There were two regions of variability within the MSP spectra of the various samples at 450 nm and between 550 – 600 nm which facilitated sample discrimination. Example spectra are presented in Figures 2 and 3 where the regions of interest have been highlighted.

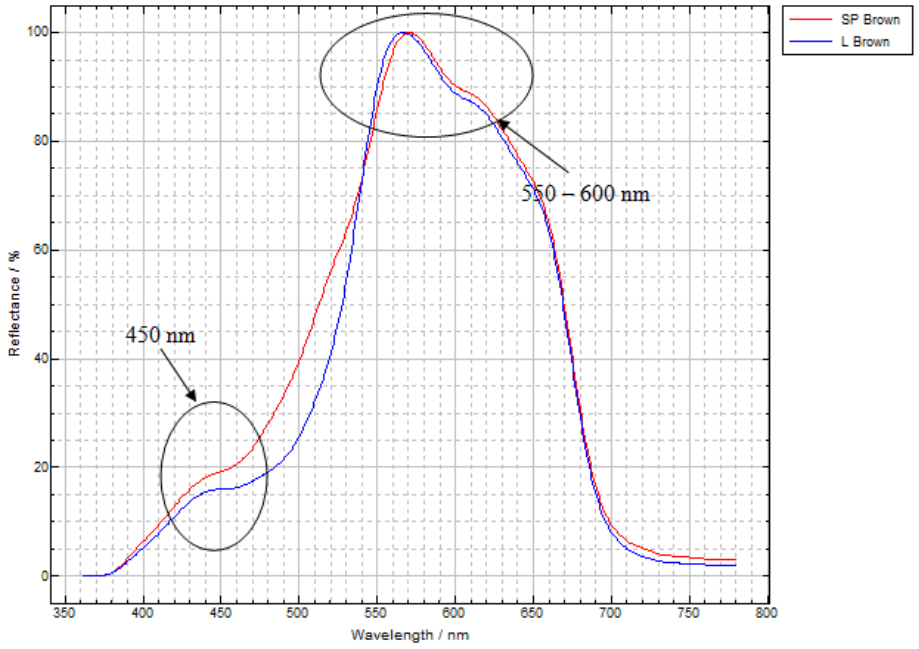


Figure 2 Example of reflectance spectra of brown wax based products (SP = Shoe Polish, L= Lipstick)

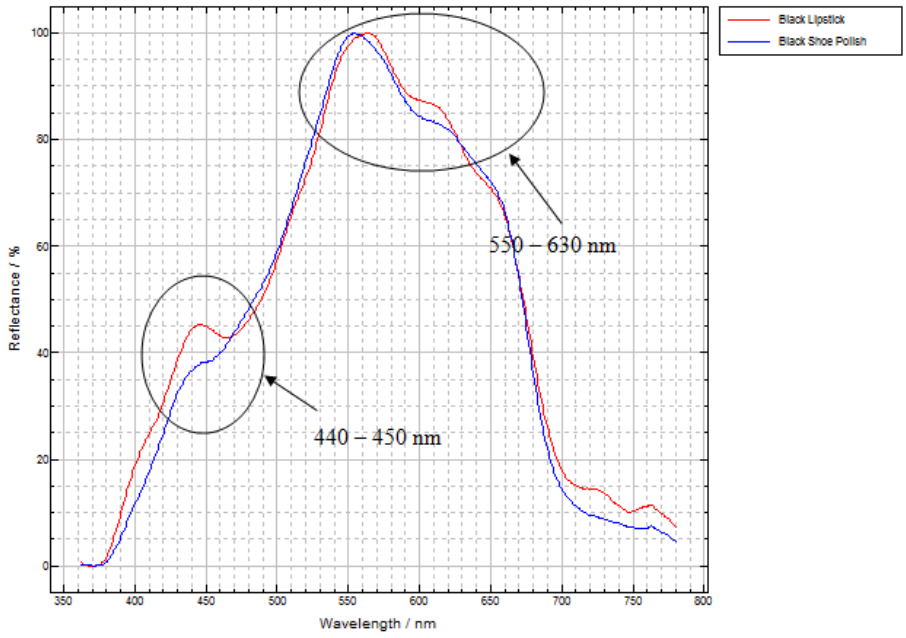


Figure 3 Example of reflectance spectra of black wax based products (SP = Shoe Polish, L= Lipstick)

