CRANFIELD UNIVERSITY

MICHAEL CAUCHI

DATA ANALYSIS TOOLS FOR
SAFE DRINKING WATER PRODUCTION

Cranfield Health

PhD Thesis

2006

CRANFIELD UNIVERSITY

Cranfield Health

Analytical Science and Informatics

Doctor of Philosophy

2006

MICHAEL CAUCHI

DATA ANALYSIS TOOLS FOR
SAFE DRINKING WATER PRODUCTION

Supervisors: Dr S. J. Setford, Dr C. Bessant

April 2006

This thesis is submitted in partial fulfillment of the requirements for the Degree of PhD

ABSTRACT

Providing safe and high quality drinking water is essential for a high quality of life. However, the water resources in Europe are threatened by various sources of contamination. This has led to the development of concepts and technologies to create a basis for provision of safe and high quality drinking water, which had thus resulted in the formation of the *Artificial Recharge Demonstration* project (ARTDEMO). The overall aim of this thesis in relation to the ARTDEMO project was to develop a real-time automated water monitoring system, capable of using data from various complementary sources to determine the amounts of inorganic and organic pollutants.

The application of multivariate calibration to differential pulse anodic stripping voltammograms and fluorescence spectra (emission and excitation-emission matrix) is presented. The quantitative determination of cadmium, lead and copper acquired on carbon-ink screen-printed electrodes, arsenic and mercury acquired on gold-ink screen-printed electrodes, in addition to the quantitative determination of anthracene, phenanthrene and naphthalene have been realised. The statistically inspired modification of partial least squares (SIMPLS) algorithm has been shown to be the better modelling tool, in terms of the root mean square error of prediction (RMSEP), in conjunction with application of data pre-treatment techniques involving range-scaling, filtering and weighting of variables. The % recoveries of cadmium, lead and copper in a certified reference material by graphite furnace atomic absorption spectrometry (GF-AAS) and multivariate calibration are in good agreement.

The development of a prototype application on a personal digital assistant (PDA) device is described. At-line analysis at potential contamination sites in which an instant response is required is thus possible. This provides quantitative screening of target metal ions. The application imports the acquired voltammograms, standardises them against the laboratory-acquired voltammograms (using piecewise direct standardisation), and predicts the concentrations of the target metal ions using previously trained SIMPLS models.

This work represents significant progress in the development of analytical techniques for water quality determination, in line with the ARTDEMO project's aim of maintaining a high quality of drinking water.

ACKNOWLEDGEMENTS

I would like to thank Dr Steven John Setford and Dr Conrad Bessant for their support, knowledge, expertise and guidance throughout the course of this work, and in allowing me the opportunity to work on this project.

Thanks also go to Mr Paul Knight and Mr David Pitts for their enthusiastic approach to assisting me in times of need.

Additional thanks go to the European Commission for funding the ARTDEMO project in addition to attending conferences and meetings. Particular thanks go to the project coordinator Mr Bjoern Jensen and all the other partner representatives for its smooth running.

I also take the opportunity to thank the following:
- Dr Joanne Cooper for easing me into the flow of work in my early days
- Mr Guillaume Saint-Pierre for his valuable advice in the laboratory
- Mr Richard Dudeney for permitting me to employ his batch of gold-ink screen-printed electrodes
- Mr Richard Andrews for his valuable assistance, time and expertise in employing the atomic absorption spectrometer
- Mr Darren Oakley for his time in assisting me with computing issues
- My friends and office colleagues (Ian, Daz, Fady, Jenny, Judit, Jez, Nick, Nikki and Marke) for making my time here at Cranfield enjoyable, pleasant and rewarding

Finally, I wish to thank my parents and my brother for their continual support, patience, and advice, especially through dark times.

DEDICATION

This thesis is dedicated to my parents, Louis and Marie, my sister Isabelle, and especially my brother, Daniel.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A	Ampere
Ag/AgCl	Silver/silver chloride
ANN	Artificial Neural Network
ARTDEMO	Artificial Recharge Demonstration
ASV	Anodic Stripping Voltammetry
CLS	Classical Least Squares
CRM	Certified Reference Material
CSV	Cathodic Stripping Voltammetry
%CV	% Coefficient of Variation
CV	Cyclic Voltammetry
DME	Dropping Mercury Electrode
DOC	Dissolved Organic Carbon
DPASV	Differential Pulse Anodic Stripping Voltammetry
F-AAS	Flame Atomic Absorption Spectrometry
GF-AAS	Graphite Furnace Atomic Absorption Spectrometry
HG-AAS	Hydride Generation Atomic Absorption Spectrometry
HMDE	Hanging Mercury Drop Electrode
HPLC	High Performance Liquid Chromatography
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
ILS	Inverse Least Squares
ITSV	Independent Test Set Validation
ITTFA	Iterative Target Transformation Factor Analysis
LOD	Limit of detection
LV(s)	Latent Variable (LVs)
MFE	Mercury Film Electrode
MLR	Multiple Linear Regression
mg kg^{-1} (mg/kg)	Milligrams per kilogram
mg l^{-1} (mg/l)	Milligrams per litre (ppm)
mol l^{-1}	Mole per litre
MS	Mass Spectrometry
mV	Millivolts
N	Number of electrons or number of data points

NIPALS	Non-linear Iterative Partial Least Squares
NIR	Near-Infrared
NL-PLS	Non-linear Partial Least Squares
NN-PLS	Neural Network Partial Least Squares
NMR	Nuclear Magnetic Resonance
NOM	Natural Organic Matter
ng l^{-1}, μg l^{-1} (μg/l)	Nanograms per litre (ppt), Microgram per litre (ppb)
nm	Nanometers
PC or PC(s)	Personal Computer or Principal Component(s)
PCA	Principal Component Analysis
PCR	Principal Component Regression
PDA	Personal Digital Assistant
PLS	Partial Least Squares
PDS	Piecewise direct standardisation
ppb, ppm	parts per billion (μg/l), parts per million (mg/l)
PSA	Potentiometric Stripping Analysis
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error of Cross Validation
RMSEP	Root Mean Square Error of Prediction
RSD	Relative Standard Deviation
SCE	Saturated Calomel Electrode (Hg|Hg$_2$Cl$_2$|KCl)
SHE	Standard Hydrogen Electrode
SG (or S-G)	Savitzky-Golay
SIMCA	Soft Independent Modelling of Class Analogies
SPE	Screen-printed electrode
SVD	Singular Value Decomposition
SWASV	Square Wave Anodic Stripping Voltammetry
s	Second
US-EPA	United States Environmental Protection Agency
UV-VIS	Ultraviolet - Visible
μA	Microampere
μl	Microlitre
V	Voltage
v/v, w/v, w/w	Volume/volume, weight/volume, weight/weight

XRF	X-ray Fluorescence

As	Arsenic	Cu	Copper	AgCl	Silver Chloride
Ag	Silver	Hg	Mercury	HCl	Hydrochloric Acid
Bi	Bismuth	Pb	Lead	KCl	Potassium Chloride
Cd	Cadmium	Zn	Zinc	HNO_3	Nitric Acid
Cr	Chromium	Au	Gold	H_2SO_4	Sulphuric Acid

NOTATION

The datasets described in this thesis are manipulated as matrices or vectors. A *matrix* is seen as a two-dimensional array, for example a 4 x 5 matrix will contain 4 rows and 5 columns and a total of 20 *elements*. Conversely, a one-dimensional array is a *vector*, for example, if the dimensions are **1** x 7 then this is labelled as a *row vector*; if 7 x **1**, then it is labelled as a *column vector*; both contain 7 elements.

The terminology often employed is to assign letters to matrices, vectors and elements (also known as *scalars*), often the letter X.

- A matrix with dimensions I × J, is denoted by a bold upper-case letter such as $X_{I,J}$.
- A row vector is denoted by a bold lower-case letter with indices such as $x_{i,J}$, for example, $x_{3,J}$ is the entire third row of the matrix.
- A column vector is denoted by a bold lower-case letter with an index as $x_{I,j}$, for example, $x_{I,4}$ is the entire fourth column of the matrix.
- An element (or scalar) is denoted by an italicised lower-case letter with the appropriate indices such as x_{ij}, for example $x_{5,2}$ gives the value of the element in the 5th row and 2nd column of the matrix.

In general, X denotes independent variables such as stripping current at varying potentials (in anodic stripping voltammetry (Section 1.4.2.4)) and Y denotes dependent variables such as the concentration. Therefore each row (vector) in the dataset (matrix) is, for example a voltammogram, and where each element within the vector is a *response* (variable) at a specific potential. The corresponding row in the Y matrix is the concentration(s) of the particular constituent(s) generating the response in X.

CHAPTER 1: INTRODUCTION

1.1 Overview

This chapter gives a general introduction to the project, and the need for artificial recharging of ground water. The purpose of the ARTDEMO project is explained followed by a general discussion of three analytical techniques: electrochemical, fluorescence and atomic absorption. An in-depth discussion on the origins and application of chemometrics follows with the onus on multivariate calibration. Finally, the chapter concludes with the main aims and objectives of the thesis, finishing with an outline of the structure and format of the remainder of the thesis.

1.2 General

Water is a vital commodity for every living entity on the planet from humans to plants and animals. Most of the rain falls on to the ground where it will either settle into stagnant pools and later evaporate back into the air, or seep down beneath the ground via processes such as leaching and elution, and over time settle into vast pockets of water sometimes deemed "underground lakes" (Stuyfzand, 2004). Rain will also fall on to higher ground, running-off into rivers which flow into large lakes or into estuaries and eventually out to sea.

Water companies exploit the presence of these large lakes by extracting the water from them, purifying it, and distributing it to households and businesses in the immediate area. However, there are some countries that are unable to store huge amounts of water in lakes, whether man-made (reservoirs) or natural. Such reasons include the extreme temperatures that can be found in developing regions such as in Africa hence leading to lakes drying-out, or due to the natural terrain, such as mountainous regions. Such regions will rely on water that has seeped into the underground lakes, known as *ground water*.

Coupled with the extensive growth of the world population, even lakes and reservoirs are succumbing to the extensive demands. Some water companies are turning to

ground water sites to keep up with demand. However, uptake of this water is much greater than replenishment resulting in these underground lakes being depleted. It is thus imperative that more efforts are made to capture and harvest the rainfall seeing as the process of natural recharging takes many years. With this in mind, it is necessary to *speed up* this recharging process, and thus the concept of *artificial recharging* has evolved.

Artificial recharging is defined as the *rapid* replenishment of water sources such as underground lakes (ground water) where the natural process takes decades to complete. This can involve re-directing surface run-off and channelling rain water into aquifers which filter the water and complement the ground water already present. Figure 1.1 illustrates the two processes.
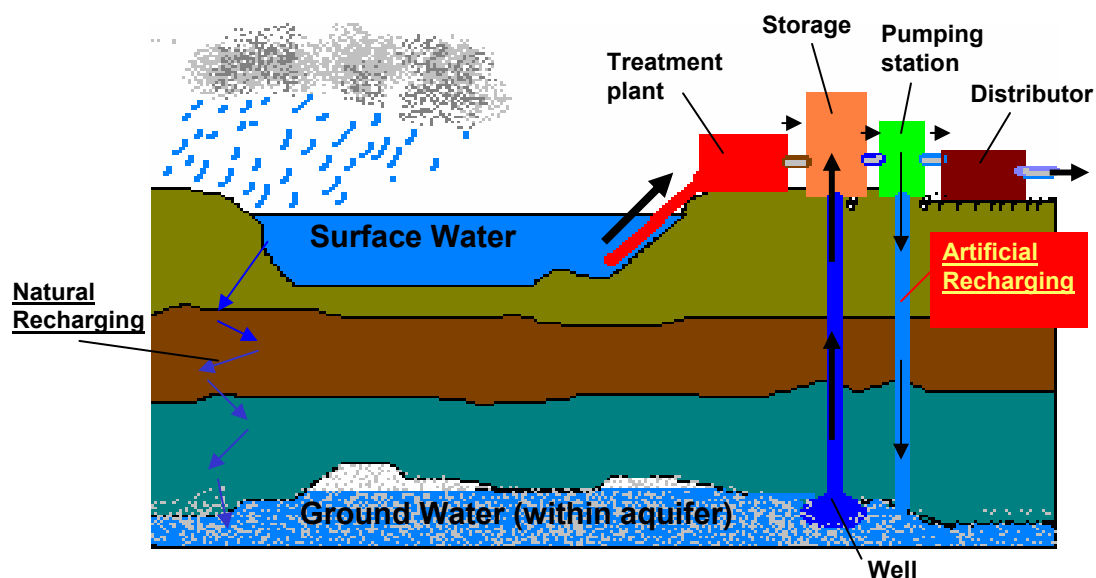
Figure 1.1: The processes involved in natural and artificial recharging

Such harvesting of the water can occur from ponds, ditches, farm lands, storm drains and even the reclamation of sewage water. Many water companies are looking to develop concepts and technologies which will create a basis for the provision of safe and high quality drinking water.

1.3 The ARTDEMO Project

As the world population grows, so does the ever-increasing demands on the Earth's natural resources. In addition to this, the generation of waste materials continues to greatly affect the environment in numerous ways, for example greenhouse emissions and river pollution. The effect on the latter applies more pressure on relevant water companies to maintain the quality of the drinking water from its source (surface or ground waters) to final distribution. Development of reliable, low-cost, and robust tools for the quantitative determination of target pollutants has therefore been one of the main aims of the ARTDEMO project. These tools range from portable hand-held devices to at-line or on-line sensors.

The Artificial Recharge Demonstration Project (ARTDEMO) was a European Commission funded project established to research the management and maintenance of the quality of drinking water via the process of artificial recharging. The overall objective extracted from the "Description of Work" (final version, 14 August 2002) was to "demonstrate a management tool for artificial recharge (AR) plants, which use sophisticated monitoring systems linking automatic real-time data acquisition, available on-line/at-line sensor systems and fast field analysis kits with intelligent decision support software, which compile and communicate processed data to specific action protocols. The advanced technology will enhance the state-of-the-art in AR markedly and extend the capacity of existing plants and general applicability of the technology, thereby ensuring a stable and safe drinking water production".

There were six institutions involved in this project each contributing a level of expertise to the requirements, as detailed in Table 1.1, and undertaking a series of projects to achieve the desired goal.

Table 1.1: The ARTDEMO Project partners and their expertise

No	Partner	Acronym	Type	Main expertise and role in ARTDEMO
1	Danish Hydraulic Institute: Water & Environment, Denmark	DHI	R&D	Project coordination, Modelling, Decision support systems, Fate and transport
2	Consejo Superior de Investigaciones Cientificas (Spanish National Research Council), Barcelona, Spain	CSIC	R&D	Contaminant chemistry, Sensor systems
3	Cranfield University, UK	CU	R&D	Sensors, mobile test kits
4	The Gothenburg Region Association of Local Authorities, Sweden	GR	End User	Provision of test sites
5	Kiwa Water Research, Holland	KIWA	R&D,	Bed regeneration, operation of AR-plants
6	Copenhagen Energy, Denmark	KE	End User	Provision of test sites

The overall aim was to develop a *management tool* which would contain a suitable monitoring system that was able to detect both organic pollutants such as polynuclear aromatic hydrocarbons (PAHs), estrogens, progestogens, antibiotics and volatile organic compounds (VOCs), in addition to inorganic pollutants such as the heavy metals (arsenic, lead and cadmium). Ideally, automatic real time data acquisition would be integrated with *intelligent decision software*. On-line and at-line sensor systems coupled with fast field "hand-held" analysis kits in the form of *personal digital assistants* (PDAs) were also a requisite of the management tool.

The main role of Cranfield University was to integrate the measurement test kits designed and developed in-house for on-line/at-line monitoring with advanced data handling software incorporating chemometric techniques (discussed in Section 1.5). The contribution of this thesis to the overall ARTDEMO project is the development and application of these techniques. This therefore leads to the qualitative detection and quantitative determination of a range of organic and inorganic pollutants, and thus

the overall quality of the water. Section 1.6 details the full aims and objectives of the work performed in this thesis.

1.4 Analytical Techniques for the Detection of Heavy Metals and Hydrocarbons

This section will briefly introduce typical contaminants common in the water industry such as heavy metals (Section 1.4.1.1) and hydrocarbons (Section 1.4.1.2). Three analytical techniques employed in this study will be presented: electrochemical methods (Section 1.4.2), fluorescence spectroscopy (Section 1.4.3), and atomic absorption spectrometry (Section 1.4.4).

1.4.1 The Main Contaminants in the Water Industry

1.4.1.1 Heavy Metals

The heavy metals which are of great concern to environmental monitoring bodies are cadmium (Cd), lead (Pb), copper (Cu), zinc (Zn), mercury (Hg), arsenic (As), selenium (Se) and chromium (Cr) (Lambert, 1997). Due to their wide applications in industry, soil and water contamination is high thus presenting a threat to both the ecosystem and humans. Table 1.2 lists some of the analytical techniques employed to analyse them in addition to the advantages and disadvantages (Alloway, 1990; Thomas, 2003).

Table 1.2: The main advantages and disadvantages of some of the analytical techniques employed for the detection of heavy metals

Technique	Metals	Advantage	Disadvantage
Flame AAS	Cd, Pb, Cu, Cr	Wide analytical working range	Different metals require different flames and sources
HG-AAS	As, Se	Increased sensitivity compared to flame only	Qualified personnel required and lab-based
ICP-MS	All	Simultaneous qualitative and quantitative determination of all metals	High consumption of argon and qualified personnel required
Voltammetry	All	Can be employed in the field and is cheap	Different instrumental parameters for different metal
XRF	All	Direct analysis of solids in addition to field applications	Poor resolution and low sensitivity

Table 1.2 indicates that ICP-MS is superior to the atomic absorption techniques due to the simultaneous analysis of the target metal ions. However, just like the absorption techniques, the cost of operation is high due to the consumption of argon for ICP-MS and acetylene for flame AAS. The voltammetric analytical technique offers a cheaper and more convenient approach to metals analysis in addition to field (at-site) applications. (Wang *et al.*, 1999). Voltammetry can also be expanded to measure hydrocarbons (Richards, 2003; Guiberteau *et al.*, 1995).

Selenium and chromium were not analysed in this work. Further details of cadmium, lead and copper are given in Section 2.3.1 whilst arsenic and mercury are given in Section 3.3.1. Voltammetry is discussed further in Section 1.4.2. Atomic absorption spectroscopy is generally employed as a validation technique to support the findings of field-instruments (Cooper, 2004; Samek *et al.*, 2001). As this technique was employed in this work to do just that, a brief discussion on the basic theory is given in Section 1.4.4.

1.4.1.2 Hydrocarbons

In addition to heavy metals there is contamination by a whole host of hydrocarbons. In relation to the ARTDEMO project pesticides, estrogens, phenols and phthalates have been analysed in water, soil and sewage sludge samples by a range of laboratory-based instruments such as HPLC-MS, HPLC-MS-MS, and GC-MS (Borba da Cunha *et al.*, 2004; Diaz-Cruz *et al.*, 2003; Kuster *et al.*, 2004; Rodriguez-Mozaz *et al.*, 2004). The main advantages of these instruments are the selectivity, specificity, high resolution and simultaneous quantitative determination of the target compounds. The main disadvantages are the operational costs and the inapplicability to field (at-line/on-line) measurements.

Another important class of compounds is the polynuclear aromatic hydrocarbons (PAHs) of which 16 have been listed to be of great environmental concern (Section 5.3.1). In addition to employment of the instruments in the above paragraph, the PAHs can be analysed with fluorescence spectroscopy (Skoog *et al.*, 1996; Rouessac & Rouessac, 2000). The main advantages are its non-destructiveness to the sample,

minimal sample preparation and reagent consumption, rapid analysis and minimal operational costs. The main disadvantage is poor resolution of overlapping peaks. However these can be overcome by application of chemometric techniques (Section 1.5 and 5.4). The basic theory of fluorescence spectroscopy is presented in Section 1.4.3.

1.4.2 Voltammetric Methods in Electrochemistry

A brief description of electrochemistry followed by an explanation of the theory and some of the commonly used methods in voltammetry, for example differential pulse anodic stripping voltammetry (DPASV) will be presented.

1.4.2.1 Electrochemistry

Electrochemistry is a branch of chemistry that focuses on chemical reactions that occur at the surface of an *electrode*, namely in the *reduction* (gain of electrons) or *oxidation* (loss of electrons) of a given species when an electrical *current* is passed through a solution containing an *electrolyte* (Sharp, 1990). These reactions are deemed *redox* reactions. Two types of reactions are *electrolytic* and *galvanic*. The former involves the application of an external voltage to drive the reactions whilst the latter results in a spontaneous reaction dependent on the ions present in the reaction medium.

In electrolytic systems, the electrode should be a highly conductive material ideally fabricated from inert materials such as graphite carbon, metallic gold, and platinum. When oxidation occurs at the surface, it is termed an *anode*; when reduction occurs, a *cathode*. In solution-based systems, the electrolyte is an ionic compound that partially or completely dissociates in solution permitting the electrical conduction between two or more electrodes. Common electrolytes employed are potassium chloride, sodium chloride and sulphuric acid (Atkins, 1992).

The reactions that occur at an electrode are either *irreversible* or *reversible*. The former implies that the original reactants will not be re-generated under appropriate electrochemical conditions. The latter implies that the same reactants will always be

attained whether or not a potential is applied (Skoog *et al.*, 1996). The electrode potential (*E*) arises when an electrode (eg Cu) is placed into a solution containing the same ions (Cu^{2+}). The size of the potential is dependent upon the concentration of these ions (Sharp, 1990). The standard electrode potential (E^O) is the same as the electrode potential but where the *activities* of the reactants and product are at unity, and is measured against a hydrogen electrode at 1M solution, 1 atm and 25°C. The activity is generally employed in place of concentration (for solutions) and pressure (for gases) when non-ideal situations arise as in the deviation from *Raoult's Law* (the dissolution of a solute in a solvent results in the solvent's vapour pressure being reduced in proportion to the mole fraction of the solute present (Atkins, 1992)).

Table 1.3 shows the standard electrode potential for a number of selected electrodes and their respective half-cell reactions.

Table 1.3: Standard electrode potentials (E^O) of a number of selected electrodes

Electrode	Reaction (Reduction)	E^O (V) at 25^OC
Ag	$Ag^+ + e^- \rightarrow$ Ag (s)	+0.799
Cu	$Cu^{2+} + 2e^- \rightarrow$ Cu (s)	+0.337
Hg	Hg_2Cl_2(s) + $2e^- \rightarrow$ 2Hg (l) + $2Cl^-$	+0.268
H$_2$	**$2H^+ + 2e^- \rightarrow$ H$_2$**	**0.000**
Pb	$PbSO_4$ (s) + $2e^- \rightarrow$ Pb (s) + SO_4^{2-}	-0.350
Cd	$Cd^{2+} + 2e^- \rightarrow$ Cd (s)	-0.403
Zn	$Zn^{2+} + 2e^- \rightarrow$ Zn (s)	-0.763

In relation to the standard hydrogen electrode, positive values imply that the corresponding electrode is a better oxidant (acceptor of electrons) than hydrogen or any other electrode with a greater negative potential, in other words, copper is a better oxidant than lead. From Table 1.3 a positive standard potential indicates the reduction is spontaneous whilst negative values indicate that oxidation is spontaneous.

Given an electrode reaction such as:

$$C_O + ne^- \leftrightarrow C_R \qquad \qquad \textbf{1.1}$$

The electrode potential is related to the concentration of the corresponding ion via the *Nernst* Equation (Equation 1.2).

$$E = E^O - \frac{RT}{nF} \ln \frac{C_R}{C_O}$$ **1.2**

where E and E^O are the electrode potential and standard electrode potential respectively, R is the universal gas constant (8.314 J K^{-1} mol^{-1}), T is the temperature in Kelvin (K), n is the number of electrons involved in the half-reaction, F is the Faraday constant (96,485 C), ln is the natural logarithm, and C_R and C_O are the concentrations of the reduced and oxidised components containing the target ions respectively.

Electrochemical measurements initially involved two electrodes: the *working electrode*, where the electrochemical reactions in the form of oxidation or reduction occurred, and a *counter electrode* which was large enough not to be polarised during the measurement (Skoog *et al.,* 1996). The potential of the working electrode was thus calculated from the difference between the applied potential and the potential of the counter electrode. However, distorted voltammograms were obtained when the resistance of the electrolyte was high. A third electrode was thus introduced whose potential remained constant throughout the entire measurement. This was known as the *reference electrode*. Such electrodes used were the saturated calomel electrode (SCE) and the silver/silver chloride (Ag/AgCl) electrode. Figure 1.2 illustrates a schematic configuration of a voltammetric instrument.
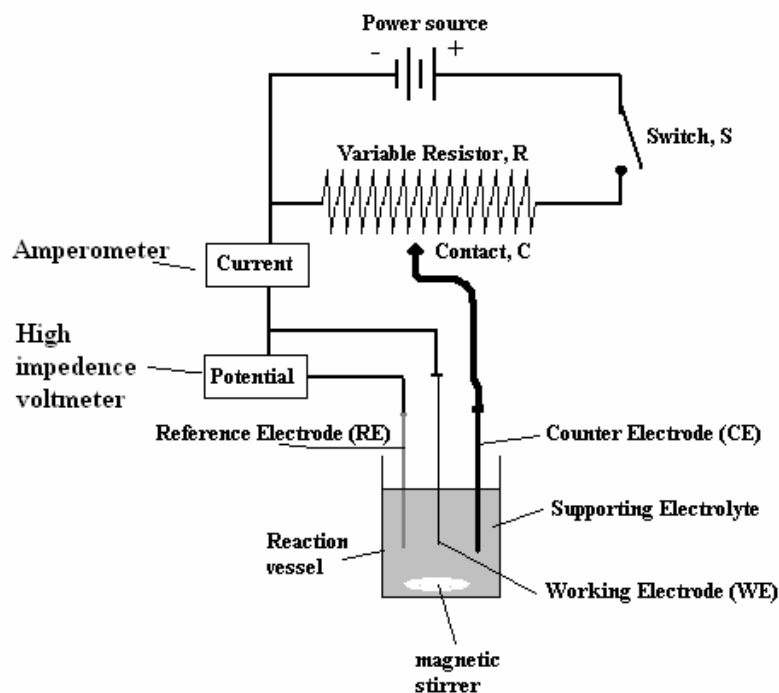
Figure 1.2: A typical set up of an instrument to perform linear-scan voltammetric (LSV) measurements.

In predominately aqueous systems, the supporting electrolyte is typically a Group I (alkali) metal salt added in excess to the solution containing the metal ions, an example being potassium chloride (KCl). This is known not to undergo chemical reaction with the commonly used working electrode at the potentials used for measurement. The dimensions of the working electrodes are generally smaller when compared to the other electrodes to ensure polarisation. The potential at the working electrode is varied linearly with respect to time. The potential of the reference electrode remains constant. The counter electrode allows the flow of electric current from the power source to the working electrode. It acts to complete the electrochemical circuitry in addition to offering a counter reaction to the process occurring at the working electrode.

The power source itself consists of a variable resistor, R, and the potential is selected by sliding the contact, C, along it. The actual resistance of the circuit is very high (> $10^{11}\Omega$) which implies that there is no current present, and thus all the current generated at the source is transferred from the counter electrode to the working electrode. The voltammogram is thus obtained by moving the contact, C, along R at a

constant rate (*linear scan voltammetry*). The current obtained is recorded as a function of the potential between the working electrode and the reference electrode.

The working electrodes commonly employed are highly conductive and inert materials such as mercury (in the form of a dropping mercury electrode (DME)), carbon, gold and platinum, with which the potential is varied to induce a redox reaction (Skoog *et al.,* 1996). Two types of carbon-based working electrodes commonly employed are solid-state glassy-carbon and carbon-inks as employed in screen-printed electrodes (Cooper, 2004). Section 2.2 discusses in greater detail carbon-based working electrodes, particularly in relation to the application of a mercury film on to the carbon surface. Section 3.2 discusses gold working electrodes and introduces the novel application of gold-ink screen-printed electrodes.

The relationship between current and concentration has led to the development of two important electro-analytical techniques: *amperometry* and *voltammetry*. The former involves measuring the change in current as a function of time when potentials are applied in order to reduce or oxidise target components at the surface of the electrode. The latter is explained in the next section.

1.4.2.2 Voltammetry

Voltammetry is defined as "a method based on the measurement of current as a function of the potential applied to an electrode" (Skoog *et al.,* 1996). It originates from *polarography* which is a measure of current as a function of potential when the working electrode is comprised of a liquid conductor such as mercury implying that polarography is a reductive process. It was discovered by Jaroslav Heyrovsky in the 1920s. Voltammetry has now become widely used in biological, physical and inorganic chemistry where such processes such as electron transfer, adsorption, oxidation and reduction are studied.

During a voltammetric measurement, as the applied potential attains either a positive or negative value dependent on the electroactive species present in solution, electrolysis of the species occurs leading to the generation of a current which is deemed the *faradaic current* (Wang, 1994). This current varies as the potential

changes and is proportional to the concentration. In addition to the faradaic current there is the *non-faradaic current* which comprises of the background current inherent of the circuit and is caused by an "electrical double layer" on the surface of the working electrode (Wang, 1985). The size of the current affects the overall detection limit of the technique.

From the late 1950s to the early 1960s, the advent of atomic spectroscopic methods saw the decline in utilisation of voltammetric techniques. This was coupled with the expensive equipment available at the time which relied on vacuum tubes (valves), for example a differential pulse polarographic instrument cost $25,000 USD in the early 1960s (Osteryoung & Osteryoung, 1985). Revival of these techniques occurred from the mid-1960's when the "classical" methods were modified, coupled with the availability of cheaper amplifiers, semi-conductors, and hence making these techniques more sensitive. Consequently, the same instrument cost $2,000 USD in the early 1970s mainly due to the advent of transistors.

1.4.2.3 Stripping Voltammetry (SV)

Stripping voltammetry involves a two-step process. The first step involves pre-concentration in which the target analytes are deposited on to the surface of the working electrode via either reductive or oxidative means. This results in the concentration on the working electrode far exceeding that of the ions in the solution. The second step involves the consecutive removal of the target analytes from the surface of the working electrode dependent on the applied potential and the standard electrode potential of the target analytes. Three methods of stripping commonly utilised are: anodic, cathodic and adsorptive. The *anodic* method is normally employed for the detection of species with a positive charge such as positively charged metal species. The *cathodic* method is generally employed for the detection of negatively charged species, such as the halides.

The *adsorptive* method does not rely on an electrical potential to deposit the analyte(s) on to the working electrode but rather is based on physical adsorption. Examples of analytes are organic molecules such as clinical and pharmaceutical compounds, such as riboflavin, and also for the determination of inorganic cations at

very low concentrations which have been associated to complexing agents such as dimethylglyoxime and catechol. An example is the determination of chromium and tungsten ions with carbon electrodes (Brainina, 1995).

1.4.2.4 Anodic Stripping Voltammetry (ASV)

ASV typically involves the application of mercury films to inert working electrodes instead of drops as used in polarography. The diffusion path length from the mercury film to the solution interface is much shorter compared to a drop, and thus analyte ions are "stripped" much quicker. This quicker stripping leads to better resolved peaks, and implies that solutions containing multiple target analytes can be studied. This makes ASV one of the most sensitive techniques in electrochemistry (Skoog *et al.,* 1996).

ASV is used to measure concentrations in the sub-parts per billion (sub-ppb) range ($\mu g\ kg^{-1}$). ASV is typically a two-step process:

1. <u>Deposition</u>: The working electrode behaves as a *cathode* whence the metal ions are reduced and deposited on to the electrode surface at a suitable fixed negative potential for a specific length of time. For example, cadmium is reduced on to the electrode as follows: $\mathbf{Cd^{2+} + 2e^{-} \rightarrow Cd^{0}}$

2. <u>Stripping</u>: The working electrode now behaves as an *anode* as the metals are re-oxidised ("stripped") from the electrode back into the solution as a function of increasing potential, eg -1.2 to -0.2V. The re-oxidation of cadmium from the electrode is as follows: $\mathbf{Cd^{0} \rightarrow Cd^{2+} + 2e^{-}}$.

The most common voltammetric techniques employed in ASV are normal pulse (NPASV), differential pulse (DPASV) and square wave (SWASV).

Normal Pulse and Differential Pulse Anodic Stripping Voltammetry (NPASV and DPASV)

Figure 1.3 displays a linear-scan voltammetry (LSV), an NPASV and a DPASV voltammogram of a solution containing three metal ions: cadmium, lead and copper,

recorded by an Autolab PSTAT10 (Eco Chemie, Holland) using a carbon-ink screen-printed electrode (C-SPE).
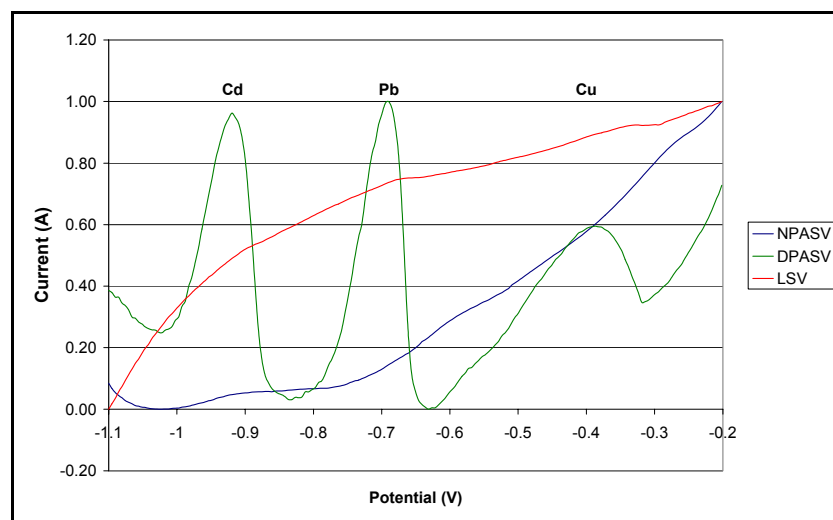
Figure 1.3: LSV (red), NPASV (blue) and DPASV (green) voltammograms recorded for 200 ppb of Cd, Pb and Cu in 200 ppm mercuric (II) nitrate, 1% nitric acid, 0.1M potassium chloride acquired on carbon-based screen-printed electrodes. Deposition potential: -1.1V (vs Ag/AgCl); Deposition time: 165s; Scan rate: 10mV/s. DP amplitude: 50mV; NP and DP pulse durations: 50ms. The currents are range-scaled

The LSV plot (red) clearly displays two plateaux for lead (Pb) and copper (Cu), the latter having a wider potential window due to the copper ions continually being reduced as the potential increases. The plateau for cadmium (Cd) is therefore shorter. The plateaux indicate that the reduced metal species are being oxidised back into the solution. The length of the plateau leads to quantitative determinations of the respective metal ions but at much greater concentrations (>10ppm). The NPASV plot (blue) clearly defines the presence of cadmium compared with LSV.

With regards to DPASV (green), deposition has occurred at a potential of -1.1V versus the stable reference electrode potential (of silver/silver chloride) for a fixed length of time (165s). The potential is increased linearly at a rate of 10mV/s with a differential potential pulse of 50mV applied at the last 50ms of the period of 500ms. The cadmium peak (left) with a stripping potential of ~-0.9V, and lead peak (~-0.65V) are both well-defined but as for copper (right) it is not so well resolved. Some reasons offered by (Eriksen *et al.,* 2001) are with regards to the poor solubility of copper in mercury, copper existing in more than one oxidation state, and the actual

stripping peak of copper occurring at the same potential where certain organic species may also be adsorbed on to the working electrode such as fulvic or humic acids which are both naturally occurring complexing agents in the environment (Bott, 1995). In this situation, multivariate calibration (discussed in Section 1.5.4) may be used to determine the true concentration of the copper and hence overcome these matrix effects.

Linear-scan methods suffered from low scan rates, inconvenient apparatus and poor limits of detection (Skoog *et al.,* 1996). As advances in electronics grew, signals could be better manipulated. NPASV and DPASV differ from LSV in that during the last 50ms of the waveform period, the current is recorded at the end of a potential pulse which grows with each pulse (for NPV), or the current is recorded both before a fixed 50mV pulse is applied and again at the end of the pulse duration (for DPV). The NPASV and DPASV waveforms are displayed in Figure 1.4.
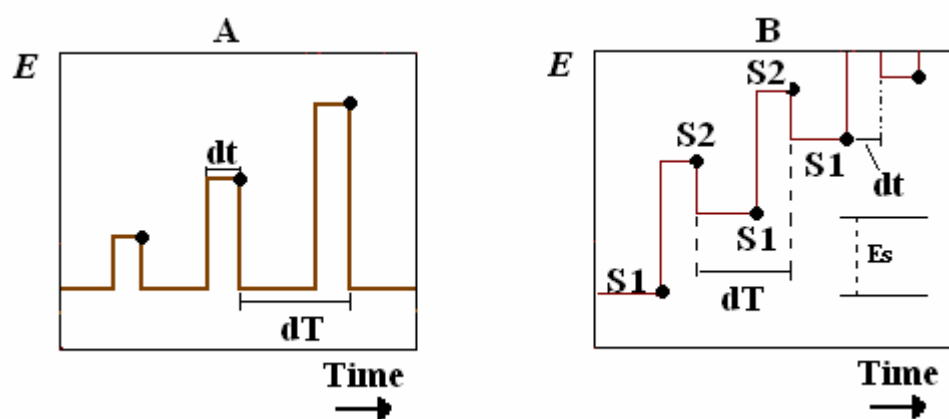
Figure 1.4: Normal pulse (A) and Differntial pulse (B) waveforms. Currents are recorded at ●

The NPASV waveform (Figure 1.4A) leads to a smoother curve than LSV. This is a consequence of the *faradaic current* being highly stable and the *non-faradaic current* negligible (Rouessac & Rouessac, 2000). The DPASV waveform (Figure 1.4B) is combined with a pulse output from a staircase signal where **Es** is the height of the step. The difference in the currents between **S1** and **S2** (**S2 − S1**) is recorded as a function of the linearly increasing potential. This produces more highly resolved voltammograms with peaks similar to those obtained on a chromatogram where the height of the peak is directly proportional to the concentration of the analyte (Figure

1.3). Referring to Figure 1.4B, the scan rate is calculated by dividing the step potential (ΔE) by the period (dT). One of the advantages of DPASV is that the number of maximum peaks generally corresponds to the number of target analytes, unless two or more analytes possess the same redox potential. However, the resolution is much improved with electrode potentials of 0.04 to 0.05V for DPASV compared to 0.2V for LSV.

Square Wave Anodic Stripping Voltammetry (SWASV)

This is a highly sensitive technique which acquires voltammograms rapidly compared to DPASV (Osteryoung & Osteryoung, 1985). It combines a staircase (step) waveform with a "pulse train" to produce the SW waveform illustrated in Figure 1.5.
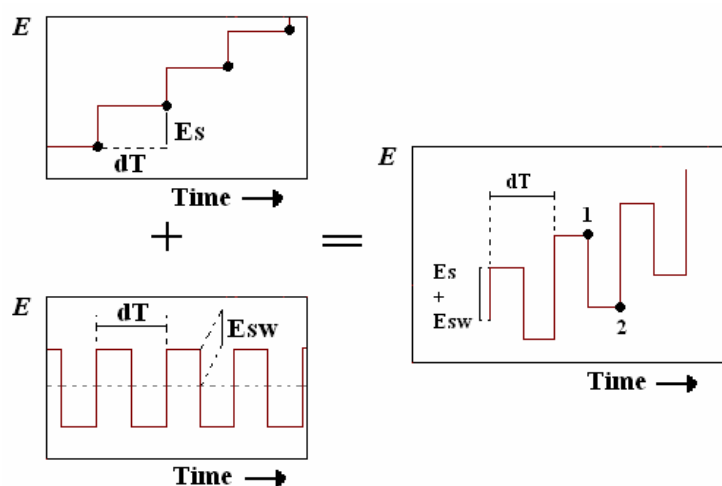
Figure 1.5: Waveforms contributing to a square-wave form (right): staircase (top-left) and a train pulse (bottom-left)

The current recorded at each waveform period is calculated from the current at point **2** subtracted from the current at point **1**. Voltammograms can be recorded in under 10ms.
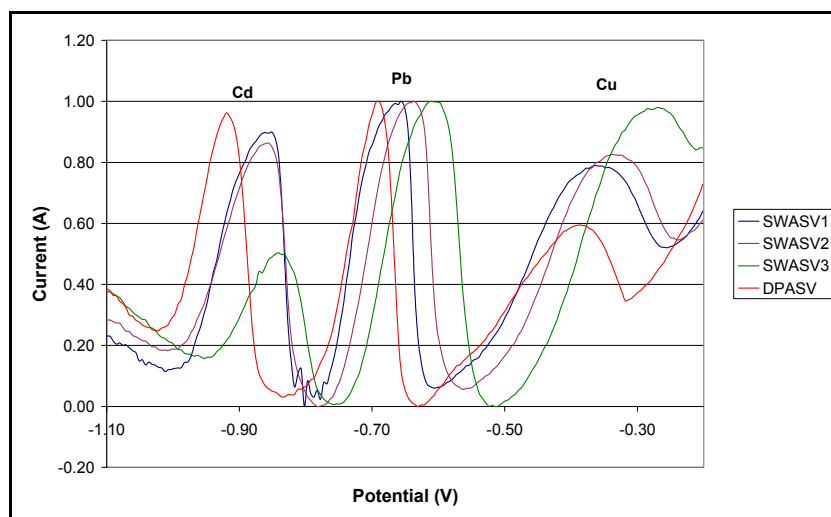
Figure 1.6: SWASV voltammograms of 200 ppb Cd, Pb and Cu in 200 ppm mercuric (II) nitrate, 1% nitric acid, 0.1M potassium chloride. Dep potential: -1.1V (vs Ag/AgCl), Dep time: 165s. Scan rate (DPASV (red)): 10mV/s.
Key: SWASV1 (blue) = Amplitude: 200mV; Frequency: 25Hz; Scan rate: 124mV/s
SWASV2 (purple) = Amplitude: 50mV; Frequency: 25Hz; Scan rate: 124mV/s
SWASV3 (green) = Amplitude: 50mV; Frequency: 50Hz; Scan rate: 248mV/s
The current was range-scaled.

Typical parameters include waveform periods of 5ms, 10mV for **Es** and 50mV (2×25mV) for **Esw**. With a pulse frequency corresponding to 200Hz (1/0.005s), a voltammogram recorded over a 1V range will take 0.5s. Typical voltammograms are displayed in Figure 1.6 (above) and clearly resemble that of Figure 1.5 for DPASV.

The quality of the voltammograms is maintained even with the increase in scan rate. Decreasing the amplitude from 200mV (SWASV1) to 50mV (SWASV2) leads to a slight positive potential shift and peak broadening. Increasing the frequency and thus the scan rate (SWASV2 to SWASV3) leads to further broadening of the peaks and loss of sensitivity, especially for cadmium and copper. SWASV1 resembles DPASV but is greatly affected by noise.

Overall, the DPASV voltammogram is better resolved with narrower and sharper peaks. With regards to the SWASV voltammograms, SWASV2 offers the better voltammograms. It is reported that detection limits for SWV are in the region 10^{-7} to 10^{-8} M for a range of inorganic analytes (Skoog *et al.,* 1996). Further applications are reviewed in Section 2.2 and 3.2.