SOFM Data Analysis:

The SOFM output map for the combined TLC and MSP dataset is illustrated in Figure 4 and clearly separates the samples into two distinct groupings based upon their product types. The first group occupying the bottom neurons of the output map group all of the lipstick samples whereas the entire shoe polish samples are grouped in the top portion of the output map. This clearly demonstrates that the use of SOFM demonstrates that the combination of TLC data and MSP spectra can discriminate the samples by type where as each technique failed to do this individually.

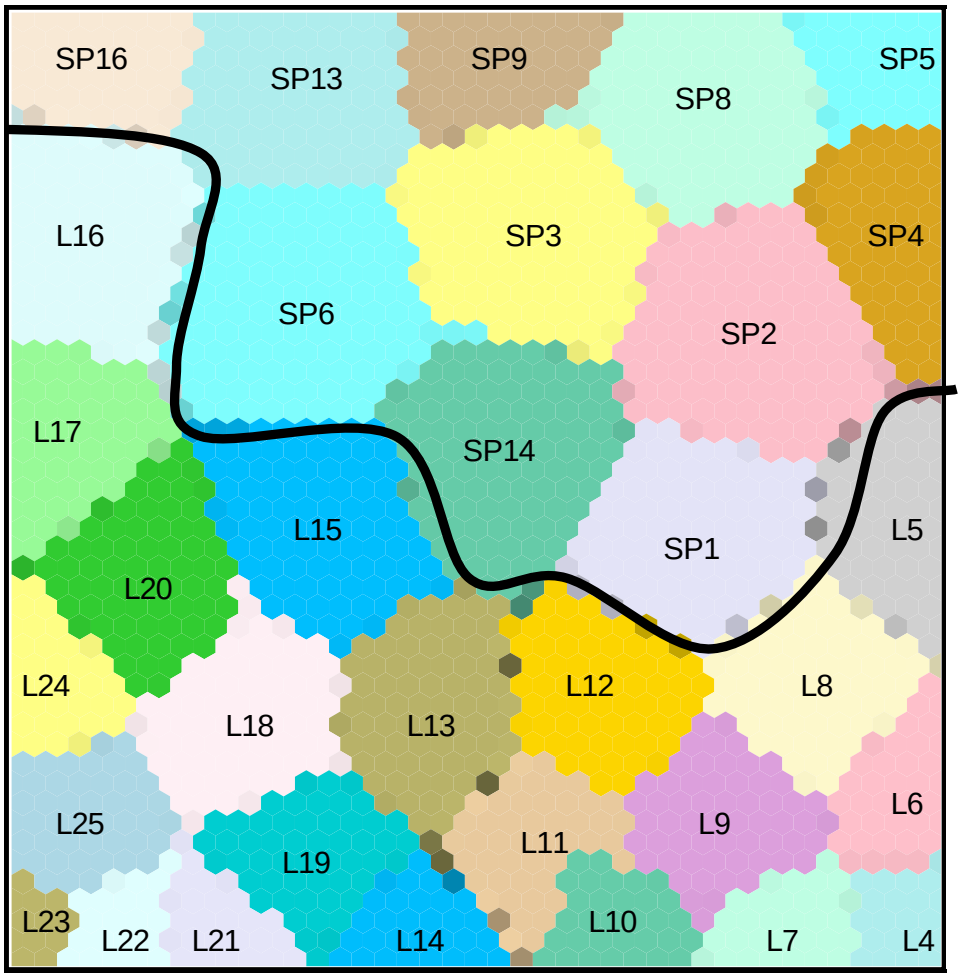


Figure 4 SOFM output map for the TLC and MSP combined dataset (Sp= shoe polish, L = lipstick)

Conclusions:

The results of TLC analysis of wax based products can be directly subjected to MSP to produce reflectance spectra. Wax based products having similar colour and shade which were not distinguishable by TLC were resolved using direct MSP analysis in reflectance mode of the resultant TLC spots. This combination of information obtained from both TLC and direct MSP can be used as a low cost and rapid means to distinguish smears from different type of wax based products having similar colours. SOFM was successful and grouped the wax based products according to type even with basic data derived from a combination of TLC and MSP analysis. This provided a robust mathematical method for verifying the application of TLC and MSP for the wax based product discrimination.