1.4.2.5 Potentiometric Stripping Analysis (PSA)

PSA differs from ASV in that as soon as the metal ions have been pre-concentrated, potential changes are recorded against time at a constant (null) current. This gives the advantage that the signals obtained are independent of the surface of the electrode in which the size of the electrode is therefore not a factor (Estela *et al.,* 1995). Other advantages include "detection in the presence of some electroactive organic species, analysis in solutions with lower ionic strength (such as acetic acid and propanol at 10^{-4}M) and lower background contributions" (Hocevar *et al.,* 2002). Furthermore, a single potential at for example -1.25V vs SCE suffices to reduce all of the metals in solution and thus avoid hydrogen evolution in water-based systems. It is easier to measure time than micro-currents (DPASV). However, peak resolution and detection limits are improved for DPASV than with PSA (Estela *et al.,* 1995). Furthermore, acquisition times for PSA are greater than for DPASV because of the constant potential as the target analyte is stripped. This is more evident when varying concentrations of the other analytes are present in the same cell. This implies that as the target analyte is stripped the other analytes continue to be deposited.

1.4.3 Fluorescence Spectroscopy

1.4.3.1 The Phenomenon of Fluorescence

When a compound is irradiated with a high-energy (low wavelength) beam of electromagnetic radiation, the molecules or atoms become excited due to the electrons lying in the ground state absorbing the energy and transferring to higher energy states (Atkins, 1992). On removal of the excitation beam, the excited electrons return to the ground state at a lower energy (higher wavelength) emitting light (*fluorescence*) in the process. This shift in wavelength is termed the *Stokes shift* and is due to a portion of the absorbed energy being dissipated to the surrounding medium in the form of heat as a result of the electrons undergoing *vibrational relaxation* (Skoog *et al.,* 1996). Figure 1.7 illustrates the phenomena via the so-called Jablonski diagram.
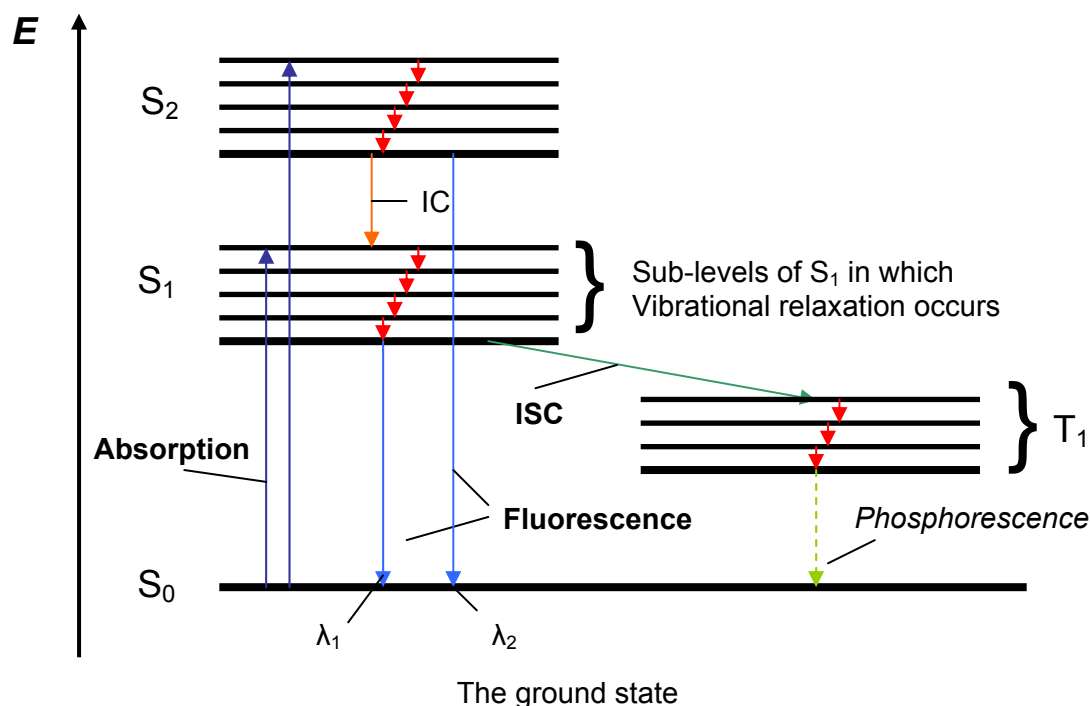
Figure 1.7: Jablonski diagram illustrating the electronic transitions that result from the irradiation of a compound with high-energy beam of photons

The term internal conversion (IC) describes the process by which electrons lose excess energy in the form of heat by transferring from the lowest sub-level of an excited state (S_2) to a lower excited state (S_1). Intersystem crossing (ISC) occurs when paired electrons from the excited singlet state (S_1) transfer to the more energy-stable triplet state (T_1) (Atkins, 1992). However, the electrons attain parallel spin which prevents them from rapidly returning to the ground state (S_0) and thus do so via the longer-lived phenomenon of *phosphorescence* (Kemp, 1991).

All compounds which absorb high-energy photons possess the ability to fluoresce. However many do not due to radiationless relaxation processes which occur at a faster rate than fluorescence (Skoog *et al.,* 1996). This can be attributed to the structure of the compound. Aliphatic hydrocarbons containing double or triple bonds tend to fluoresce, but fluorescence is more common in non-substituted aromatic hydrocarbons. The quantity of fluorescence is described as the *quantum yield*, Q_e (Equation 1.3):

$$Q_e = \frac{r_f}{r_f + r_r} \qquad\qquad \textbf{1.3}$$

where r_f is the rate of relaxation via fluorescence and r_r is the rate of relaxation via radiationless processes. If a compound possesses a Q_e value close to unity, it will be highly fluorescent; the opposite is true if Q_e is close to zero. With regards to un-substituted aromatic rings, the Q_e increases as the number of fused rings increases. The *rigidity* of a compound also affects the quantum yield in that the more rigid a compound, the greater is the yield. For example, fluorene possesses a Q_e close to 1 whilst biphenyl is ~ 0.2 (Figure 1.8).
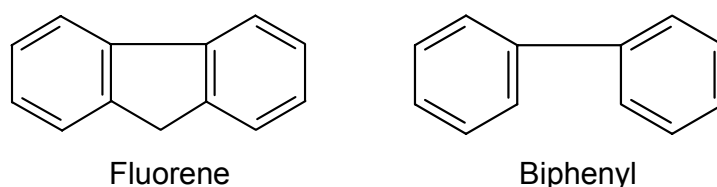
Fluorene   Biphenyl

Figure 1.8: The molecular structures of fluorene and biphenyl illustrating the rigidity of the respective compounds

Furthermore, any substitutions made on the rings can result in a shift of the fluorescence wavelength and subsequent lowering of Q_e.

The phenomenon of fluorescence is applicable in a wide number of areas such as the detection of leaks in water systems and markers for rescue operations performed at sea (both employing *fluorescein*). Another application is the addition of "brighteners" to detergents. Whilst fabrics are being washed, the brightening compounds adhere to them. When the fabrics are thus illuminated by sunlight (absorbing in the UV region) they fluoresce in the blue region and hence give the effect of "whiter than white" (Kemp, 1991).

1.4.3.2 Relating Concentration to Fluorescence

Fluorescence intensity is proportional to the intensity of the excitation energy that has been absorbed by the target species. We thus have Equation 1.4.

$$I_F = K'(I_O - I)$$ **1.4**

where I_F is the intensity of the fluorescence, K' is a constant dependent on the quantum efficiency of the fluorescence occurring, I_O is the intensity of the incident light, I is the intensity of the light after passing through a sample of path length l, and therefore the difference $(I_O - I)$ relates to the intensity of the light absorbed. The Beer-Lambert Law relates the absorbance (A) to the concentration (c) of a given compound (Equation 1.5).

$$A = \varepsilon c l$$ **1.5**

where ε is the molar extinction coefficient (or molar absorptivity) and l is the path length of the sample cell. The absorbance is obtained from the natural logarithmic ratio of the intensity of the incident light (I_O) over the intensity of the light (I) after emergence from the cell (Equation 1.6).

$$A = \ln\left(\frac{I_O}{I}\right)$$ **1.6**

Combining Equations 1.5 and 1.6 and substituting into Equation 1.4 gives:

$$I_F = K'I_O\left(1 - e^{-\varepsilon c l}\right)$$ **1.7**

The intensity of fluorescence is proportional to the concentration. If the absorbance remains below 0.05, a linear calibration curve will be attained. However, on absorbance values greater than 0.05, linearity is lost due to the phenomenon of *self-quenching* in which the molecules of the given analyte absorb the fluorescence

produced by other neighbouring analytes. The effect therefore increases as does the concentration of the given analyte (Skoog *et al.,* 1996).

1.4.3.3 2D Fluorescence

A spectrum which contains the fluorescence intensity values of a given target compound over a specific wavelength range is classed as two-dimensional (fluorescence intensity versus excitation or emission wavelength). Three types of spectra are attainable: *emission*, *excitation* and *synchronous*.

Emission Spectroscopy

The emission spectrum of a target compound is obtained after the sample has been irradiated with a fixed wavelength excitation source. Figure 1.9 displays the main components of a 2D fluorimeter in the acquisition of an emission spectrum.
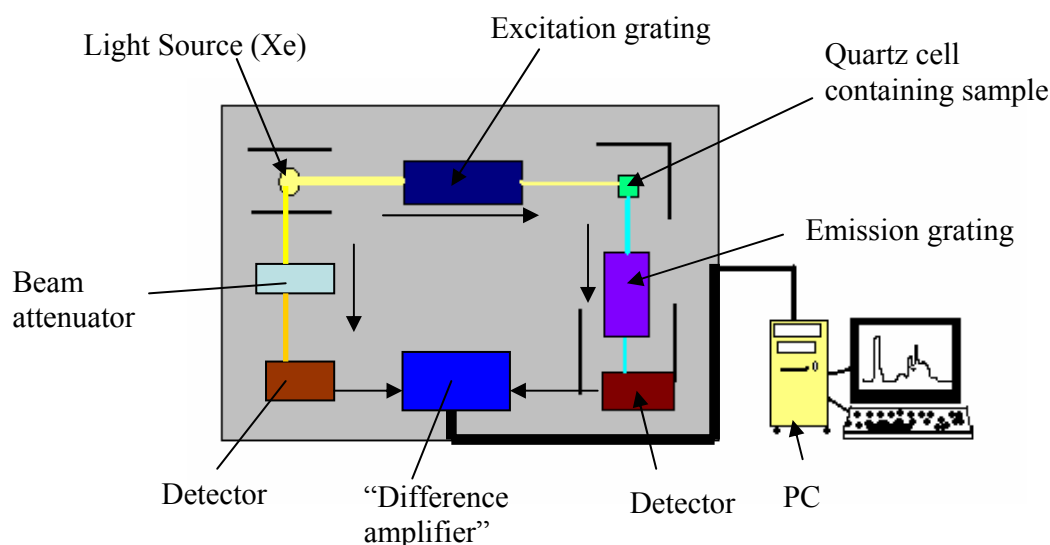
Figure 1.9: The main components of a typical 2D fluorimeter leading to the acquisition of an emission spectrum

The light source in this case is a Xenon arc lamp, which has a wavelength range of 200 to 600nm. The light beam passes through an excitation grating or filter, which produces an excitation beam across a narrowly defined wavelength (i.e. a narrow bandwidth). This then passes through the sample. The fluorescence occurs in all

directions. However, common practice is to acquire it at right-angles to the excitation beam. If at other angles, the effect of scattering from the quartz cell walls and the sample solution itself can lead to large errors in the intensity values recorded.

The emitted beam passes through the emission grating which can be focused to select one particular wavelength, or scanned over a range of wavelengths. These are collected at a detector (for example a photomultiplier). In many fluorimeters light from the source is also passed through a beam attenuator which reduces the beam intensity and is collected at another detector. The outputs from the two detectors are combined and processed by the difference amplifier followed by transference to a processor for further processing and displaying. This type of instrument is classed as *double-beam* due to the presence of the two detectors which serves to take into account power fluctuations in the source. This also serves in calibrating the instrument for day-to-day use.

Figure 1.10 displays the emission spectrum of 400ppb anthracene in HPLC grade acetonitrile.
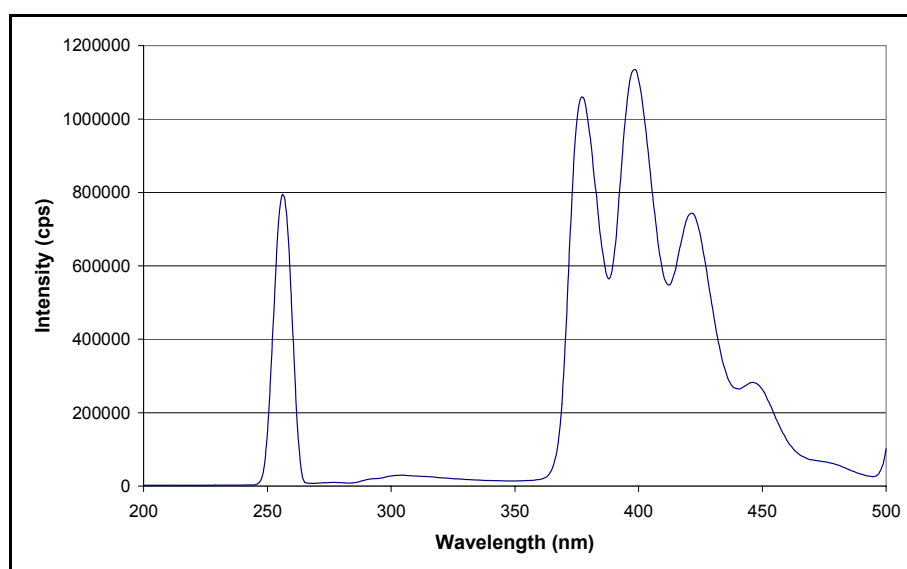
Figure 1.10: The fluorescence emission spectrum for 400 ppb anthracene in HPLC grade acetonitrile. Start: 200nm; End: 500nm; Excitation: 254nm; Increment: 1nm; Integration: 0.1s

The peak at 254nm is due to the phenomenon of *Rayleigh scattering*. This is where the particle size of the given sample is smaller than the wavelength of the excitation

source (Atkins, 1992). This leads to elastically back-scattered light from the excitation source. The intensity is dependent on $1/\lambda^4$ and is more effective at shorter wavelengths. Four emission peaks are clearly visible at 380, 400, 420 and 450nm. This would lead to the deduction that there are four sub-levels in the excited state (S_1). However if this was the case then four sharp lines would be observed at each of the above wavelengths. The reality is that there are a large number of sublevels which leads to the broad spectrum in Figure 1.10.

Excitation Spectroscopy

The excitation (absorption) spectrum of a compound is acquired by scanning the excitation grating at a constant emission wavelength. The excitation beam of varying wavelengths traverses the sample and is reflected by an arrangement of mirrors into the emission grating as illustrated in the simplified diagram shown in Figure 1.11.
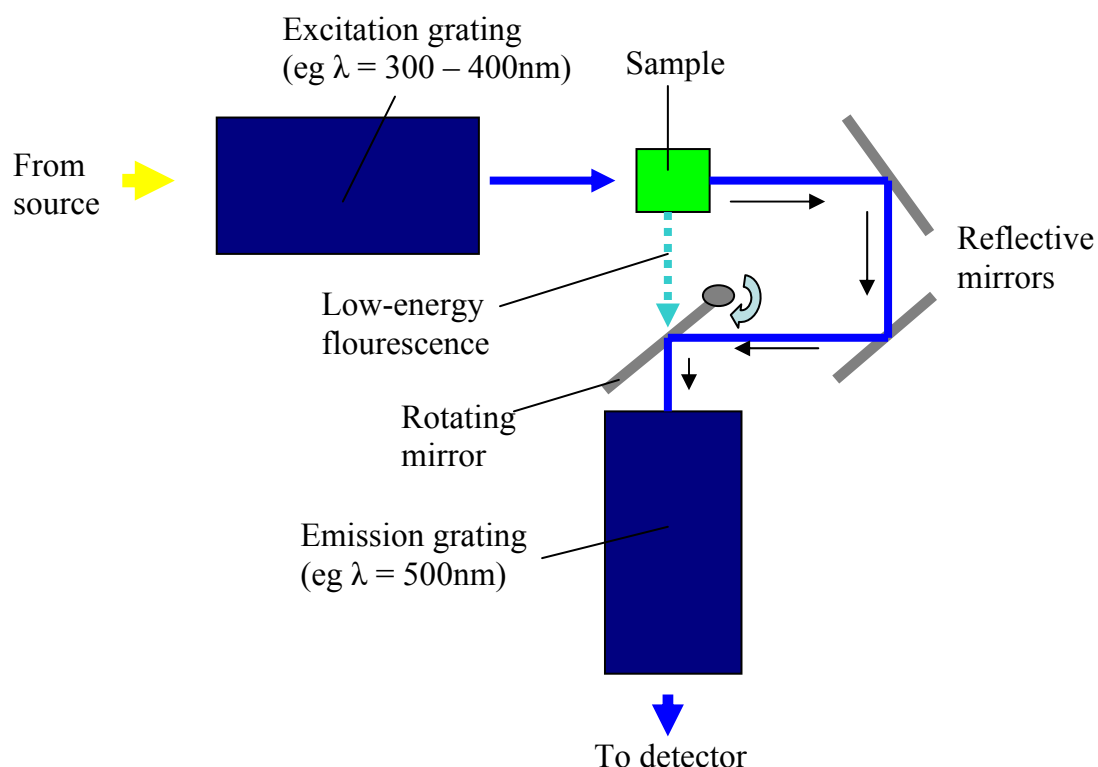
Figure 1.11: Typical instrument set up for acquisition of an excitation (absorbance) spectrum

The emission grating is set at a high wavelength in order to reject the fluorescent light which accompanies the excitation light. These fluorescent beams are of lower energy,

and with wavelengths less than 500nm (in this instance). The rotating mirror can change position so that the instrument can be set to "emission" mode. In some systems, additional shutters in front of the mirrors can prevent any stray fluorescent light from affecting the measurement and thus preventing interference.

Figure 1.12 displays both the excitation and emission spectra for 400 ppb anthracene in HPLC grade acetonitrile.
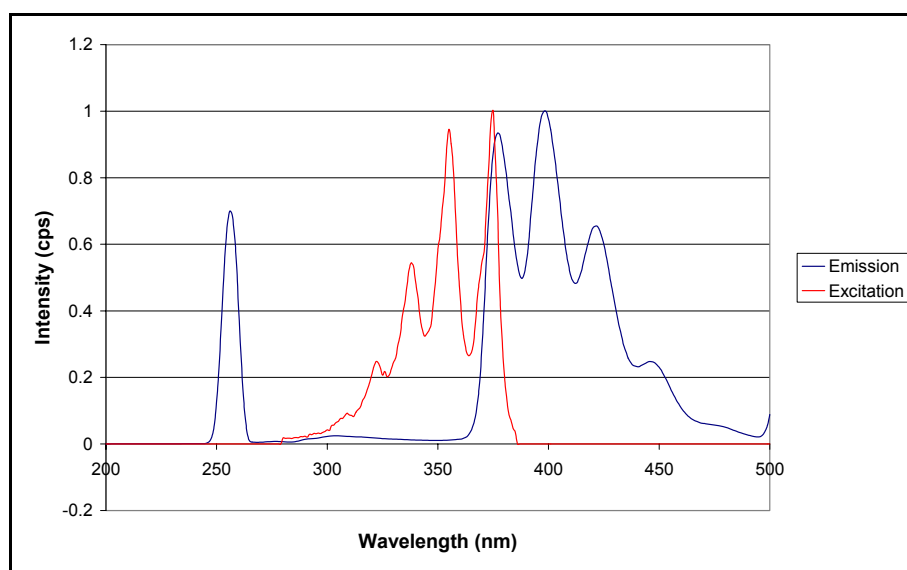
Figure 1.12: The excitation (absorption) spectrum (RED) and emission spectrum (BLUE) for 400 ppb anthracene in HPLC grade acetonitrile. Emission parameters detailed in Figure 1.10. Excitation parameters: Start: 280nm; End: 385nm; Emission: 400nm; Increment: 1nm; Integral: 0.1s

The absorption spectrum is a "mirror image" of the emission spectrum (Figure 1.10). This illustrates the energy absorbed by the electron as it is promoted from the ground state (S_0) to the higher energy excitation state (S_1, S_2, etc). Furthermore the difference in energy between the sublevels in both the ground and excited states are the same (Skoog *et al.,* 1996). The peak at 375nm appears in both spectra and is a product of "resonance fluorescence". This occurs when the emission wavelength is identical to the excitation wavelength that generated it.