References:

1. Barker, A.M.L. and Clarke, P.D.B., *Examination of Small Quantities of Lipsticks.* Journal of Forensic Science Society, 1972.
2. Cole, M.D. and Thorpe, J.W., *The Analysis of Black Shoe Polish Marks on Clothing.* Journal of Forensic Science Society, 1972. **12**: p. 449 - 451.
3. Harry, R.G., *Harry's Cosmeticology*. 6th Ed. 1973, London: Leonard Hill Books.
4. Williams, D.F.and Schmitt, W.H., *Chemistry and Technology of The Cosmetics and Toiletries Industry*. 2nd Ed. 1992, London: Chapman and Hall.
5. Elvers, B.and Hawkins, S., *Ullman's Encyclopedia of Industrial Chemistry*. 5th Ed (A28). 1996, Weinheim: VCH.
6. Griffin, G.M.E., Doolan, K., Campbell, M., Hamill, J. and Kee, T.G., *Analysis of Wax Based Products by Capillary Gas Chromatography-Mass Spectrometry.* Science and Justice, 1996. **36**.
7. Kohonen, T., *Self Organising Maps*. 1991, Berlin: Springer-Verlag.
8. Zupan, J. and Gasteiger, G., *Neural Networks in Chemistry and Drug Designs*. 1999, Weinheim: Wiley-VCH.
9. Zupan, J., Novic, M. and Ruisanchez, I., *Kohonen and Counterpropagation Artificial Neural Networks in Analytical Chemistry.* Chemometrics and Intelligence Laboratory Systems, 1997. **38**: p. 1 - 23.
10. Beale, R. and Jackson, T., *Neural Computing: An Introduction*. 1990, Bristol: Hilger.
11. Marini, F., Bucci, R., Magri, A.L. and Magri, A.D., *Artifial Neural Networks in Chemometrics: History, Examples and Persepctives.* Microchemical Journal, 2008. **88**: p. 178 - 185.

APPENDIX F

The Application of Multivariate Analysis and Artificial Neural Network Techniques to Analytical Data of Forensic Relevance.

By
Ismail Dzulkiflee

Centre for Forensic Science
Department of Pure and Applied Chemistry
University of Strathclyde

Supervisor
Dr. Niamh Nic Daeid

Friday, 13th August 2010

The Research Objectives

Investigate the applicability of pattern recognition techniques (both unsupervised and supervised) on datasets generated from forensic analysis.

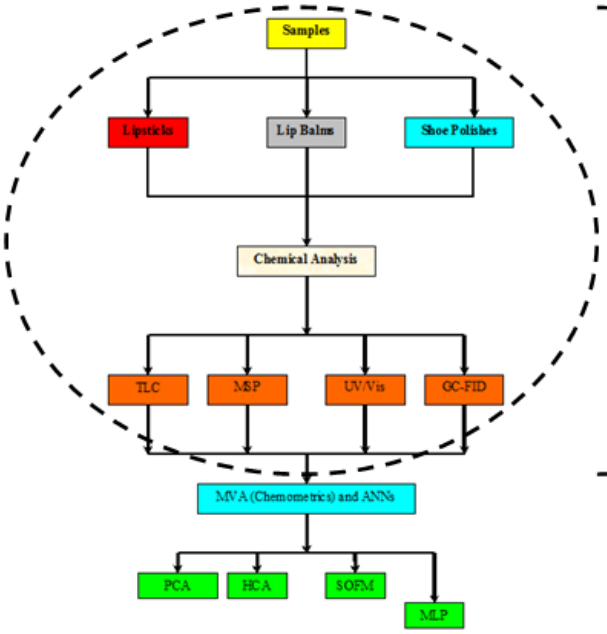
Wax based products analysis was used as a 'proof of concept'

Dataset gathered from four different analytical techniques:
TLC (single data points per sample)
MSP, GC-FID and UV/Vis (multiple data points)

Unsupervised techniques – comparison of conventional PCA and HCA with neural network based technique SOFM

Supervised techniques – Neural network based MLP to interpret unknown samples. MLP can also be used as a **predictive** tool for data.