Synchronous Spectroscopy

Synchronous fluorescence spectroscopy (SFS) involves the simultaneous scanning of both the excitation and emission gratings at a constant scan rate (Patra & Mishra, 2002). This involves scanning the excitation and emission gratings at the same constant rate to give a constant wavelength interval ($\Delta\lambda$). SFS can be better described in a Jablonksi diagram (Figure 1.13).
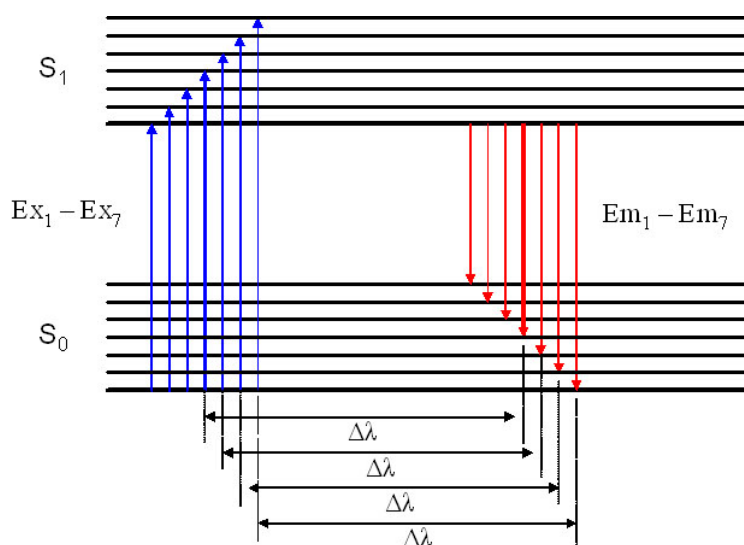
Figure 1.13: Explaining the phenomenon of SFS at constant scan rate via a Jablonski diagram

Normal emission spectroscopy would involve excitation at Ex_4 (where in this instance the absorption would be at its maximum) and scanning the emission grating from Em_1 to Em_7. With SFS, the wavelength interval is set constant so that $\Delta\lambda = Em_4 - Ex_4 = Em_5 - Ex_5$ and so on. Additionally, no fluorescence is detected until the excitation and emission gratings reach Ex_4 and Em_4 respectively. Excitation then occurs at Ex_5, Ex_6, etc resulting in the respective fluorescence emissions at Em_5, Em_6, etc. The full spectrum is thus acquired in this manner. This therefore results in narrower spectral bands compared to normal emission spectra. Depending on the choice of $\Delta\lambda$ the peak intensities can be increased further due to the simultaneous scan of both the gratings.

Figure 1.14 displays the synchronous spectrum of 400 ppb anthracene in HPLC grade acetonitrile.
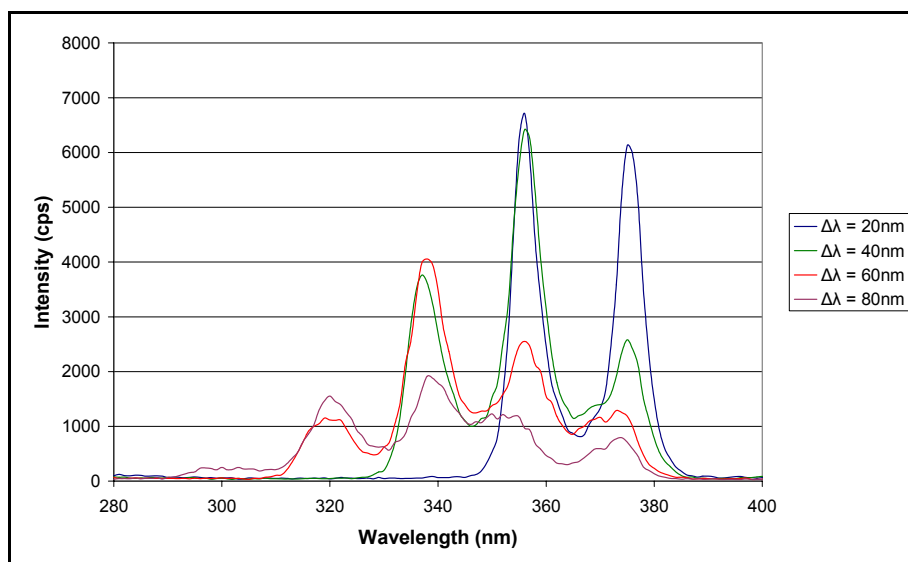
Figure 1.14: Synchronous fluorescence spectra of 400 ppb anthracene in HPLC grade acetonitrile at varying synchronous wavelengths. Start: 280nm; End: 400nm; Increment: 1nm; Integration: 0.1s

In this instance the application of a small synchronous wavelength difference ($\Delta\lambda$ = 20nm) has led to the acquisition of well resolved and narrower peaks, the quality and resolution decreasing as $\Delta\lambda$ increases. This can be related to the Jablonski diagram in Figure 1.13 in that fewer electron transitions occur from the excitation state (S_1) down to the ground state (S_0). This technique has therefore made the peaks much sharper and narrower compared to emission and excitation spectroscopy. It will thus allow for better deconvolution of samples containing a mixture of target analytes leading to the extraction of much useful data.

1.4.3.4 3D Fluorescence

A disadvantage of 2D fluorescence is the requirement to perform a number of measurement steps in order to determine the optimum excitation and emission wavelengths for a particular sample (Rouessac & Rouessac, 2000). These steps involve the following:

- Acquiring a UV-VIS (absorption) spectrum via a spectrophotometer. The maximum intensity obtained determines the *initial* excitation wavelength

- The fluorescence (emission) spectrum is acquired at the above *initial* excitation wavelength. The maximum peak obtained determines the emission wavelength

- The excitation spectrum is acquired at the above emission wavelength. This leads to the determination of the *actual* excitation wavelength to be employed for the particular sample or compound.

- The emission spectrum is acquired again at the *actual* excitation wavelength

These steps are cumbersome if they are to be followed for every sample containing one compound. However, if a sample contains a mixture of compounds with different fluorescent properties, the optimum excitation wavelength for one compound may not be so for another compound, and will lead to 2D spectra which are difficult to interpret. This issue may be addressed, at least in part, by 3D fluorescence.
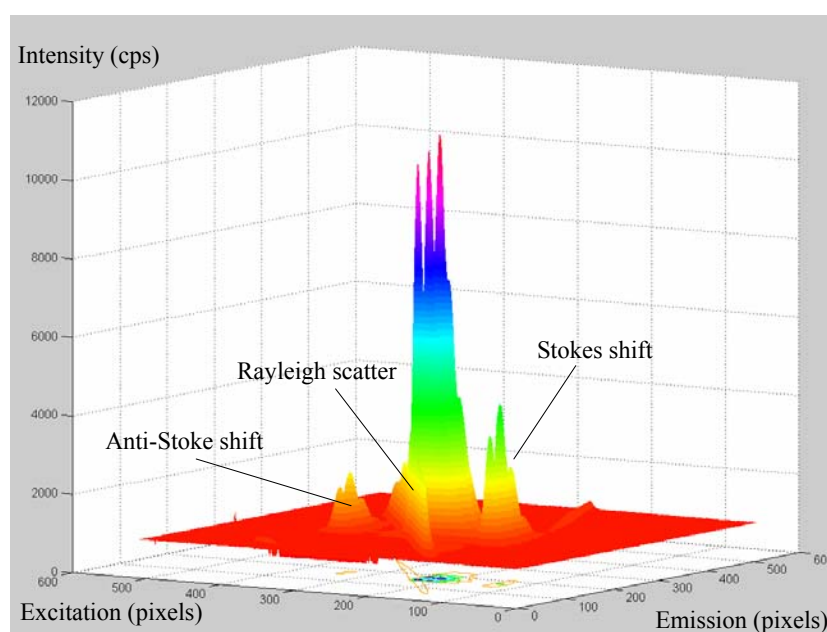
Figure 1.15: EEM (3D) spectrum of 400 ppb anthracene in HPLC grade acetonitrile acquired on a CCD detector. Excitation (Z): Start: 73.9nm (1); End: 691.4nm (511); Resolution: 1.2 Emission (X): Start: 227.7nm (1); End: 724.5nm (511): Resolution: 1.0; Exposure: 1s

Figure 1.15 displays an excitation-emission matrix (EEM) which is a 3D plot with the emission wavelengths in the x-axis, the excitation wavelengths in the z-axis and the intensities in the y-axis. 3D fluorescence involves acquiring emission spectra ranging from specified wavelengths over the desired excitation wavelength range.

This arrangement permits the instantaneous acquisition of the spectrum and thus immediate interpretation of the data. With a sample containing a mixture of fluorescent compounds it will be easier to determine the excitation wavelengths to employ for a specific compound on the 2D fluorimeter. However, with the advent of powerful computers and advanced data handling software applications, the 3D data itself can be quantitatively processed (Chapter 6). The anti-Stokes emission shown in Figure 1.15 originates due to the collection of energy from already-excited sample molecules and thus emit at higher frequency (energy) (Atkins, 1992). Figure 1.16 illustrates the instrumentation employed for the acquisition of an EEM.
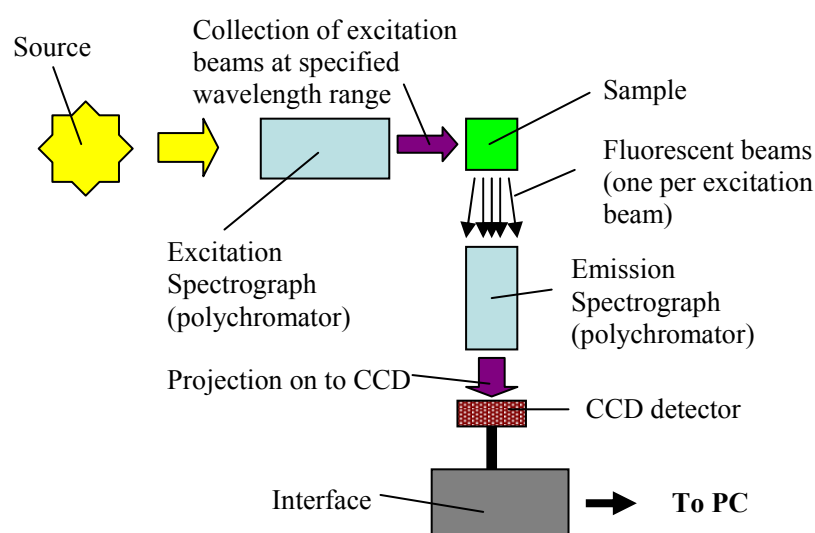
Figure 1.16: Instrumentation for 3D fluorescence spectroscopy

The excitation spectrograph generates excitation energy over a specified wavelength range and dispenses it in a manner such that they strike the cuvette in a vertical plane. The emission spectrograph disperses the resultant emission fluorescence in a horizontal plane before projection on to the *charge-coupled device* (CCD) detector. The CCD employed to generate Figure 1.13 had a dimensionality of 511 rows each with 511 elements, implying a total of 261,121 elements (pixels) each of which can be seen as individual detectors (Skoog *et al.,* 1996). Each pixel can have a surface area of 20 - 30μm^2 (Epperson & Denton, 1989).

When a CCD pixel is illuminated, a charge packet develops within the *p*-type semi-conducting metal oxide beneath the insulating silica surface. Electrons are thus stored in the packet. The charge packets are then transferred along the row towards a low-capacitance diode leading to the induction of a change in voltage which is proportional to the amount of charge. The charge from each pixel can also be combined (binning). Two types of binning are *serial* and *parallel*. The former involves combining a number of charge packets from a given number of pixels, whilst the latter involves combing rows of pixels into one row and combining a specified number of elements in this row. This leads to an increase in sensitivity, a decrease in readout noise, and a faster acquisition due to fewer ADC conversions performed. Shutters are also employed to reduce the exposure of the CCD to the light and thus reduce the risk of charge-overflow from one pixel into a neighbouring one.

CCDs were invented in the early 1970s. They were primarily designed for imaging but later saw application with spectroscopic techniques. CCDs are highly sensitive due to factors such as a high quantum efficiency, charge integration for a few hours, negligible dark current due to cooling at low temperatures (-125°C), and the most important factor is the low read-out noise. CCDs are employed in astronomy, Raman spectroscopy, and capillary zone electrophoresis (CZE) coupled with laser-induced fluorescence (LIF) which led to the differentiation of fluorescent species by their emission spectra and rates of migration due mostly to the time-delayed integration (TDI) performed on the CCD output (Sweedler *et al.,* 1991).

1.4.3.5 Applications

2D fluorescence techniques, in this instance emission spectroscopy, are reported to be more sensitive than absorption techniques due to the ability to enhance the excitation beam in the former in addition to amplifying the signal from the detector (Skoog *et al.,* 1996). Fluorescence spectroscopy offers a relatively inexpensive means of accumulating data compared to GC-MS and HPLC-MS. Important applications include the analysis of food products, clinical samples, pharmaceutical formulations and environmental pollutants, for example polynuclear aromatic hydrocarbons such as anthracene.

Over the last decade, 3D fluorescence techniques have seen a marked increase in application to environmental analysis coupled with multivariate analysis. Two examples are the analysis of humic substances ((Antunes & Esteves da Silva, 2005)) and determining triphenyltin in sea-water ((Saurina *et al.,* 2000)). Further applications include detecting and identifying plasticizers in polymer chemistry, and studying proteins that contain fluorescent groups either naturally in the form of aromatic amino acids or which have been added as labels (Kemp, 1991). Further applications are reviewed in Section 5.2.

1.4.4 Atomic Absorption Spectroscopy

The analytical technique of atomic absorption spectroscopy (AAS) is primarily employed for the quantitative determination of metals (Sharp, 1990). When a sample containing a specific target analyte is irradiated with an appropriate source, the atoms of the analyte absorb the energy and become excited (Rouessac & Rouessac, 2000). In each atom, the outer electrons in the ground state are "promoted" to higher excited energy states. The total loss of intensity recorded by the detector therefore corresponds to the total absorption of energy.

The initial conception of the technique came as a result of a set of experiments performed by Kirchhoff (in the late 1800s) in which sodium chloride was sprinkled on to the flame from a Bunsen Burner which itself was situated between a source of white light and a slit which directed the light on to a prism which dispersed the light producing a continuous spectrum. When this spectrum consisted of black lines on a white background, the phenomenon of atomic *absorption* was conceived. Omission of the light source led to white lines on a dark background, and thus the phenomenon of atomic *emission* was conceived.

The absorption recorded by an instrument is dependent upon the number of atoms (currently in their ground-state) that are in the optical path. This in turn can be related to the concentration of a given element via a "Beer-Lambert Law-like" equation; "law-like" because the molar absorptivity (ε) is not calculated (Rouessac & Rouessac, 2000).

$$A = k \cdot C \qquad\qquad \textbf{1.8}$$

where A is the absorbance and C is the concentration. The constant k is a unique value of a specific element at a specific wavelength.

AAS instruments are purely laboratory-based because of the need for a constant gas supply (acetylene or nitrous oxide) as for flame AAS or in some cases water for constant cooling (as in graphite-furnace AAS). They can therefore be employed to validate field-based instruments such as voltammetric in conjunction with other laboratory techniques such as inductively coupled plasma mass spectrometry (ICP-MS). A common disadvantage with AAS is the requirement to swap the lamps over whenever a new element is about to be studied. This can be deemed a cumbersome and time-consuming process. However, modern instruments can contain an array of lamps which can be automatically rotated into position.

Three types of AAS techniques will be briefly discussed: flame AAS, flame-less AAS and hydride generation. The application and thus choice of these techniques is dependent upon the target analytes and the expected concentrations within a given sample.

1.4.4.1 Flame AAS

In brief, when the sample solution is aspired into the system, it is *nebulised*, ie. made into a fine aerosol, prior to being carried over to the *atomiser* where it is vaporised (Skoog *et al.,* 1996). At this point, any compounds present will be broken down producing the free metallic atoms. When the light from the lamp hits these atoms, they are excited by absorbing the energy of the light, hence leading to the phenomenon of *absorption*. The *photomultiplier detector* detects the change in the intensity of the light at a specific wavelength (specific to the metallic element) corresponding to the energy absorbed, which is represented as absorbance.

At an atomic level, an electron, which is orbiting a nucleus within a given orbital shell (for example 3s) can absorb energy and hence move to a higher orbital shell (for example 3p or 3d) leading to *excitation* (Skoog, 1985). If the electron attained enough

energy to break away from its orbit of the nucleus, the atom would be *ionised*. An electron in a higher orbital shell therefore possesses more energy than an electron in a lower orbital shell. The transition of the electron from one orbital into another occurs at a specific wavelength.

A commonly employed flame atomiser is the *laminar flow burner* (LFB). The sample is nebulised by the flow of the oxidant past a capillary tip. It is then mixed with fuel, which then flows past a series of baffles that remove the very fine droplets. A major drawback here is that most of the sample is wasted away. The remainder is taken up by the oxidant and fuel, and burnt in the flame, which is usually 5 to 10 cm long. There is a risk of an explosion in the mixing chamber if a flashback occurs. The sensitivity is very high but sample wastage is large (90 − 95%). The flame can be made up of a mixture of gases and the temperature of the flame varies depending on this mixture. For example, a combination of natural gas and air produces a flame with a temperature ranging from 1700 to 1900°C whereas a combination of acetylene and oxygen produces a flame with a temperature ranging from 3050 to 3150°C. Other combinations include hydrogen and air (2000 – 2100°C), acetylene and air (2100 – 2400°C), and acetylene and nitrous oxide (2600 – 2800°C).

The most commonly used light sources are the *hollow cathode lamps* (HCL). The cathode is constructed of the metal whose spectrum is to be measured. The inert gas is ionised when a potential of 300V is applied across the electrodes. The corresponding current is between 5 and 10mA. Gaseous cations are generated and have enough energy to knock metal atoms from the cathode forming a cloud. This phenomenon is known as *sputtering*. These atoms are in their excited states and when they fall to their ground states, they emit the necessary radiation at the specific wavelength to excite the atoms in the flame. The metal inside the lamp also gets re-deposited on the cathode or the glass. Modern designs ensure that re-deposition occurs on the cathode.

The higher the potential, the greater the intensities but this in effect increases the *Doppler broadening* (DB) of emission lines. DB occurs due to the rapid motion of the atoms in the flame (Atkins, 1992). When the atoms move *towards* the detector, the wavelength of the emitted or absorbed radiation *decreases*. When the motion is *away* from the detector, the wavelength *increases*. If the flame temperature is increased, the

atoms move faster and thus the DB is heightened. When an increase in current occurs, more unexcited atoms are generated which will absorb some of the radiation from the excited atoms and hence reducing the intensity. This is known as self-absorbance (Skoog *et al.,* 1996). The life of a hollow cathode lamp is related to the maximum current applied. Figure 1.17 shows the schematic instrumentation of flame AAS.
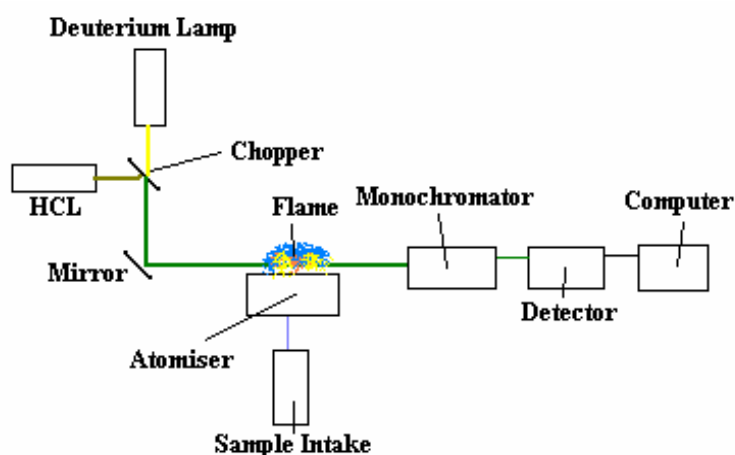
Figure 1.17: Schematic of the instrumentation for Flame AAS

Flame AAS tends to suffer from *matrix effects* (Skoog *et al.,* 1996). Normally, when standards are prepared, they differ from the sample in that they do not contain any interfering species as do the samples. These interfering species can also be present in the flame absorbing or even emitting radiation thus increasing or decreasing the absorbance observed.

Matrix effects can be eliminated by the use of internal standards or by performing standard additions (Kennedy, 1990). Instrumentally, these effects can also be eliminated. One such method involves a *continuous source method* in the form of a deuterium (D_2) lamp (Figure 1.17). This covers the wavelength range of 190-425nm (Lajunen, 1992). A rotating chopper causes the two beams (one from the HCL, the other from the DL) to alternately pass through the flame containing the analyte atoms to the detector. The slit-width is kept large (~0.2-0.7nm) so that any absorption of the continuous radiation by the analyte is neglected. This implies that any attenuation in the intensity of the continuous source is due to the interfering species. As the detector records the intensities from the HCL and the DL it computes a ratio of I_{DL}/I_{HCL}. This

leads to the background being corrected. In order to avoid errors, the two beams must be aligned exactly. Note that the more the intensity decreases, the greater the absorbance increases.

The *monochromator* allows the passage of light of a sufficiently narrow wavelength range. Interference filters can be used but diffraction gratings are more common. The process of *constructive* or *destructive* interference occurs due to the angle of deflection of the light beam. Normally, the band pass of the monochromator is usually greater than the line-width of the source. The *photomultiplier tube detector* is located after the monochromator. The internal construction consists of well-focussed dynodes, which are connected to each other via a resistor. The dynodes closer to the anode have a more positive potential than do the other dynodes nearer the cathode. When the light hits the cathode, the emitted electrons get attracted to the first dynode since it is more positive. This leads to the emission of a number of secondary electrons for every primary electron striking the dynode. The whole process is repeated until it reaches the anode. The signal at the anode gets amplified and the built-in computer represents the information as absorbance or as concentration (e.g. in ppm).

1.4.4.2 Flameless AAS

Electrothermal atomisation is a "flameless" technique (Skoog *et al.,* 1996; Lajunen, 1992). It has an advantage over the flame methods since it does not waste the sample as in LFB or suffer from incomplete atomisation (TFB). Sensitivity is improved for the sample is atomised in a short period. Furthermore, only a minimum amount of solvent is required. The sample is first heated until it is dried before subjection to ashing. The current is increased in order to atomise the sample. The absorbance of light is measured at this point.

A major disadvantage tends to be the reduction in precision due to the matrix effects being more severe. Another disadvantage is that the sample can be contaminated by the graphite tube due to some of the previous sample "slipping" into the pores. This is avoided by coating the graphite tube with a thin layer of *pyrolytic carbon* which seals the pores in the tube. Use of a *L'vov platform* controls the atomisation by delaying the

heating of the sample until the walls and vapour have reached a steady state temperature.

1.4.4.3 Hydride Generation AAS (HG-AAS)

This method is useful for those elements whose atoms are not well excited in a flame - even a nitrous oxide/acetylene flame - due to poor sensitivity (Lajunen, 1992). Instead, the *hydride* of the element is formed; the hydride tends to be volatile so the atoms will be easily excited - sensitivity is in fact improved greatly (at least 2-3 orders of magnitude higher).