The Works



The data sets generated in proof of concept study

Multi-Dimensional Dataset

Analysis of wax based products produced multi-dimensional datasets.

Multi-dimensional (multi-variables) datasets provide more information than univariate and bivariate dataset.

Within multi-dimensional datasets are PATTERNS which can convey useful information relating to the interrelatedness of the samples.

Dealing with multi-dimensional datasets (m x n matrix) in order to find any relationships that might exist within and between samples is difficult, due to large numbers of variables and also samples.

Unsupervised Pattern Recognition Techniques

What is an Unsupervised Pattern Recognition technique?

Techniques that do not require in their "learning process" any output example.

Used to determine natural groupings or clusters that might exist in a given dataset based on iterative analysis using a **fixed** algorithm (generally).

PCA and HCA are the most commonly used unsupervised techniques.

Unsupervised Pattern Recognition Techniques

PCA – Original dataset (original variables) are described using new variables (known as principal components (PCs)) obtained from a process known as principal component transformation.

PC transformation process involves a series of complicated mathematical calculations to produce PC loadings used to calculate PC scores for each sample within a given dataset such that;

$$PCn = an1X1 + an2X2 + \ldots + anpXp$$

Where a = loading and X = original / standardised value of the original variables.

University of Strathclyde Science

Unsupervised Pattern Recognition Techniques

Graphical Display of PCA

Arrangement of samples on a score plot based on their PC scores (normally PC1 and PC2 scores - x and y coordinates).

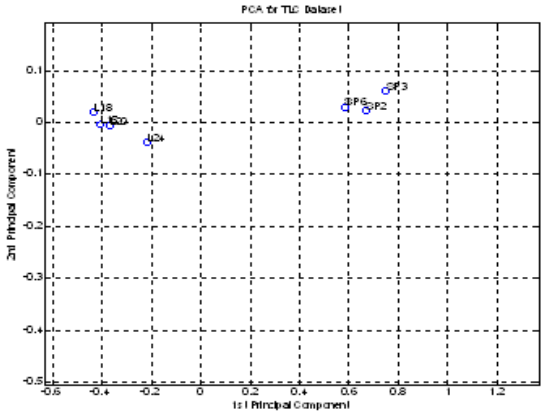
Samples having similar PC scores have similar characteristics.

University of Strathclyde Science

Unsupervised Pattern Recognition Techniques

Graphical Display of PCA

Score Plot arranged in 1, 2 or 3 dimensional form.



University of Strathclyde Science

Unsupervised Pattern Recognition Techniques

HCA - Similarity between objects i.e. samples are measured using a proximity measure determined by a 'fixed' algorithm - most commonly the Euclidean distance is used to calculate the distances between samples.

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}$$

Where d_{ij} is the distance between object i and j

The closer the distance between objects, the similar they are and vice - versa.

University of Strathclyde Science

Unsupervised Pattern Recognition Techniques

Graphical Display of HCA

Arrangement of samples on a tree diagram or dendrogram based on the similarity of the samples using the Euclidean distance proximity measure.

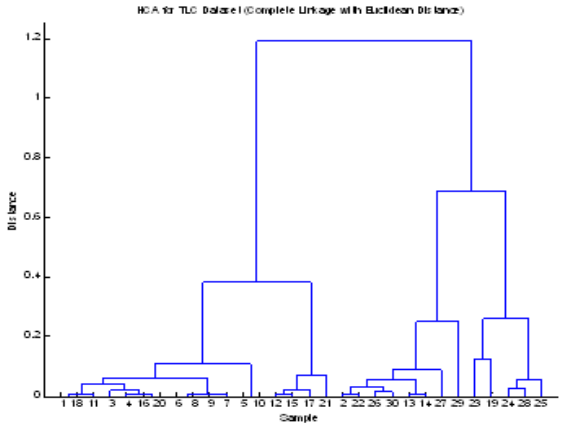
Cluster linkage - simple, average or complete linkages can be used to determine how clusters within dataset are linked.

University of Strathclyde Science

Unsupervised Pattern Recognition Techniques

Graphical Display of HCA

Dendrogram or tree diagram



University of Strathclyde Science

Unsupervised Pattern Recognition Techniques

Self Organising Feature Maps (SOFM) or Kohonen Maps is an unsupervised pattern recognition technique based on Artificial Neural Network learning.

Self Organising Feature Maps (SOFM) – have been used in other fields e.g. Engineering, Marketing, Finance, Environmental and Ecology – yet to be explored for data of forensic science relevance.

The idea of SOFM is the same as PCA and HCA - dimensionality reduction (representing multi-dimensional dataset into much lower dimensional) and clustering i.e. finding any natural groupings within a given dataset.

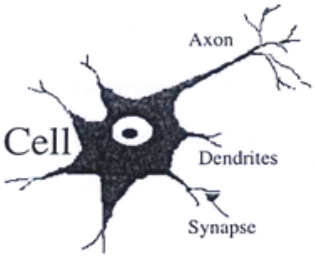
The big difference is in the nature of the algorithm generation in Artificial Neural networks.

University of Strathclyde Science

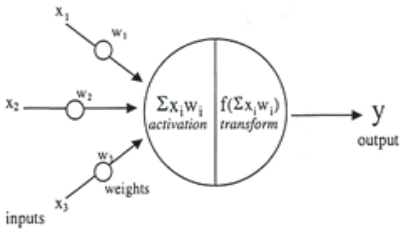
What is Artificial Neural Networks (ANNs)

Is a '**machine**' either it be hardware or software that tries to simulate the most basic function of human brain i.e. ABILITY TO LEARN.

Structure - made up of a connection of artificial 'neurons' (perceptron) which function in a manner similar to biological neurons.



Biological Neuron



Artificial Neuron a.k.a Perceptron

University of Strathclyde Science

What is Artificial Neural Networks (ANNs)

Biological Neuron Vs Perceptron

Biological Neuron	Perceptron
Signal receives by **Dendrites**	Input Pattern ($X_1, X_2,\ldots,X_n$)
Synaptic Strength at **Synapse**	Weight Vectors ($W_1, W_2,\ldots,W_n$)
Cell Body	Summation Block
Cell Body Output	Perceptron Output

University of Strathclyde Science

What is Artificial Neural Networks (ANNs)

Human Learning

Information – from a sensor (biological neuron) to a synapse passes through a series of neurons in the brain.

ANNs Learning

Difficult process – involves the adjustment of connection weights between perceptrons using a systematic application of algorithms until the correct response is found for a given piece of information.

University of Strathclyde Science

What is Artificial Neural Networks (ANNs)

Analogy of ANNs Learning Process – Weight Adjustments

1 + 1 = 2
1 + 1 = 0
1 + 1 = 1

Which is the correct answer?

University of Strathclyde Science

What is Artificial Neural Networks (ANNs)

Analogy of ANNs Learning Process – Weight Adjustments

1 + 1 = 2
(Do nothing)

1 + 1 = 0
(Subtract 1 to the answer, 1 - 1 = 0)

1 + 1 = 1
(Add 1 to the answer, 1 + 1 = 2)

The process of do nothing, subtraction and addition is analogous to the weight adjustment (i.e. learning process) in ANNs

University of Strathclyde Science

Unsupervised Pattern Recognition Techniques

Mathematical Background of SOFM

Learning utilises the adjustments of weight vectors so that they will become close to the input vectors after every introduction (iteration) of the input vectors to the network.

Similar to HCA, the Euclidean Distance is used but in this case in an iterative alternating manner to assess the similarity between input and weight vectors.

Best Matching Unit (BMU) has weight vectors similar to input vectors i.e. Groupings or clusters of similar objects within the dataset are created (around the BMU).

University of Strathclyde Science

Unsupervised Pattern Recognition Techniques

Graphical Display of SOFM

Best Matching Units (BMU) as the centre - objects similar to BMU are arranged around the BMU.