The acidified solution is passed through a 1% solution of *sodium borohydride*; the hydride of the element is thus formed and as it passes through the heated quartz tube (~750°C) it is vaporised; the vapour then gets carried into the light path by an inert gas. As matrix effects are completely eliminated, this technique can be employed for the determination of elements such as tin (Sn), silicon (Si), and arsenic (As).

1.4.5 The Acquisition of Multivariate Data

The techniques described in Sections 1.4.2 and 1.4.3 leads to the generation of vast quantities of data. Whereas the absorbance in AAS is acquired at one specific wavelength (univariate), a fluorescence emission spectrum consists of intensity values covering a specified wavelength range (multivariate). This implies that specialised data analytical tools are required to process and interpret the data. Such tools are employed in the field of *chemometrics* which is discussed in the next section.

1.5 Chemometrics

Chemometrics has fast become an important tool in the field of analytical chemistry spanning areas such as NIR spectroscopy, Raman, NMR, UV-VIS spectrophotometry, fluorescence spectroscopy, liquid and gas chromatography (LC and GC respectively) and electrochemistry. The growth in the latter has become substantial over the last two decades, which has seen the development of highly sensitive techniques, like square-wave anodic stripping voltammetry (SWASV) offering a faster acquisition time coupled with the generation of larger datasets. As laboratory instruments have become more complex and sophisticated in acquiring data, it is important to have the means to be able to interpret the data and hence to extract the most relevant characteristics of the data, which will thus lead to the production of comprehensible and reliable results.

Chemometrics is employed in many fields such as biology, organic and inorganic chemistry, pharmaceuticals, food, and industrial processes (Brereton, 2003). It is not restricted merely to the natural sciences; application to sociology, management, marketing and business are now commonplace. In 1972, Svante Wold from Sweden and Bruce R. Kowalski from the United States first introduced the term *Chemometrics* (Otto, 1999b). It is a way of obtaining useful information from raw data such as chromatograms, spectra, and voltammograms. Due to the complex calculations performed, subject areas such as mathematics, statistics, and information technology play a major integral part of chemometrics.

Chemometrics can be split into four distinct fields: *pattern recognition*, *experimental design*, *signal processing*, and *multivariate calibration* (or *modelling*). A brief description of the first three groups will be given followed by an in-depth discussion on the latter.

1.5.1 Pattern Recognition

Pattern recognition originated in the *biological* and *psychiatric* disciplines (Brereton, 2003). In brief, it can be split into two sections: exploratory data analysis (EDA) and classification.

1.5.1.1 Exploratory Data Analysis (EDA)

EDA is used to "make sense" of a large complex dataset by extracting the major characteristics of the elements within the dataset and combining them into a smaller and simplified dataset in which the inherent relationships between the samples can be observed visually. EDA assists in the following:

- The identification of groups of samples providing there are significant differences between these groups
- The variability within the individual groups of samples
- The correct labelling of the individual samples
- Any samples which do not pertain to any of the groups (outliers)

Principal Component Analysis (PCA)

PCA "reduces the number of variables in the original dataset into fewer factors (or *principal components*) without loss in the total variance of the data" (Lam *et al.,* 1997). The matrix comprising the original dataset is reduced to a smaller matrix with fewer variables than the original number of variables. These new variables are called principal components (PC) and are generated from linear projections of the decomposed original matrix. The first PC results from the largest variance observed in the original data. The next PC is orthogonal to the first and captures the next set of variance. Each consecutive PC is orthogonal to the previous PC which results in fewer PCs than the original number of variables due to all the variance of the data captured in conjunction with redundant and/or correlated data variables within the original data matrix.

A full spectral decomposition is thus carried out on the original dataset (*X*). This leads to two matrices: *Scores* (*T*) and *Loadings* (*P*) where *T* relates the rows of *X*, that is the measurement or experiment number, and *P* relates the columns of *X* that is the wavelengths or potentials. (Brereton, 2003; Otto, 1999b; Wise & Gallagher, 1998b; Otto, 1999a; Wise & Gallagher, 1998a). Figure 1.18 illustrates the decomposition of the original dataset.
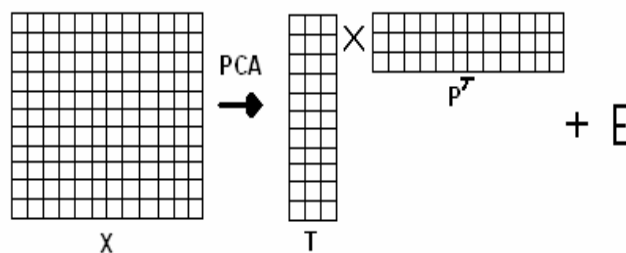
Figure 1.18: Graphical representation of PCA on X.

The number of rows in *T* is equal to the number of measurements (rows in *X*). The number of columns in *T* is the number of PCs. The number of rows in *P* is equal to the number of potential measurements (columns in *X*). The number of columns in *P* is the number of PCs. We thus have Equation 1.9.

$$X = TP^{\mathrm{T}} + E \qquad\qquad \textbf{1.9}$$

The residual matrix *E* possesses the same dimensionality as *X*. It follows that the total variance is generally captured in the first three PCs. This results in uncorrelated data in the form of noise captured in the later PCs. Plotting columns from the scores matrix (*T*) against one another yields scatter plots as shown in Figure 1.19. This allows the observation of the relationships between the samples.
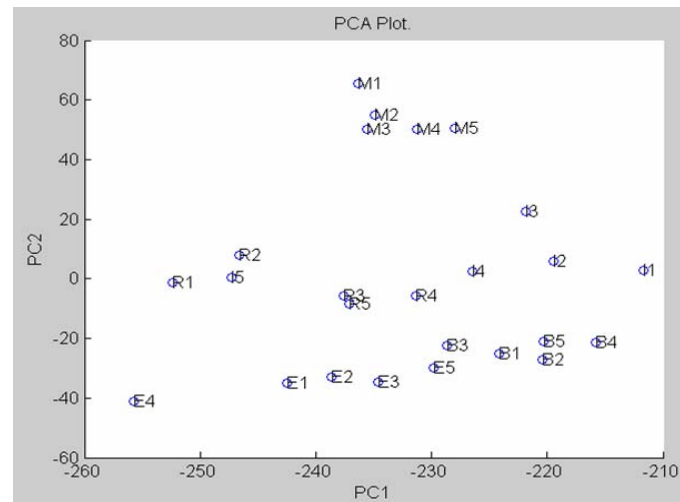
Figure 1.19: PCA score plot of the second principal component (PC2) against the first (PC1) of a randomised dataset comprising of 4 species of pathogenic fungi (E, I, R and M) and 1 blank (B) control. Dataset supplied by _Sahgal & Magan_, Cranfield University, UK (2005).

On close inspection of Figure 1.19 it is observed that the variability within the M and B groups is small, whereas within the others it is large, which is why they are widely distributed within the plot. Furthermore, E5 is more similar to B3 than to the other E samples. Concurrently, I5 is more similar to the R group, particularly R1 and R2, than any other I sample. This could imply mislabelling of the particular sample, and thus lead to the identification of human error. Finally, there appears to be a linear relationship between the two components with regards to the E samples which could be attributed to instrumental drift.

Cluster Analysis

Cluster analysis is another form of exploratory data analysis which detects similarities between groups such as features in organisms which can be body length, age and sex. It involves the creation of a "tree-diagram" which links the similar groups together. This "tree-diagram" is called a _dendrogram_ in which the branch length pertaining to a particular sample depicts its similarity with the other samples. The branch lengths are obtained by the calculation of the relative distances between each sample. The distance is generally calculated via the Euclidean method although other methods are employed (Otto, 1999b). Figure 1.20 displays a dendrogram for the same dataset employed in Figure 1.19.
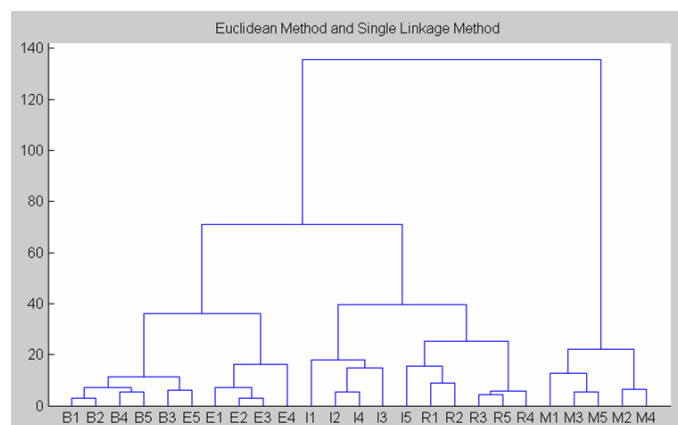
Figure 1.20: A dendrogram obtained after hierarchal cluster analysis of a randomised dataset comprising of 4 species of pathogenic fungi (E, I, R and M) and 1 blank (B) control. Dataset supplied by *Sahgal & Magan*, Cranfield University, UK (2005)

On closer inspection of the dendrogram, five groups (or classes) are observed. Again, it is observed that E5 is more similar to B3 than any other E samples, as is I5 to the R samples. PCA can also be employed prior to cluster analysis if the original dataset contains highly correlated variables. The resulting dendrogram is thus constructed from the samples contained in the first PC (first column if the scores matrix).

1.5.1.2 Classification

This involves the use of "training sets" where the outcomes are already known so that unknown samples can then be "classed" into groups by regression on to the appropriate localised models (Otto, 1999b). A commonly employed method is SIMCA (soft-independent modelling of class analogies). This creates individual models (classes) for the groups observed in PC plots (for example in Figure 1.17). An unknown sample is then classed by regression on to each model until the better model is found. Another method is PLS-DA (partial least squares discriminant analysis). These are employed only if a qualitative determination is required, for example whether the sample is classed as polluted; quantitative determination of the individual components is thus achieved via multivariate calibration (discussed in Section 1.5.4).

Another classification method is support vector machines (SVM). These function by locating a hyperplane within the space of sample inputs. The hyperplane splits the positive samples from the negative samples. The size of the split is dependent on the

greatest distance between the hyperplane and nearest of the positive and negative samples. When classification occurs, the closeness of the test sample to the training samples is computed. SVMs were recently employed to classify meat and bone meal in animal feeds via NIR imaging spectroscopy (Fernandez Pierna *et al.,* 2004).

1.5.2 Experimental Design

Experimental design (also known as *design of experiment* (DoE)) is very important in chemometrics especially when *training data* is required. It ensures that the relationship between the instrument and the target analytes is sufficiently described in addition to any interfering phenomena originating from the instrument and the samples (Martens & Naes, 2001). Furthermore, it is costly to perform numerous experiments to simply obtain a dataset that will consist of the maximum combination of parameters, such as the concentration of four metal ions in different ratios. Experimental design assists in designing a procedure in which the overall number of measurements is reduced.

1.5.3 Signal Processing

The application of chemometrics to the acquired data with respect to signal processing involves background (blank) subtraction, scaling, derivatisation, filtration (smoothing), or a combination of the above (and more). *Background subtraction* can lead to a more linear baseline, and thus enhance the signals attained (Otto, 1999b). *Scaling* of the data generally involves resizing the axes of the variables to the same length. Three examples of scaling techniques are *mean-centring, auto-scaling* and *range-scaling* which are displayed in Equations 1.10, 1.11 and 1.12 respectively.

$$X_{(I,j)}^{MC} = X_{(I,j)} - \overline{X}_{(I,j)} \qquad\qquad \textbf{1.10}$$

$$X_{(I,j)}^{Auto} = \frac{X_{(I,j)}^{MC}}{stdev\left(X_{(I,j)}\right)} \qquad\qquad \textbf{1.11}$$

$$X_{(I,j)}^{Range} = \frac{X_{(I,j)} - X_{(I,j)}^{\min}}{X_{(I,j)}^{\max} - X_{(I,j)}^{\min}} \qquad\qquad \textbf{1.12}$$

Equation 1.10 calculates the mean of each column in a matrix (X) and subtracts it from every element in the respective column vector. Equation 1.11 performs Equation 1.10 followed by division of the standard deviation of each column in the matrix. Equation 1.12 forces all of the elements in the respective column vectors to be in the range of 0 for the minimum value and 1 for the maximum value.

Filtration improves the signal-to-noise ratio (S/N). A number of different algorithms can be employed from *moving average* (MA), *Savitzky-Golay* (S-G), or *Fourier Transform* (FT) of which the S-G function is discussed further. The S-G function can be employed to smooth spectra that suffer from systematic variations in the baseline using a second-order polynomial and a 13-point window (Galvao *et al.*, 2004). The S-G function is superior to MA due to information lost in the latter, for example, on going through a sharp peak, the end of the peak is degraded by the function (Savitzky & Golay, 1964). The S-G filter function is also known as a *least squares smoothing filter* (Wise & Gallagher, 1998b). It is applied to frequency data and spectroscopic peak data. In the latter it is very efficient at preserving such properties as the line width of a peak. However, it can be less effective at removing noise than MA. Furthermore, the S-G filter function requires uniform spacing of the predictor data.

Differentiation (first or second order) usually assists in the resolution of overlapping peaks, but tends to decrease the S/N ratio, leading to differences of opinion as to whether it should be applied; *convolution* involves the combination of two or more functions leading to a "smoother" function. *PCA* can also be utilised to process signals by removing the "undesirable systematic variation in the data" which illustrates the inter-dependency of the groups (Wold *et al.*, 1998).

Orthogonal Signal Correction (OSC) is an alternative approach as a pre-treatment tool (Wold *et al.*, 1998). It was initially developed to find an alternative to pre-processing NIR spectra where much key information is not uncovered. OSC simply removes elements of the X matrix that are completely unrelated (orthogonal) to the Y matrix. The conception, development and application of OSC in conjunction with PLS to NIR data led to improved predictions compared to PLS only (Fearn, 2000).

In general, a combination of data pre-treatment techniques is applied to the data prior to multivariate calibration (Section 1.5.4). However, whatever the dataset is comprised of, a form of pre-treatment may be required (Richards *et al.,* 2002). The main objective of data pre-treatment is to reduce the contribution of noise present in a dataset to a minimum. An increase in the signal-to-noise (S/N) ratio is desired and is expressed as "the height of the most intense peak *divided by* the root mean square of noise where there is no signal". The greater the S/N ratio, the better is the signal (Brereton, 2003).

The term *noise* originates from audio and electronic engineering. It is the unwanted fluctuations observed in signals that appear as static ("hiss") (Skoog *et al.,* 1996). These fluctuations are always present in the output of an analytical instrument. They are a combination of a large number of uncontrolled *random* and *systematic* variables, not only in the instrument, but also within the chemical system under analysis. They also lead to a reduction in sensitivity of the instrument.

The majority of modern instruments can increase the S/N ratio by applying analogue filters, amplification, smoothing and Fourier transforms (Skoog *et al.,* 1996).

1.5.4 Multivariate Calibration

Multivariate Calibration (or Modelling) is a major part of chemometrics. In short, this involves relating a dataset containing, for example, concentration values for a series of analyte solutions to a larger, analytical dataset comprising, for example, the respective spectra, chromatograms or voltammograms. This relation can be applied to, for example, the spectrum of an unknown sample to predict the concentration of the analytes within that sample. Large datasets can be comprised of few rows but numerous columns. Such datasets can be found in NIR spectra, optical such as emission or fluorescence spectroscopy (Skoog *et al.,* 1996), GC chromatograms and voltammograms where each row (vector) in the dataset (matrix) is, for example a voltammogram, and where each element within the vector is a *response* (variable) at a specific wavelength, time or potential.

Numerous multivariate calibration modelling tools are available such as classical least squares (CLS), principal component regression (PCR), partial least squares (PLS) and artificial neural networks (ANNs). They are applied to spectra, chromatograms and voltammograms. The methods, advantages and disadvantages of these tools will be briefly discussed in the following sections.

1.5.4.1 Classical Least Squares (CLS)

This originates from the *univariate* approach where the response of a detector, for example, the *absorbance*, is directly proportional to the *concentration* of one analyte at a given wavelength (Erickson *et al.,* 1992). Equation 1.5 related the absorbance of an analyte with its concentration (*Beer-Lambert Law*):

$$A = \varepsilon c l \qquad\qquad\qquad \textbf{1.5}$$

where A is the absorbance, ε is the molar extinction coefficient (or molar absorptivity), c is the concentration and l is the path length of the cell containing the sample. The absorbance is obtained from the natural logarithmic ratio of the intensity of the incident light (I_O) over the intensity of the light (I) after emergence from the cell (Equation 1.6).

$$A = \ln\left(\frac{I_O}{I}\right) \qquad\qquad\qquad \textbf{1.6}$$

Equations 1.5 and 1.6 are combined into Equation 1.13 leading to a linear relationship between the intensity of the light emerging from the sample and the concentration.

$$\ln(I) = \ln(I_O) + \frac{1}{\varepsilon c l} \qquad\qquad\qquad \textbf{1.13}$$

This relates to the equation of a straight line as denoted in Equation 1.14:

$$y = b_0 + b_1 \times x \qquad\qquad\qquad \textbf{1.14}$$

where y relates is the detector response (ln (I)), x is the reciprocal concentration ($1/c$), b_0 is the intercept through the Y-axis (ln (I_O)) and b_1 is the slope of the line ($1/\varepsilon l$). For a solution that contains multiple components, Equation 1.17 can be rewritten in *matrix* form as shown in Equation 1.15:

$$Y = B \times X \qquad\qquad\qquad 1.15$$

where *Y* is the response matrix containing the absorbance values for the individual components, *X* is the concentration matrix and *B* is a matrix containing the response factors for every component in the solution.

1.5.4.2 Multiple Linear Regression (MLR)

Also known as *inverse least squares* (ILS), MLR relates the concentrations of a specific component (*y*) to a series of recorded spectra or voltammograms (*x*) via a regression vector (**b**) (Wise & Gallagher, 1998b; Otto, 1999b; Otto, 1999a). Observe now that the *Y* and *X* have swapped definitions in comparison to CLS, and from now on, *Y* will refer to concentration or some other property, and *X* will refer to spectra, chromatograms or voltammograms. Each row of *X* will contain a response as a function of wavelength (in the case of spectra) or potential (in the case of a voltammogram), which itself will be related to the same row in *Y*. Therefore:

$$y = Xb + E \qquad\qquad\qquad 1.16$$

where *E* is a "residual" matrix with the same dimensions as *X*. If there is more than one component, Equation 1.16 becomes:

$$Y = XB + E \qquad\qquad\qquad 1.17$$

Figure 1.21 illustrates a situation where three analytes are present in the same solution at different concentrations. Each row of *X* corresponds to a voltammogram as illustrated in Figure 1.3. Each row of *Y* corresponds to the concentrations of each analyte in the solution at different ratios. Each column of *B* is the response factor to the corresponding column in *Y*. It is calculated as shown in Equation 1.18:

$$\boldsymbol{B} = (\boldsymbol{X}^{\mathbf{T}}\boldsymbol{X})^{\mathbf{-1}}\boldsymbol{X}^{\mathbf{T}}\boldsymbol{Y} \qquad\qquad\qquad \textbf{1.18}$$

where $\boldsymbol{X}^{\mathbf{T}}$ is the transposed matrix of $\boldsymbol{X}$, that is to say the rows in $\boldsymbol{X}$ have become the columns and vice versa; a rotation about the main diagonal (left to right). The "**-1**" term indicates that the terms within the brackets have been *inversed* (Hirst, 1991).
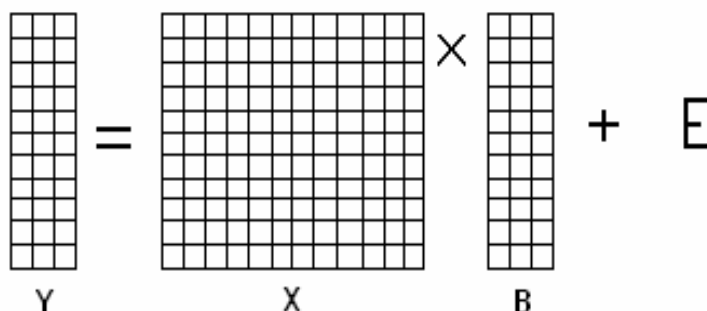
Figure 1.21: Graphical representation of Equation 1.17

Prediction of the concentration of analytes from a new dataset simply involves multiplying the new dataset, $\boldsymbol{X_{unk}}$ by $\boldsymbol{B}$ (Equation 1.19):

$$\boldsymbol{Y_{pred}} = \boldsymbol{X_{unk}}\boldsymbol{B} \qquad\qquad\qquad \textbf{1.19}$$

An advantage of MLR (or ILS) over CLS is that it "minimises the squared errors in concentration" (Torralba *et al.,* 1994). Another advantage is that the number of analytes does not directly affect the analysis. A disadvantage of the technique is that only a small number of wavelengths can be simultaneously looked at (Ni & Gong, 1997). If a large number, the problems of *Collinearity* come into play. This is when some of the columns in $\boldsymbol{X}$ are linear combinations of other columns, that is, there is a "linear dependence among the variables" (Otto, 1999b).

1.5.4.3 Principal Component Regression (PCR)

PCR differs from MLR in that principal components analysis (PCA) is first performed on the $\boldsymbol{X}$ matrix before being regressed to the $\boldsymbol{Y}$ matrix. PCA extracts the "dominant

patterns" in the X matrix (Wold *et al.,* 1987). The regression of T and P to Y is executed via Equation 1.20:

$$B = P((T^{T}T)^{-1}\mathbf{T}^{T}Y) \qquad\qquad \textbf{1.20}$$

Prediction of unknown concentrations occurs as for Equation 1.19. However, if all of the PCs are included, then the same result as for MLR will be obtained (Richards *et al.,* 2002).

1.5.4.4 Partial Least Squares (PLS) Regression

A drawback associated with both MLR and PCR is that they both assume that there are no errors associated with the concentration values, in other words, the Y matrix is not taken into account. PLS, developed by H. Wold, does take this information into account. Both the X and Y matrices are decomposed at the same time so that the *covariance* between them is maximised.
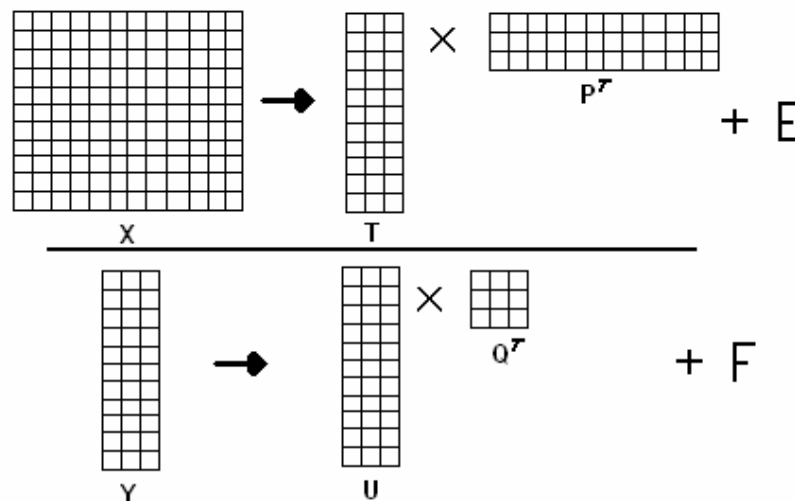
Figure 1.22: Graphical representation of PLS.

U, Q, and F are analogous to T, P, and E respectively where F has the dimensions of Y. In this way, any important information that is obtained from the X matrix is

directly associated with the appropriate portion of the **Y** matrix. The regression matrix, **B**, also known as the "inner relationship" is calculated in Equation 1.21:

$$B = W(P^{T}W)^{-1}Q^{T}$$

1.21

where **W** is the *PLS Weights* which "re-orientates" the **X** matrix "with respect to" the **Y** matrix. The dimensions of **W** are the number of columns of **X** by the number of columns of **Y**. Prediction at the appropriate *latent variable* (LV) is illustrated in Figure 1.23 where an LV is analogous to a PC but incorporates information from the **Y** matrix.
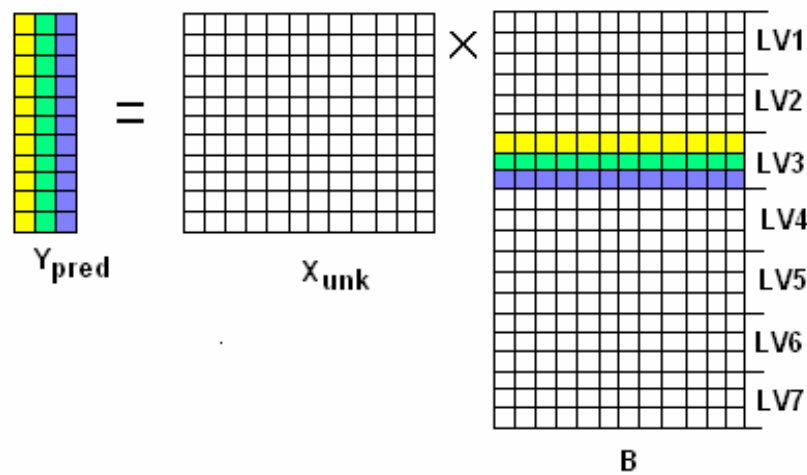
Figure 1.23: Prediction of unknown concentrations at the 3rd latent variable (LV) as denoted by rows 7 to 9 of the regression matrix, B.

In this instance, each row in X_{unk} produces an equivalent row in Y_{pred} from the third latent variable in the **B** matrix.

There are many variants of PLS such as *non-linear iterative PLS* (NIPALS), *statistically inspired modification of PLS* (SIMPLS), and a number of variants including non-linear PLS algorithms. These are described below.

Non-linear Iterative Partial Least Squares (NIPALS)

NIPALS was originally introduced by Wold in 1966. It consists of two methods: PLS1 and PLS2 (Brereton, 2003). The difference is that PLS1 only performs regression on all of the X-data with ONE Y-column (analyte at varying concentrations) whilst PLS2 performs regression on all of the X-data with the entire Y-matrix (ALL of the analytes at varying concentrations). It is quite common to find that predictions via the PLS1 algorithm are better than for PLS2 since independent models are obtained for each analyte under investigation. However, PLS2 is easier to compute when all of the variables corresponding to the analytes within the concentration matrix are modelled.

The NIPALS algorithm is listed in Appendix A1.

Statistically Inspired Modification of PLS (SIMPLS)

This was introduced by Simjen de Jong in 1993 (de Jong, 1993). It differs from NIPALS in a number of ways. Firstly, "deflated data matrices" are not built. A *deflated matrix* is the residual matrix that is calculated from the subtraction of the regressed variables on the score vector from the original matrix during a single iteration. Equation 1.22 illustrates the calculation of a deflated matrix for the first iteration which occurs in NIPALS.

$$X_1 = X_0 - [t_1(t_1^T X_0)/(t_1^T t_1)] \qquad\qquad 1.22$$

X_1 is the deflated matrix; X_0 is the original matrix; t_1 is the score vector.

Secondly, the PLS factors (latent variables) are calculated as direct linear combinations of the original variables. This is in fact analogous to PCA. Thirdly, the SIMPLS algorithm is computationally faster than NIPALS. However, if only one *Y* variable is present, then SIMPLS will be equivalent to PLS1. A difference is however observed when more *Y* variables are present, and thus SIMPLS will differ from PLS2. Compared to NIPALS, the covariance criterion in SIMPLS is fully maximised.

The proposed modification was to by-pass the calculation of the PLS-weights, W, as used in Equation 1.21, and instead to calculate a different set of weights, R, which have not been subjected to any inverse calculations. Furthermore, whether X is centred or not, the resulting score matrix (T) is automatically centred. The corresponding score matrix for the Y matrix (U) can also be centred but it is not a requisite.

The SIMPLS algorithm is listed in Appendix A2.

Other PLS Methods

PLS is a powerful tool and is widely used in many fields (Otto, 1999b). There have been a number of PLS variants which have been developed and applied over the years. One example is the *Kernel PLS* method for application to matrices with more rows (>1000) than columns (Lindgren *et al.,* 1993). The dimensions are reduced by multiplying the transpose of the matrix by the original matrix (for example, X^TX) and this new square matrix is the *kernel*. NIPALS is then performed on the kernel matrix. However, in comparing the Kernel PLS with SIMPLS, the latter was still found to be computationally faster (de Jong, 1993). Other modifications have included compressing a model with *A* components into a model with only 2 components leading to easier interpretation of the data (Ergon, 2003).

CLS, MLR, PCR, NIPALS, and SIMPLS work well for linear datasets but PLS has the added advantage of being applied to non-linear datasets. PCR can also be applied to non-linear data sets (in the form of *polynomial PCR*) however this is limited by the assumption that the relationship between the response modelled and the components is a simple one. It does not take into account that "some components may be linear combinations of the original variables" (Despagne & Massart, 1998).

Other non-linear regression methods exist such as non-linear PLS (NLPLS), smooth multiple additive regression technique (SMART) and spline partial least squares (SPL-PLS) (Frank, 1995). These methods are able to perform well on non-linear data but are more computationally demanding and highly complex than the linear methods.

They also have a tendency to "overfit" if not carefully trained which is a limitation also shared by neural networks (Section 1.5.4.6).

1.5.4.5 Iterative Target Transformation Factor Analysis (ITTFA)

ITTFA can be utilised to process data obtained from "second-order bilinear instruments" such as HPLC-UV and GC-MS, as well as to calculate missing data points (Ni & Gong, 1997). It involves two steps. The first performs *singular value decomposition* (SVD) on the *X* matrix to form three new matrices: *U*, *S* and *V* where *V* is related to the *loadings*, *P* and the product *US* is related to the *scores*, *T*. *S* is in fact a diagonal matrix which contains singular values that are placed in descending order, that is $S_{11} > S_{22} > S_{33}$, etc (Martens & Naes, 2001; Wise & Gallagher, 1998b; Wise & Gallagher, 1998a). The last row in *X* contains the voltammogram of a sample containing the target analytes of unknown concentration. The last row in the concentration matrix (*Y*) initially contains a vector of ones. The second step involves calculating the pseudo-inverse of *X* as shown in Equation 1.23.

$$X^+ = V[\mathrm{diag}(1/S_{\mathrm{ij}})]U^{\mathrm{T}} \qquad\qquad \textbf{1.23}$$

X[+] has a reduced dimensionality as determined by the principal components retained implying the removal of noise. This leads to the calculation of a projection matrix, *PM* (Equation 1.24).

$$PM = X^+X \qquad\qquad \textbf{1.24}$$

The original concentration matrix (*Y*) is then subjected to *PM* to create a projected concentration matrix (*Y*proj). This new matrix replaces *Y* for the next iteration, which ceases once convergence is achieved. The concentrations of the target species in the unknown sample is obtained from the last row of the projected *Y* matrix.

1.5.4.6 Artifical Neural Networks (ANN)

A brief overview of ANNs as applied to chemometrics is given by the following references (Richards *et al.*, 2002; Cabanillas *et al.*, 2000b) (Cabanillas *et al.*, 2000a)

but a concise and detailed overview is in this article (Despagne & Massart, 1998). Neural networks (NNs) originated from the field of *artificial intelligence* (AI). They were initially created to mimic the workings of the human brain, mainly its ability to learn. Although the biological neuron has a firing frequency of ~1KHz, millions of these neurons utilised in unison created a highly efficient process. Fewer neurons are required in a computer for each one has a firing frequency >100MHz (Otto, 1999b).

These days, biological and artificial neurons are no longer analogous (Despagne & Massart, 1998). A computer executing a NN algorithm will not be an intelligent machine. The computer must still be told what to do. NNs are employed to relate a set of independent input variables (descriptors) with one or more dependent output variables. Figure 1.24 (below) illustrates the structure of an NN.

The nodes (or neurons) are connected together by weights: w'_{ij} for connecting nodes in the *input layer* to nodes in the *hidden layer*, and w''_{ij} for connecting the nodes in the hidden layer to the nodes in the *output layer*. The nodes in the hidden and output layers each perform two actions. The first is to sum the weighted input signals received from the input layer. The second is to take this sum and project it through a *transfer function* to produce an "activation". This transfer function can be of several forms such as a linear, sigmoid or a hyperbolic tangent function (Otto, 1999b).

Forward pass
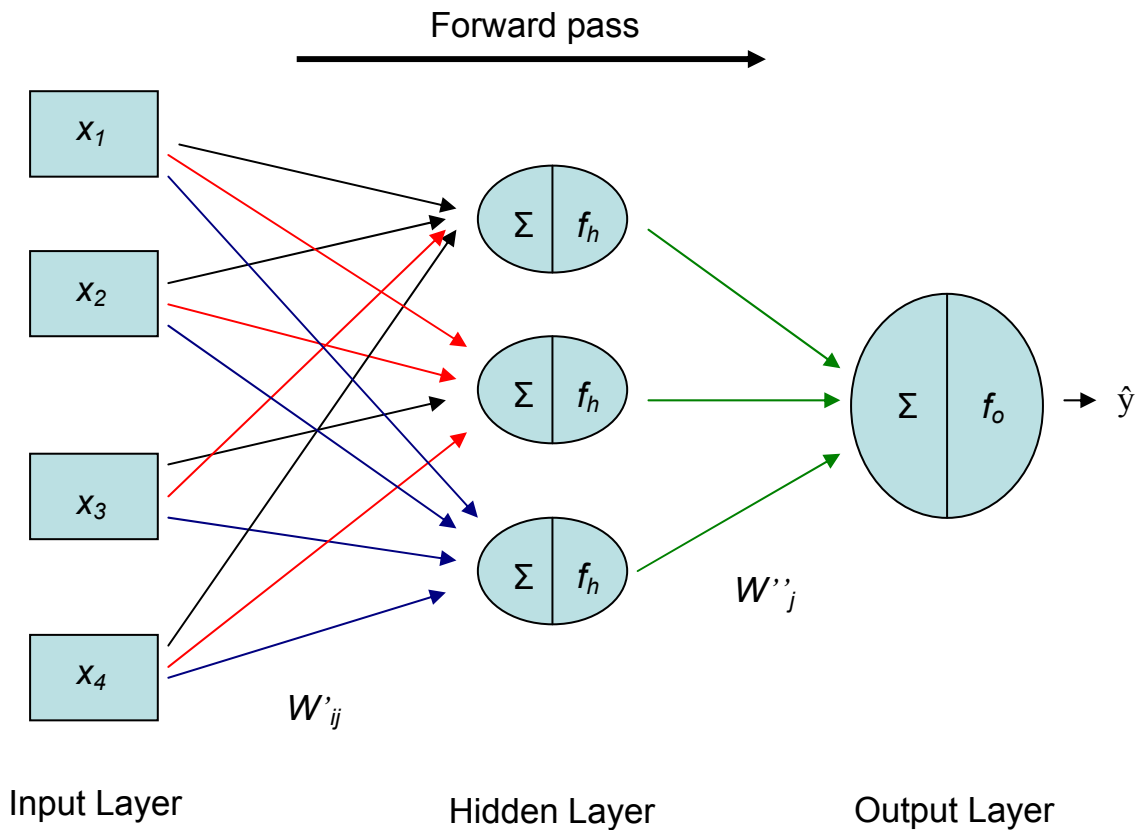
Figure 1.24: Structure of an artificial neural network (ANN)

Equation 1.25 relates the output response y_{pred} (the predicted response) with the input variables x_i.

$$y_{pred} = f_o\left[\theta'' + \sum_{j=1}^{nh} w_j'' f_h\left(\sum_{i=1}^{nd} w_{ij}' x_i + \theta'\right)\right] \qquad \textbf{1.25}$$

The transfer functions for the output and hidden nodes are denoted by f_o and f_h respectively; θ'' and θ' are *biases* which act as offset terms and are calculated during the training (learning) of the network. The weights and the biases are adjustable, and thus are assigned random values on the first iteration. Each iteration consists of two steps. The first, known as *forward pass*, is to employ training sets and predict the dependent variable which is then compared to the true value. The error between the two is used in the second step to adjust the weights, known as *error backpropagation*. The iterations are repeated until the error has reached a very low pre-specified level; this is known as *convergence*.

The advantage of this powerful tool is that it can model both linear and non-linear systems (Despagne & Massart, 1998). For linear systems, ANNs perform to the same extent as PCR or PLS, but out-performs PCR and PLS for non-linear systems. It can create an infinite number of different models due to the initial random generation of weights and biases. This can be disadvantageous in terms of optimisation partly due to the varying number of nodes that can be assigned to the hidden layer (or layers), which therefore leads to a more time-consuming procedure.

It is recommended that ANNs are only used when one is certain that the dataset is non-linear. MLR, PCR or PLS can be performed to test for non-linearity. On examination of the residuals, if a curvature or trend is observed then this is likely to indicate non-linearity. Sources of non-linearity are when signals overlap, the sample is non-homogenous, particle sizes are not constant, and stray light (when dealing with optical spectroscopy). A major advantage of ANNs over PCR and PLS is the ability to build robust non-linear models from fewer input variables in the presence of random noise (Bessant & Saini, 2000).

1.5.5 Validation of Calibration Models

Before a model can be employed to predict variables such as concentration values, it must be *validated*. This is to ensure that the model is sufficient to be able to make sensible predictions (Brereton, 2003). When the predictor variables are modelled, the most significant data will be captured within the first few PCs or LVs with noise being modelled in the latter PCs or LVs. Validating the model will thus assist in determining the maximum number of PCs or LVs required to build the final model (and thus omitting the contributions of noise) and then to employ it to make predictions. In short, the aims of these techniques are to optimise the validation model and measure the performance of the final model. Three methods of validation employed are: auto-prediction, cross validation and test-set validation.

1.5.5.1 Auto-Prediction

This is also known as the *root mean square error of calibration* (RMSEC). It is calculated on the training set. As the number of components increases, the RMSEC decreases. Figure 1.25 shows a plot of RMSEC against the number of components. This means that if there are ten components on the x-axis, then a model was built ten times, the first only having 1 component, and the last having ten.
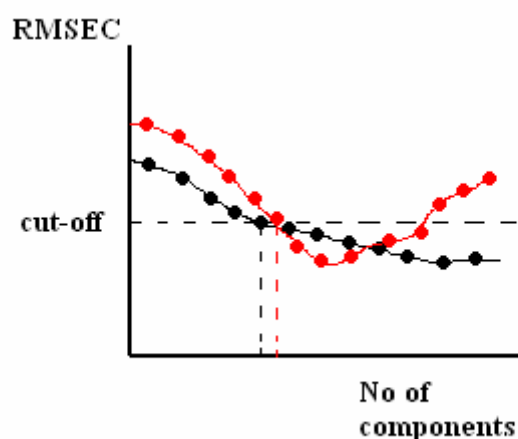
Figure 1.25: RMSEC plots.

A cut-off is normally set. The number of components chosen to build the final model is selected from when the RMSEC value reaches the cut-off (or at the component number before the cut-off). The RMSEC is calculated as:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{I}(c_i - c_i^{pred})^2}{(I - a - 1)}}$$

1.26

The above terms are defined as: c_i the true concentration for row i, c_i^{pred} the predicted concentration for row i, I the total number of rows in the training set, a the number of components in the model. The '-1' term indicates the loss of a degree of freedom due to the data having been mean-centred. This method is not normally preferred due to the data being fitted perfectly just by increasing the number of components. However,

as shown by the red curve in Figure 1.24 (above), it does assist in determining whether an error exists within the dataset due to the rise in the RMSEC value.

1.5.5.2 Cross-Validation (CV)

This is classed as an important chemometric tool (Brereton, 2003). Like auto-prediction, it too is performed on the training set, but differs from it depending on the CV method chosen, of which there are a number to select. The most common method employed is *Leave-One-Out* (LOO). Figure 1.26 graphically illustrates the method.
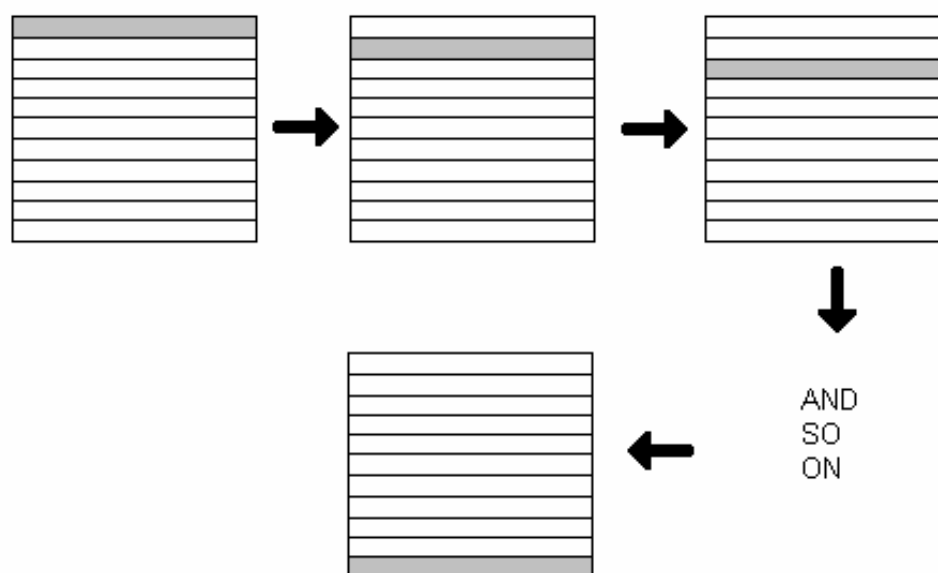
Figure 1.26: Illustration of leave-one-out cross validation (LOO-CV).

This extracts one sample (row) from both the *X* (voltammograms) and *Y* (concentration) matrices, builds a model (with *a* number of components) with the remaining samples, then uses the model (with *a* components) to predict the Y-data from the X-data originally extracted. A *predicted residual error sum of squares* (PRESS) value is calculated as follows:

$$PRESS_a = (y_i - y_i^{pred})^2 \qquad\qquad \textbf{1.27}$$

The $PRESS_a$ is simply calculated from the squared difference between the observed (true) concentration value *y* and the predicted concentration value y^{pred} of the

particular row extracted (i) at component a. This process is repeated with all of the samples (I) so that each row will have been left out and predicted once. All the $PRESS_a$ values are summed to form the *cumulative PRESS* (*CUMPRESS*) at component a as shown in Equation 1.28.

$$CUMPRESS_a = \sum_{i=1}^{I} PRESS_a \qquad\qquad \textbf{1.28}$$

The *root mean square error of cross validation* (RMSECV) is obtained via Equation 1.29 where I is the total number of rows in the training set.

$$RMSECV_a = \sqrt{\frac{CUMPRESS_a}{I}} \qquad\qquad \textbf{1.29}$$

This is repeated for every a component. It is important to stress that the dimensions of the above terms vary depending on the number of columns in the concentration matrix corresponding to the number of target analytes. As an example, if the Y matrix contained 50 concentration values for two analytes such as cadmium and lead, there would be a 50×2 matrix. If the 50 respective voltammograms were stored in X with dimensions of 50×100 and LOO cross-validation was employed in conjunction with the SIMPLS algorithm and 20 latent variables, the following dimensions for PRESS, CUMPRESS and RMSECV would vary depending on the number of analytes (columns) in the concentration matrix as illustrated in Table 1.4.