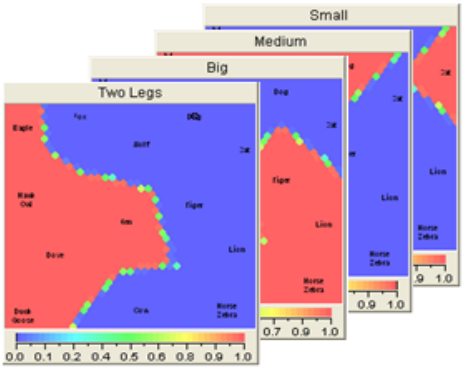
The closer the object to the BMU, the greater the similarity.

University of Strathclyde Science

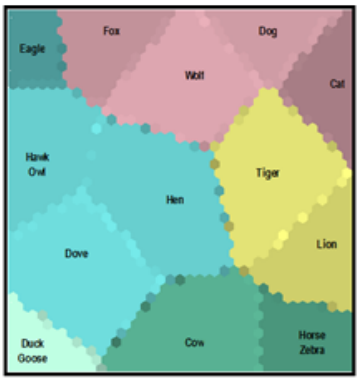
Unsupervised Pattern Recognition Techniques

Graphical Display of SOFM

Output and component maps - commonly in 2 dimensional form.



SOFM Component Maps



SOFM Output Map

University of Strathclyde Science

Supervised Pattern Recognition Technique

Multi – Layer Perceptron (MLP) also known as Back Propagation Neural Network (BP-NN).

MLP is based on Supervised Pattern Recognition Technique – as opposed to unsupervised – requires output examples in "learning" process.

MLP is a "Universal Function Approximator":
MODELS the underlying function in a given dataset,
MODELS the distribution of objects or samples in the sample space.

Useful for classification and prediction.

University of Strathclyde Science

Supervised Pattern Recognition Techniques

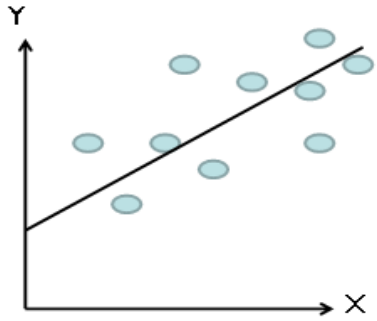
What is a model?

According to Oxford Dictionary:
A small object, usually built in scale, that represent in detail another, often larger object.

University of Strathclyde Science

Supervised Pattern Recognition Techniques

In a **mathematical sense**,



$Y = f(X)$

e.g. $Y = mX + C$

(used to model distribution of objects in two dimensional sample space)

University of Strathclyde Science

Supervised Pattern Recognition Techniques

Has **PREDICTIVE** capability

$Y = mX + C$

Given the value of X, m (the slope) and C (the intercept), can find the value of Y (an unknown).

e.g. X = 2, m = 2 and C = 2, then Y ?

$Y = mX + C$
$Y = 2(2) + 2$
$Y = 6$

University of Strathclyde Science

Supervised Pattern Recognition Techniques

Classification rules

$$Y = mX + C$$

Given that,
If $Y > 5$, then Y belongs to Group A;
If $Y < 5$, then Y belongs to Group B;

So for our example, since Y (unknown) = 6 > 5, then Group A.

University of Strathclyde Science

Supervised Pattern Recognition Techniques

For dataset which are multi-dimensional, we need 'something' to estimate the function of the multidimensional dataset - MLP neural network is a tool for such a purpose.

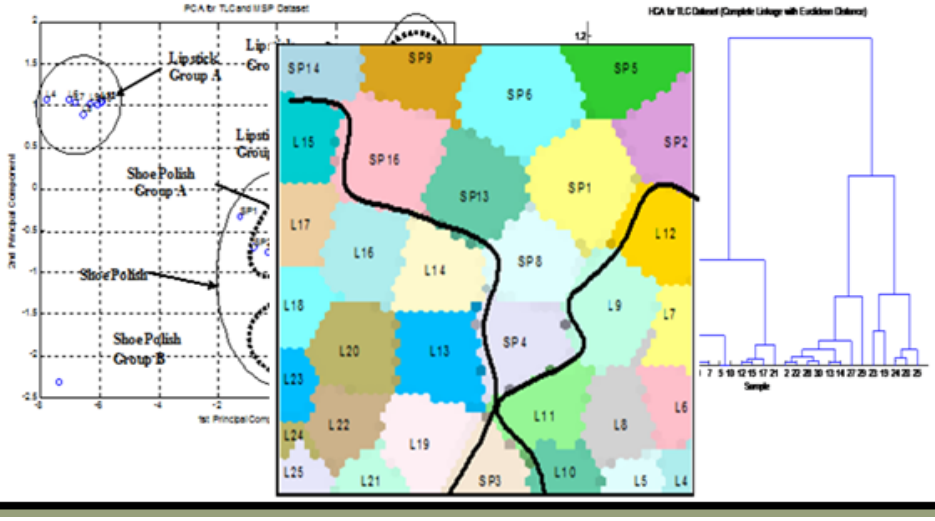
MLP determines the mathematical algorithm which models the distribution of objects/samples presented within a multi-dimensional space i.e. estimates the function.

Once found, the model can be used to interpret unknown objects/samples and classify these or predict their classification.

University of Strathclyde Science

Proof of Concept of The Use of ANN (Unsupervised (SOFM) and Supervised (MLP)

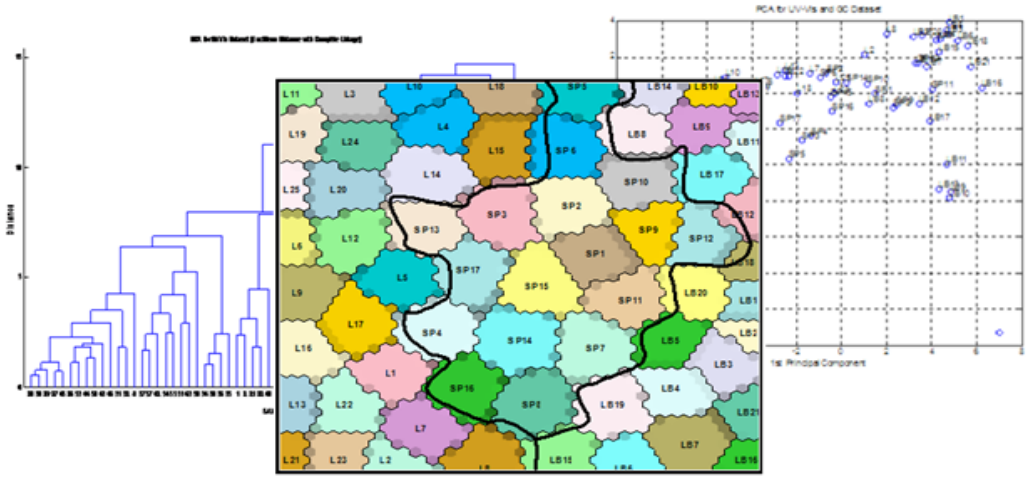
Wax based products – single data point per sample from TLC identify groupings.



University of Strathclyde Science

Proof of Concept

Wax based products – multiple data points per sample from MSP, GC-FID and UV/Vis analysis – identify groupings



University of Strathclyde Science

Proof of Concept

HCA and PCA both fail to cluster known samples based on either single point data or multiple point data for samples with similar characteristics

SOFM was capable of correctly grouping samples. The SOFM output facilitates ease of understanding with a simple graphical display.

MLP was capable of correctly classifying the samples. Also successful in predicting the class membership of number of unknown samples.

University of Strathclyde Science

Further & Potential Applications

Extension of application of SOFM and MLP to other dataset of forensic science relevance – including chromatographic data (e.g. drug impurity profiling or ignitable liquid analysis), spectroscopic data (e.g. ink and dye analysis) or combinations of data sets.

Ignitable liquids provide a very complex dataset and SOFM has been used to link un-evaporated samples to their degraded weathered samples for the first time. (Paper recently published in Analytical Chemistry*).

MLP – Offers an opportunity to be predictive and as such can be used to facilitate the optimisation of chromatographic systems for forensic analysis (e.g. in method development)

*Application of Unsupervised Chemometric Analysis and Self-Organizing Feature Maps (SOFM) for the Classification of Lighter Fuels., *Wan N. S. Mat Desa, Niamh Nic Daeid, Dzulkiflee Ismail and Kathleen Savage., Analytical Chemistry.*, **2010**, 82 (15), pp 6395–6400*