Table 1.4: Relating the dimensions of the input matrices X and Y with PRESS, CUMPRESS and RMSECV

X	Y	PRESS	CUMPRESS	RMSECV
50×100	$50 \times \mathbf{1}$	$\mathbf{50} \times 20$	$\mathbf{1} \times 20$	$\mathbf{1} \times 20$
50×100	$50 \times \mathbf{2}$	$\mathbf{100} \times 20$	$\mathbf{2} \times 20$	$\mathbf{2} \times 20$
50×100	$50 \times \mathbf{3}$	$\mathbf{150} \times 20$	$\mathbf{3} \times 20$	$\mathbf{3} \times 20$

The RMSECV is plotted as a function of the number of components. The minimum in the curve generally determines the optimum number of components (or latent variables) to utilise to build the final model. In certain situations when working with either NIPALS or SIMPLS, it has been necessary not to choose the optimum latent

variable as denoted by the minimum in the plot but to choose the previous latent variable based on the criterion "that if the difference in RMSECV value is not at least 2% for adding an extra factor" (Wise & Gallagher, 1998b). The statistical F-test can also be employed by application to the PRESS values (Haaland & Thomas, 1988a; Haaland & Thomas, 1988b). Instead of looking at the minimum PRESS value to determine the optimum latent variable with which to build the respective models, the F-ratio between the impending PRESS values and the minimum PRESS value can be determined. The first PRESS value to give an F-ratio less than 0.75 (as determined by Haaland & Thomas would thus determine the optimum latent variable. This method of choosing the optimum latent variables for the respective components is employed considerably throughout the literature (Section 5.2.1).

In practice, the model can be built with all of the components with predictions made at the optimum number of components. Figure 1.27 illustrates the plot.
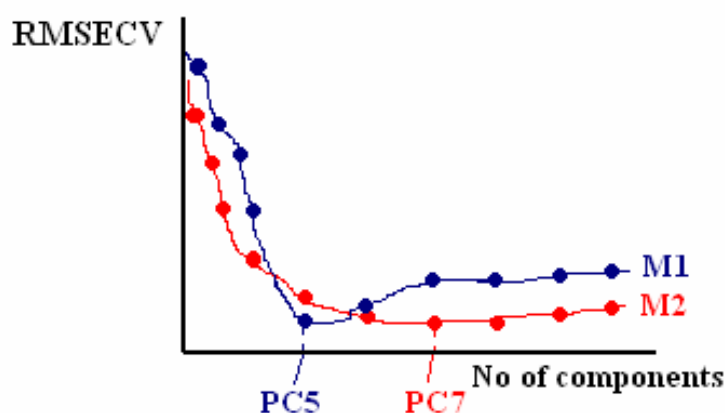
Figure 1.27: RMSECV plots.

From the above plot, the dimensions of the RMSECV matrix are 2 × 10. M1 corresponds to the first row in RMSECV and thus the first column in the concentration matrix. Predictions for both analytes would be performed at PC5 for M1 and PC7 for M2 respectively as they had attained the lowest RMSECV.

1.5.5.3 Independent Test Set Validation (ITSV)

A notable problem with CV is that it relies completely on the original dataset (Brereton, 2003). If two analytes are correlated, in that an increase in the concentration of one analyte results in the increase in the other, any model built with *a* components will not be able to predict successfully an unknown sample with a high concentration of one analyte but a lower concentration of the other. This implies a poorly built model. Although CV is able to remove the influence of internal factors such as dilution errors and instrument noise, it will not be able to account for the correlation of the analytes, which can be common in environmental samples extracted from the same source.

The models must be successful in predicting samples from unknown sources. To achieve this, two datasets are employed. The first is known as the *training* (or *calibration*) set and the second is the *test* (or validation) set. Models, as before are built with *a* components, using the training set and then validation carried out with the test set. The root mean square error of prediction (RMSEP) is calculated in Equation 1.30.

$$RMSEP_a = \sqrt{\frac{\sum_{l=1}^{L}(y_{test} - y_{test}^{pred})^2}{L}} \qquad \textbf{1.30}$$

L is the number of samples (rows) in the test set matrix. The RMSEP is plotted against the number of components (as in Figure 1.27). Whereas CV can give misleading and over-optimistic results, test set validation is more precise although it is normal to observe that test set RMSEP values are greater than CV RMSECV values. The RMSEP is also calculated after the predictions of target analytes for example in a certified reference material (CRM), to determine how well the calibration models have performed.

In industry such as in the pharmaceutical sector, test sets assist in determining if new models predict sufficiently. If not then the test set values are incorporated into the current training set forming a new training set and hence leading to the construction of

a new model which must be validated with a new independent test set. This may involve more time and money but is more economically viable than to use the current model and obtain erroneous results which could be extremely costly.

1.6 Aims and Objectives

The overall aim of this part of the ARTDEMO project described in this thesis was to develop a "real-time" automated water monitoring system, capable of using data from various complementary sources to determine the amounts of inorganic and organic pollutants. This resulted in the development of a range of *multivariate calibration* models for an electrochemical system already developed in the laboratory for the analysis of heavy metals, in particular, cadmium, lead and copper on carbon-ink screen-printed electrodes, and a novel application for the quantitative determination of arsenic and mercury on gold-ink screen-printed electrodes. Multivariate calibration models were also developed for fluorescence spectroscopy for organic compounds in particular, anthracene, phenanthrene and naphthalene.

The overall objective was to design an *analytical protocol* which will lead to the successful prediction of the amounts of pollutant compounds present, and hence the determinations of the quality of the water at the respective artificial recharge (AR) sites. Equally important is the development of software applications to accompany these tools, and to process the data that is generated from them to yield quantitative results which can be acted upon. Such applications will incorporate multivariate calibration models which will also assist in predicting, with a high degree of certainty, the respective amounts of target pollutants in the vicinity.

The main objectives of this thesis were thus:

- The determination of the overall chemometric approach (data pre-treatment, modelling, validation and prediction) leading to the successful determination of the individual concentrations of the target analytes within a given sample originating from:

- o DPASV voltammograms acquired on carbon-ink screen-printed electrodes (C-SPE)
  - o DPASV voltammograms acquired on novel gold-ink screen-printed electrodes (Au-SPE)
  - o 2D and 3D fluorescence spectra
- Data acquisition and processing on a personal digital assistant (PDA)
- A proposed automated delivery system for screen-printed electrodes

Chapters 2, 3 and 5 will discuss the chemometric approaches applied to the C-SPE voltammograms, Au-SPE voltammograms, and fluorescence spectra respectively. Chapter 4 will discuss the findings with regard to data acquired and processed on a PDA in addition to the proposed delivery system for the screen-printed electrodes. Finally, Chapter 6 will give a general discussion to the findings along with overall conclusions and suggestions for future work.

CHAPTER 2:

QUANTITATIVE DETERMINATION OF CADMIUM, LEAD AND COPPER ON CARBON-INK SCREEN-PRINTED ELECTRODES

2.1 Overview

This chapter focuses on the application of multivariate calibration to DPASV voltammograms acquired on carbon-ink screen-printed electrodes for the quantitative determination of cadmium, lead and copper. A general introduction will discuss the application of multivariate calibration to a number of different electrochemical techniques. The discussion will move on to the different types of electrodes employed in this study before focusing on the utilisation of the calibration models with screen-printed electrodes. The materials and methods employed will be detailed followed by an in-depth discussion of the results obtained which include the development of a custom-built data analysis package, development of the calibration models, and application to a real soil sample. Finally, an overall conclusion is drawn.

2.2 Introduction

Differential pulse anodic stripping voltammetry (DPASV) has been described as a highly sensitive technique (Section 1.4.2.4). The voltammograms generated closely resemble chromatograms in that one or more distinct peaks are obtained in the presence of a detectable analyte. However, there are instances in DPASV in which overlapping peaks may arise, for example due to two or more analytes exhibiting similar electrochemical properties. Therefore, the simple calculation of peak areas will not be sufficient in building a basic regression model. Examples of voltammetric peak overlap provided in the next Section are cadmium with indium, and lead with thallium. These types of issue necessitate the use of multivariate calibration techniques to account for the overlapping peaks and thus confidently predict the concentrations of a mixture of target metal ions.

2.2.1 Multivariate Calibration Applications

PLS regression has been successfully applied for the determination of thallium and lead by DPASV (Herrero & Ortiz, 1998). It had been observed that although an increase in the signal-to-background ratio is obtained, and that a discrimination against the non-Faradaic charging current ensues, the differential pulse did not correct the range of Faradaic components present in the background current. An attempt to overcome this was to subtract the background current from the recorded current. However, the background current cannot be fully compensated due to the make-up and surface area of one electrode being different from another. This is particularly apparent when working with disposable screen printed electrodes (SPEs) due to the nature of the screen-printing process, coupled to the fact that each electrode represents a unique electrode element.

The importance of being able to resolve overlapping voltammetric peaks was illustrated by Lukaszewski (Lukaszewski *et al.,* 1980). Under most electrochemical conditions, two metals can have similar oxidising potentials if their atomic structures are very similar. The authors applied a chemical approach to suppress the cadmium peak when determining the amount of thallium and lead in the presence of cadmium salts. In the case of thallium, 0.01% (w/v) of polyethylene glycol reduces the electrochemical activity of cadmium, allowing electrolysis to be performed at -0.74V (vs SCE) in 0.1M ethylenediaminetetraacetic acid (EDTA) solution, whilst the lead (and also copper) was determined in 0.1M acetic acid containing 0.1% (w/v) cetyltrimethylammonium bromide (CTAB), which shifts the cadmium peak to a much more negative value, provided that the deposition potential was in the range of -0.50 to -0.56V. The authors go on to state that "lead can be determined in the presence of ten times as much thallium". PLS was also applied for the quantitative determination of copper in the presence of iron via DPASV (Herrero & Cruz Ortiz, 1999).

To eliminate peak-shifts, the raw ASV data can be pre-treated by the addition of noise "at 10% of the maximum signal" (Donachie *et al.,* 1999). The authors showed that this approach allowed the determination of trace metals such as lead, zinc, copper and cadmium in water matrices. The data was also mean-centred. Poor results had previously been obtained when the multivariate calibration techniques were applied

directly to the raw data. The techniques compared were MLR, PCR and PLS, and the authors concluded that the PLS method gave the most accurate predictions. Results were also validated by performing measurements on ICP-MS.

A number of multivariate calibration approaches can be applied to the simultaneous determination of cadmium, copper, lead, nickel and vanadium by *differential pulse polarography* (DPP) (Ni & Jin, 1999). The approaches involved CLS, PCR, PLS and ITTFA. Differentiation of raw data was also investigated. The objective was to eliminate the residual current and the background. It was stated that the "first derivative polarograms were obtained with a $\Delta E = 4mV$". Although good results had been obtained when PLS and CLS were applied to first-derivative data, with relative prediction error values ranging from 0.5 to 1.7, there was no overall improvement in the predictive abilities of the respective models except with ITTFA (Ni & Gong, 1997; Espinosa-Mansilla *et al.,* 1992b) (Espinosa-Mansilla *et al.,* 1992a).

On performing multivariate analysis on mixtures containing cadmium, lead, indium and thallium using square wave anodic stripping voltammetry (SWASV) in an acetate-bromide electrolyte, it has been shown that the peaks produced with SWASV were narrower than with *direct current* ASV (DC-ASV) (Hassan *et al.,* 1998). However, overlap still occurs to a certain extent, especially between cadmium and indium, and between lead and thallium. In this case, inverse least squares (ILS) was sufficient to give the better separation compared with classical least squares (CLS).

Artificial neural networks (ANNs) can be employed to simultaneously determine the concentrations of cadmium and lead in a mixture (Alpizar *et al.,* 1997). Differential pulse voltammetry (DPV) coupled with a flow injection (FI) method which itself comprised of a static mercury-drop working electrode contained within the flow-through cell, had been employed. Under these experimental conditions, the voltammetric peaks pertaining to the two metals were known to interfere with each other. Two sequential pulses were applied to the three-electrode assembly. The authors explained that a hanging mercury-drop electrode inside a flow-cell will be prone to vibrations due to the "pulsating flow conditions" implying that the "geometry of the electrode" will vary and hence irreproducible results would be obtained. Their results showed that increasing the surface area of the electrode led to "poorer results"

mainly due to the "increased physical instability of the mercury drop on increasing its size under flow conditions". Since ANNs are able to model non-linear and interactive effects, this chemometric technique was employed giving better results with recoveries ranging from 78-104% for lead compared to the standard calibration methods (univariate approach) with a recovery ranging from 142-191%.

ANNs can also be used to solve the interferences caused by the formation of inter-metallic compounds such as interactions of copper and zinc when measured by ASV on a hanging mercury drop electrode (Lastres *et al.,* 1997). The authors would take each solution, make five measurements, average out the voltammograms, apply the Savitsky-Golay algorithm (2nd order polynomial, 5-point window). They concluded that ANNs can be applied to other inter-metallic compounds, and that faster convergence occurs when the least number of units are in the input layer. However, they do warn that when simplifying the "architecture" of the network, great consideration is required when selecting the amount of information needed to input into the network along with the actual signal points. This can result in a reduction of prediction errors.

A vast range of chemometric tools have been successfully applied to the determination of trace heavy metals, and demonstrated to be of considerable use, especially in environmental assays. PLS regression is the more powerful and thus employed to a greater extent than other multivariate regression techniques like PCR and CLS due to reasons given in Section 1.5.4.

2.2.2 Working Electrodes

Commonly employed electrodes are presented (Section 2.2.2.1) followed by a brief discussion on the miniaturisation of working electrodes (Section 2.2.2.2).

2.2.2.1 Commonly Employed Electrodes

Historically, the most common working electrode has been the *Hanging Drop Mercury Electrode* (HDME) (Skoog *et al.,* 1996). The drop is formed by action of a piston on the mercury reservoir forcing it through a very fine capillary tube. The

piston is driven by a micrometer screw. An example of such an electrode is illustrated in Figure 2.1.
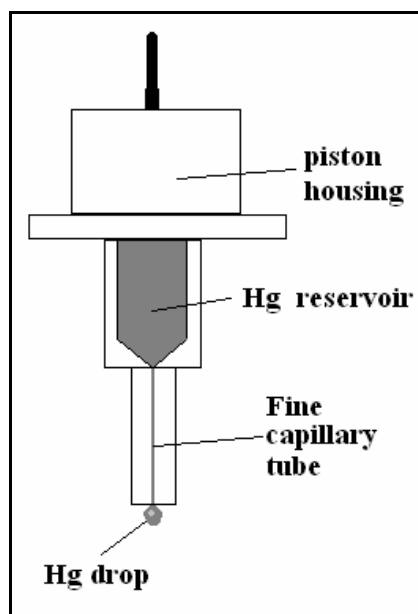
Figure 2.1: A hanging drop mercury electrode (HDME)

The HMDE has the advantage that laborious cleaning of fouling species on the working electrode in between measurements is not required, since a new mercury drop will contain a clean surface (Prado *et al.,* 2002). Although the volume of mercury employed is small in relation to the bulk solution, recent years have seen researchers undertaking the challenge to eliminate the use of mercury due to its hazardous nature to both the environment and the health of the personnel exposed to it.

An alternative to employing ANNs to calculate the concentration of inter-metallic compound mixtures was proposed. A solid electrode such as a *Boron-Doped Diamond* (BDD) solid electrode in a "mercury free" environment can be employed to determine the concentrations of lead and copper (Prado *et al.,* 2002). The BDD is quoted by the authors as being "structurally and chemically robust". It can be operated over "a wide potential window" of 3.5V in aqueous media as a result of the very low background interference. The authors reported that hydrogen evolved during SWV stripping of the lead due to the freshly exposed copper sites on the working electrode. However, the BDD electrode is not prone to surface reduction or oxidation, as are glassy carbon

electrodes and some noble metals. The BDD can be used in ASV to detect for lead, manganese, cadmium, copper and silver over a potential range of -1.0 to +1.75V (vs a SCE).

The authors also analysed the surface of the BDD electrode with *atomic force microscopy* (AFM) after firstly having measured a solution containing copper ions (A) at a deposition potential of -1.0V (vs a saturated calomel electrode (SCE)), and then a solution containing both lead and copper ions (B) at the same deposition potential. Well-defined deposits 75nm high were observed in (A) compared to numerous and smaller deposits in (B). This was due to the lead ions more readily depositing on to the copper nuclei instead of the BDD active surface, and thus the copper deposits are unable to grow as in (A). This further leads to an actual "decrease in the amount of copper detected".

The analyses of lead, cadmium and copper with PSA on a (mercury-free) *graphite-epoxy composite electrode* (G-ECE) have been reported (Serradell *et al.,* 2002). A limit of detection (LoD) for lead of 200ppb was attained at 60s deposition time, but on increasing the deposition time, a lower LoD was attained. Concurrently, a greater sensitivity was observed with the G-ECE compared with the normal glassy carbon electrodes as better defined PSA curves were obtained. The initial concept of utilising carbon-polymer composites was looked at earlier (Wang & Varughese, 1990). These materials were cheaper to attain and easier to work with than the conventional electrodes (gold, platinum and graphite). The authors here performed PSA without having to modify the surface of the composite electrode (Serradell *et al.,* 2002). They also observed that the G-ECE was more sensitive for copper (50 times that of Cd; 20 times that of Pb). The peaks for cadmium and lead were highly resolved, and thus implied that G-ECEs could be used to accurately measure the presence of both metals in the same solution. This contrasts to gold-coated SPEs where overlap was observed (Wang & Tian, 1993a).

The authors also looked at the surface of the G-ECE with *Scanning Electron Microscopy* (SEM) (Serradell *et al.,* 2002). Comparing to a glassy carbon electrode (GCE), clusters were observed on the G-ECE which implied that the surface-area of the electrode was increased, and this agreed with the increased adsorption of lead and

hence the greater sensitivity for the metal. The authors conclude that due to the ease of preparation of the G-ECEs, SPE technology can also be applied leading to the formation of disposable electrodes.

2.2.2.2 Miniaturisation of Electrodes

The utilisation of micro-electrodes in voltammetric systems can be dated as far back as the mid-1980s. An example is the construction of a micro-HPLC system coupled to a micro-flow-cell containing a silver/silver chloride reference electrode, a platinum wire counter electrode and a carbon-fibre working electrode for the determination of *catecholamines* such as adrenaline and dopamine (Goto & Shimada, 1986). They were able to construct three-dimensional voltammo-chromatograms clearly illustrating the separation of the compounds (not shown). The rapid potentiometric technique served as an aid to identifying each chromatographic peak.

Efforts have been made to miniaturise solid-state reference electrodes (Desmond *et al.,* 1997). This has been seen as an important step if miniature systems were to be designed seeing as the use of internal liquid electrolytes were impractical. They compared a normal screen-printed silver/silver chloride reference electrode, which they termed "Type I" with a modified silver/silver chloride electrode covered with a chloride-doped vinyl ester resin (termed "Type II"), the idea being to see if the concentration of chloride ions in the analyte solution could be reduced. However, the authors found that the "Type I" electrode gave more stable and consistent results compared to the "Type II" electrode. The addition of non-polar polystyrene or divinylbenzene particles to "enhance the adsorption properties of the surface" of the working electrode had been attempted when analysing for *nifuroxazide* with AdSV, but without success, warranted by an increase in background noise (Buchberger *et al.,* 1998).

2.2.3 Screen-Printed Electrodes (SPEs)

The conception, physical attributes and application of screen-printed electrodes is described (Section 2.2.3.1) followed by further enhancement by modification of the working electrode (Section 2.2.3.2).

2.2.3.1 Conception

A great advance which has been seen in recent years is the advent and application of disposable screen-printed electrodes. The advent of these electrodes coupled with their use in voltammetric techniques, such as DPASV is seen as revolutionary (Palchetti *et al.,* 1999). The convention to move away from the use of bulky electrodes and the need to de-aerate solutions was of paramount importance if portable devices were to be designed for on-site analysis. The great advantage of these electrodes is that they are easy to employ, cheap to mass-manufacture and hence can be disposed of after a single measurement implying that the laborious cleaning step akin to common solid electrodes is eliminated. This further implies that there is no risk of electrode-fouling by impurities coupled with "the memory effect" not affecting the next measurement. Memory effects are caused by the residual presence of one or more target analyte samples on the working electrode. Furthermore, as small analyte volumes (~100 µl) are measured, the need for de-aeration (removal of dissolved oxygen) is eliminated due to the size of the working electrode being much smaller than conventional solid state carbon electrodes, coupled with the reduction of capacitive charging currents and increase in the mass transport of analyte ions to and from the electrode (Skoog *et al.,* 1996).

As far back as 1988, SPEs have been employed in the environment and clinical sectors with over 50% in the "field of electrochemical sensors" (Palchetti *et al.,* 2001). They originated from the desire to miniaturise apparatus coupled to the need to design integrated circuits where the entire measurement system could be fabricated on a microchip. This could be achieved by screen-printing (Desmond *et al.,* 1996). An important application has been in the detection of glucose levels in blood (Wang & Chen, 1994). Figure 2.2 below shows the basic construction of an SPE.
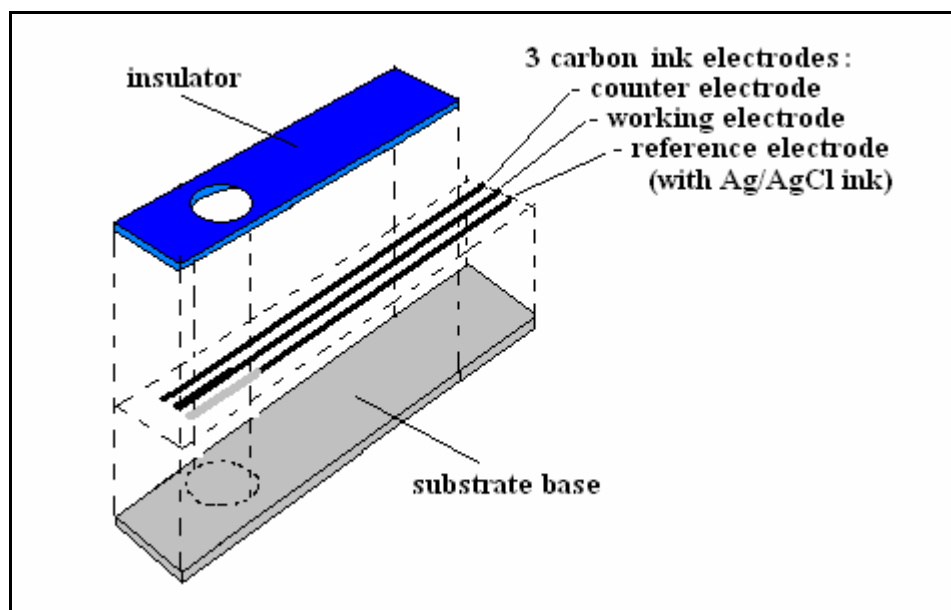
Figure 2.2: Construction of a screen-printed electrode

The graphite-carbon inks are deposited on to a polyester, polycarbonate, or an acetate substrate sheet. As shown in Figure 2.2, a silver/silver chloride ink is coated on to one of the electrodes. An insulating layer is deposited with an opening (the working area) that exposes the three terminals for contact with the analytical solution. The exposed carbon strips at the other end are inserted into an appropriately designed connector which itself is connected to the voltammetric instrument. The planar working surface area of the working electrode is 0.14cm^2 (0.7 × 0.2 cm) (Cooper, 2004).

It is important to stress that the silver-silver chloride (Ag/AgCl) reference electrode contains no internal electrolyte and thus the analyte solution should contain chloride ions, hence stipulating the need for a supporting electrolyte such as KCl (Palchetti *et al.,* 2001). It has been found that on analysis of the variation of potential of a silver-silver chloride electrode at different chloride concentrations, the potentials that had been obtained (vs SCE) were as high as 206.9mV for the lowest concentration (1 × 10^{-7} M) and -9.8mV for the highest concentration (1M) studied (Desmond *et al.,* 1997). This stipulates the need to work at a fixed chloride ion concentration in order for a steady reference potential to be generated.

Although accurate and repeatable results can be attained with SPEs incorporating mercury plating of the working electrode during the pre-concentration step, research

has been carried out to replace mercury due to its toxicity (Palchetti *et al.,* 2001). Working electrodes have been fabricated from gold inks which have thus led to the analysis of mercury itself in the environment by PSA (Wang & Tian, 1993b). The surface of the working electrode can further be modified with ion-exchangers and ligands. For example, dithizone mixed with carbon ink can be used to analyse for trace levels of lead in the absence of mercury (Palchetti *et al.,* 2001). Section 3.2 discusses further applications employed for the determination of target metal ions in the absence of mercury.

The overall performance of the SPE will be affected by the actual formation of the mercury film, the surface characteristics of the carbon working electrode, and of course the composition of the ink. The latter information can be difficult to acquire due to proprietary issues. For example, information such as the type of graphite used to manufacture the ink, the extent of any metal contaminants present in the graphite, size and porosity of the particles would be of great use (Kroger & Turner, 1997).

2.2.3.2 Modified Screen-Printed Electrodes

Screen-printed carbon electrodes (SPCEs) have also been modified with a *calixarene* in order to determine the trace concentration of lead in water (Honeychurch *et al.,* 2001). Due to the "cavity-shaped architecture" of calixarenes, they are deemed *ionophores*. The varying size of the cavity makes the calixarene selective for certain metallic ions. They are easily synthesised from the condensation of *para*-alkylphenol and formaldehyde under alkaline conditions (Gutsche & Muthukrishnan, 1978). The working electrode was modified with a range of calixarenes by simply drop coating a solution of calixarene in dichloromethane solution on to the surface of the SPCE and left to evaporate at room temperature. The thio-substituted calixarene performed much better. The authors were able to obtain a mean recovery of lead ions at 95.9% with a coefficient of variance of 8.4% (n = 5). With these encouraging results, the authors were keen to investigate the use of other calixarenes for the analyses of other metal ions.

2.3 Materials and Methods

Background information on cadmium, lead and copper is given prior to the details of the materials and methods employed. Preparation of standards in the range of 0 to 200 ppb in increments of 50 ppb as employed by Cooper is described (Cooper, 2004). Details of the acquisition of the three target analytes on carbon-ink screen-printed electrodes are given. A full factorial design was selected to ensure every combination was measured. Full details are given in Section 2.3.2. Finally, the development of a custom-built data analysis package for the multivariate calibration of the acquired differential pulse anodic stripping voltammograms is described.

2.3.1 The Target Metal Analytes

The physical attributes of cadmium, lead and copper are presented in Sections 2.3.1.1, 2.3.1.2 and 2.3.1.3 respectively.

2.3.1.1 Cadmium, Cd

Discovered in 1817 by F. Stronmeyer when he noticed that a sample of *cadmia* (now known as *calamine*) which was used in a neighbouring smelting works was yellow (CdS) instead of white (Greenwood & Earnshaw, 1990). It was shown that iron was absent thus a new element had been discovered. It has an average concentration of 0.16ppm in the Earth's crust, which is relatively low compared to zinc (76 ppm) and rubidium (78 ppm). A major use is in batteries such as rechargeable nickel-cadmium batteries. Cadmium has no positive biological role and can displace zinc in enzymes leading to storage in the bones, kidneys and liver leading to failure after around 30 years (O'Neil, 1993).

Cadmium is more readily adsorbed in to the body via the lungs than the gastrointestinal tract (Nordberg *et al.,* 1985). It can bind to sulphydryl groups on proteins and non-proteins, in addition to macromolecules such as albumin and metallothionein (Goyer, 1991; ATSDR (Agency for Toxic Substances and Disease Registry), 1989). This leads to wide distribution throughout the body via the red blood

cells leading to other heavy metals being more easily transferred into the cells via displacement. Cadmium tends to be excreted from the body via urine.

Cadmium also has a low boiling point so there are problems with atmospheric emissions. It is very mobile in the environment as it occurs in free hydrated ions thus labile in soils and sediments hence can be taken up by plants and enters the food chain. Normal intake of cadmium via food is ~0.21-0.42mg/wk. Doses of 20-30mg/kg can be fatal to humans (ATSDR (Agency for Toxic Substances and Disease Registry), 1989). Cadmium also enters the environment via pesticides and the production of super-phosphate fertilizers.

2.3.1.2 Lead, Pb

Lead has an average concentration of 13 ppm in the crust (Greenwood & Earnshaw, 1990). It is mostly present in rocks as silicate structures (O'Neil, 1993). It is not as toxic as mercury or cadmium but is more widespread. The two major uses are in lead-acid batteries and petrol additives thus leading to emission into the atmosphere. Lead has been utilised for over two thousand years. The ancient Romans used it to line their aqueducts and drinking vessels; they also added lead salts to their wine to make them taste sweeter. High levels of lead had been found in skeletons which were 100 times higher than in modern man. This is believed to be the cause of madness in emperors.

The maximum intake of lead should not exceed 430µg/day (Ferguson, 1990). The average is 150µg/day which originates from food and water. The World Health Organisation (WHO) standard is 50ppb. Lead also gets into the body by direct inhalation of air into the lungs where 40% of the lead is actually absorbed. Pb^{2+} ions in water from old lead pipes is a problem especially in soft water areas for the water is slightly acidic and since there are no Ca^{2+} ions there is thus a greater amount of Pb^{2+} ions in the water.

Lead is not very mobile in soils for it is held strongly by organic matter, and thus uptake by crops is low (Greenwood & Earnshaw, 1990). However, if fertilisers containing sewage sludge have been employed, this will lead to an increase in the lead-content of the soil. Lead has been found to be highly concentrated in the upper

layers of the soil due to the decomposition of vegetation in the humus layer coupled with the presence of lead in the atmosphere. 95% of the lead transferred to the oceans has arisen from suspended sedimentation in the rivers.

Lead has no positive biological role. *Biomethylation* does not occur extensively as for mercury therefore lead remains predominantly inorganic. It is still a neurotoxin leading to brain damage. The intelligence of lead workers had been found to be substantially decreased due to blood levels in excess of 80µg/dL (Stollery *et al.,* 1991). Lead can replace calcium in bones. High levels of lead have been found in children by analysis of their teeth due to them playing by roadsides. It can bind to haemoglobin and enzymes in the body (Bowen, 1979). It can therefore affect the gastrointestinal tract, immune, cardiovascular and central nervous systems (ATSDR (Agency for Toxic Substances and Disease Registry), 1993). Lead also enters the body via tap water pumped through lead pipes. However, this is not so apparent since these have now been replaced by copper pipes.

2.3.1.3 Copper, Cu

Copper has been used as a commodity since 5000BC. In the Middle East, ~3500BC, it was obtained by charcoal reduction of its ores then around 3000BC, tin was added to form bronze (Greenwood & Earnshaw, 1990). The average concentration is 68 ppm in the Earth's crust. It is found mainly as sulphides, oxides, and carbonates, the major ore being copper pyrite (chalcopyrite, $CuFeS_2$). It is used in pipes and electrical wiring due to its high electrical conductivity. The compounds are used in industrial chemicals, pigments, herbicides, and fungicides, due to it having anti-bacterial activity.

If exposed to high levels of inorganic copper, irritation to the eyes and skin results and if ingested, cramps, coma, and even death results. Human adults contain on average 100mg of Cu which is below the iron and zinc amounts. Daily intakes are around 3-5mg. Copper deficiency results in anaemia. However, the inability to excrete copper results in Wilson's disease which further leads to cirrhosis of the liver, neurological problems and cataracts (Goyer, 1991). Liver toxicity had also been reported as a result from long-term exposure to drinking water containing a copper concentration up to

7.8 mg/L (Mueller-Hoecker *et al.,* 1988). Copper is also found in the upper layers of soils and is believed to be due to the large amount of organic matter such as *humates* which bind tightly to heavy metals (Bowen, 1979).

2.3.2 Materials

Reagents

Double-distilled water purified by reverse osmosis (RO) was employed throughout. Stock standard solutions were prepared in the following manner:

- 10 ppm stock standard solutions of the heavy metals (Cd, Pb and Cu) were prepared by dilution of the respective 1000 ppm commercial stock standard solutions (BDH, UK) with the RO distilled water. A final volume of 10ml was attained, which also included the addition of 0.1ml of 65% nitric acid (Fluka, UK).

- A 4% w/v solution of mercuric (II) nitrate was prepared by dissolving 2g of the salt (Fluka, UK) in RO distilled water. 0.1ml of 65% nitric acid was added, and then diluted to 50ml with the RO water. This solution was further diluted to a 1% v/v solution in a final volume of 10ml.

- A 2.5M solution of potassium chloride (KCl) was prepared by dissolving 9.36g of the salt (BDH, UK) in RO distilled water to a final volume of 50ml.

- Samples were prepared in different concentrations of cadmium, lead and copper (0, 50, 100, 150 and 200 ppb) in 200 ppm mercuric (II) nitrate, 1% nitric acid, and 0.1M KCl via the aid of a full factorial experimental design generated by an in-house program developed in LabVIEW (National Instruments, USA). The final volume of each sample was 5ml. Figure 2.3 displays the design. The design was also randomised.
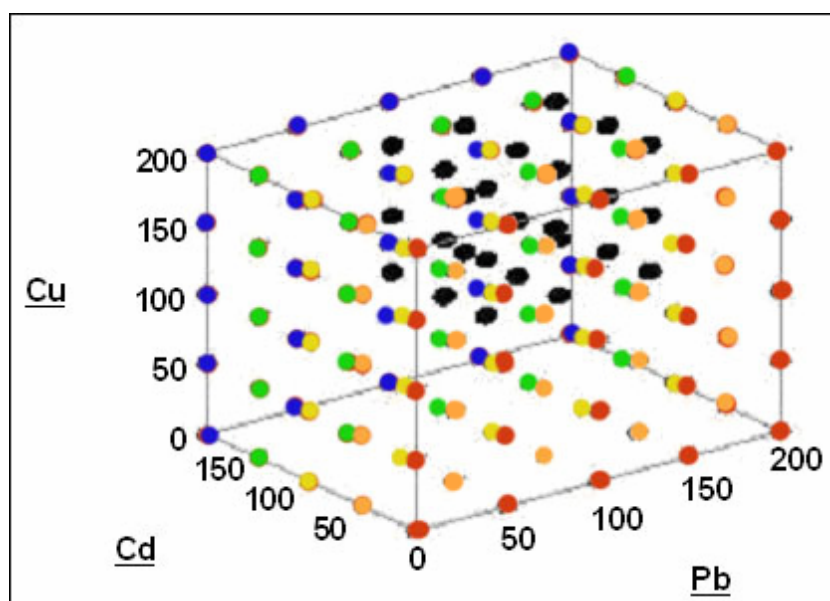
Figure 2.3: Overall experimental design: Training set (coloured with respect to Cd concentration) consists of five levels and three factors resulting in the preparation of 125 samples containing cadmium (Cd), lead (Pb) and copper (Cu) at concentrations ranging from 0, 50, 100, 150 and 200 ppb; Validation set (●) consists of three levels and three factors resulting in the preparation of 27 samples containing Cd, Pb and Cu at concentrations ranging from 90, 130 and 180 ppb

In addition to the above, a set of "unknown" samples were prepared. These were excluded from any training and validation procedures and were employed to determine the robustness of the models generated. Table 2.1 displays the concentrations of the three target metal in each of the three samples.

Table 2.1: Concentration values for the three "unknown" samples

Sample	Cd	Pb	Cu
UNK1	200	80	160
UNK2	100	20	0
UNK3	0	140	180

Screen-Printed Electrodes (SPE)

The *carbon SPEs* were printed in-house over a period of four days. The first day involves printing the carbon tracks. The second involves adding silver/silver chloride tracks. The third involves printing of the insulation layer and the last involves curing the electrodes in an oven for two hours at 120°C. The carbon-based ink (*Electrodag 423 SS*) was supplied by Acheson Colloids, Plymouth, UK. The silver-silver chloride ink (15% (w/w) AgCl was supplied by MCA Services (Cambridge, UK). The

insulating ink (in blue) was an epoxy-based polymer in the form of a resin (*242-SB*) supplied by Agmet (ESL, Reading, UK). The clear polyester sheets were supplied by Cadillac Plastics Ltd., Swindon, UK (*ST725*) with a thickness of 250 microns. An automated DEK 248 screen-printing machine (DEK Ltd., Weymouth, UK) was employed to generate the SPEs.

2.3.3 Methods

Electrochemical Measurements

For each randomised sample, measurements were made on three different carbon-ink screen-printed electrodes utilising an Autolab PSTAT10 potentiostat (Eco Chemie, Holland). The instrumental conditions employed which were optimised by Cooper ((Cooper, 2004)) are detailed in Table 2.2.

Table 2.2: The experimental parameters employed as determined by Cooper.

Parameter	Value
Electrochemical method	DPASV
Deposition potential (V)	-1.1
Deposition time (s)	165
Initial potential (V)	-1.1
Final potential (V)	-0.2
Step potential (V)	0.00488
Modulation amplitude (V)	0.05
Modulation time (s)	0.05
Interval time (s)	0.5
Scan rate (mV/s)	10

Measurements were performed at ambient temperature (20 – 25°C). For each measurement, 100µl of test solution was placed on to a fresh carbon screen-printed electrode (C-SPE) in succession, and ensuring that the electrodes were completely covered by the drop (as illustrated in Figure 2.4). The electrode was disposed of after each measurement as would be the case in a real analytical scenario.

Unless otherwise stated, measurements were performed in triplicate (n = 3) throughout. All quoted potentials are relative to Ag/AgCl reference electrodes (+0.197V vs SHE) unless otherwise stated.
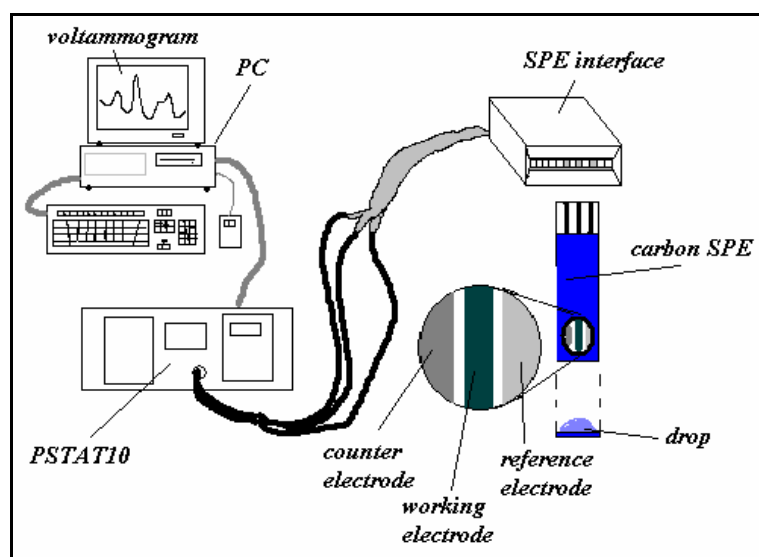
Figure 2.4: Experimental setup employing a carbon-ink screen-printed electrode

Software Applications for Data Analysis

The differential pulse anodic stripping voltammograms were acquired by the *General Purpose Electrochemical Software* (GPES) program (version 4.9, Eco Chemie, Holland). The program is also able to process the data such as to calculate the voltammetric peak areas and heights. It can also perform smoothing operations but only to the currently displayed voltammogram.

Commercial packages such as *SIMCA* (Umetrics, Sweden) and *The Unscrambler* (Camo, Norway) are dedicated data analysis packages incorporating a plethora of chemometric tools including PCA, cluster analysis, PCR and PLS. However they are limited in operation and application of specific tasks leaving the analyst with the choice of purchasing a specific plug-in or new commercial package. With the advent of programming applications such as R, Matlab and LabVIEW, the analyst is able to develop tailor-made packages to suit his/her requirements. This is particularly true for LabVIEW which permits the development of custom-built applications for personal digital assistants (PDAs) leading to data acquisition in the field (Section 4.4).

A custom-built data analysis package was constructed in both Matlab (version 6.5, Mathworks Inc., USA) and LabVIEW (version 6.1, National Instruments, Austin, TX, USA) on an Athlon 1600+ XP (AMD, USA) personal computer with 512MB RAM running Windows XP SP2 (Microsoft, USA). The package was designed to import the

acquired voltammograms, apply a number of pre-treatment techniques discussed in Section 1.5.3 such as mean-centring, range-scaling and filtering. A range of multivariate calibration regression modelling algorithms was incorporated including principal component regression (PCR), non-linear iterative partial least squares (NIPALS), statistically inspired modification of PLS (SIMPLS) and artificial neural networks (ANN). Leave-one-out cross-validation (LOO-CV) was also implemented in addition to independent test set validation (ITSV). Any generated models were saved to dedicated files including any pre-treatment parameters. This gave the added option of importing previously created models prior to predicting the concentrations of the target species. Results files were exported in text format for easy importation into dedicated spreadsheet packages such as Microsoft Excel.

The operations of both main programs including some of the main sub-programs are detailed in the form of flow charts in Appendix A3.

The *Laboratory Virtual Instrument Engineering Workbench* (LabVIEW) application is a graphical user interface that permits the user to construct user-defined programs. It consists of a *Front Panel* and a corresponding *Block Diagram* window. Digital displays, switches, buttons, and graphs are placed on the Front Panel, whilst connections between the corresponding nodes within the Block Diagram are established, either directly, or via a number of mathematical and logical functions. Matlab is a powerful application containing numerous toolboxes including *statistics*, *fuzzy logic* and *image processing* (Mathworks, 2002). Many user-defined functions have been created during the course of this project and stored in a dedicated toolbox. Additional functions were supplied by the *PLS_Toolbox 2.0* (Eigenvector Research Inc., USA). Further calculations and plots were performed on *Excel* for *Office XP* (Microsoft, USA).

2.3.4 Soil Sample

The soil sample employed originated from a certified reference material (CRM) supplied by RTC (Laramie, USA). Table 2.3 lists the amounts of the target ions present in the CRM.

Table 2.3: Amounts of target metal ions in the CRM (CRM026-050: LO# LG026, RTC, Laramie, US)

Target Metal Ion	Amount (mg/kg)
Cadmium (Cd)	11.7
Copper (Cu)	18.8
Lead (Pb)	25.6

The extraction procedure detailed below was modified from that employed by Cooper in which sonication was applied for a total of 3 minutes (Cooper, 2004). Prior to the extraction, two separate 1M solutions of nitric acid and hydrochloric acid (both of analytical grade) were prepared in HPLC water.

Extraction

- 0.93g of CRM weighed into a plastic centrifuge tube (50ml)

- 2ml of 1M nitric acid (HNO_3) added

- Sonication for 1 minute

- 6ml of 1M hydrochloric acid (HCl) added

- Sonication for 10 minutes

- Filtration through Whatman Paper No 42 into a clean centrifuge tube (50ml)

- Made up to 10 ml with RO water

Standard Preparation

For the detection of cadmium, lead and copper, 1ml of the solution was transferred into a 15ml plastic centrifuge tube:

- 100µl of HNO_3 added

- 100µl of 1% mercuric (II) nitrate added

- 200µl of 2.5M potassium chloride added

- Made up to 5ml with RO water so that the solution would contain

  - 1% (v/v) HNO_3

  - 200 ppm mercuric (II) nitrate

  - 0.1M potassium chloride (KCl)

- Another three samples were prepared in the same manner but with equal amounts of Cd, Pb and Cu added so that the final concentrations added were 100, 150 and 200ppb

Data Acquisition

Measurements were performed in triplicate on the Autolab PSTAT10 (Eco Chemie, Holland) employing the instrumental parameters listed in Table 2.2. The respective voltammograms were processed with in-house custom-built programs in LabVIEW and Matlab.

In order to validate the findings via DPASV and multivariate calibration regression, additional measurements were also made in triplicate on a Hitachi Z8100 Zeeman AAS with background correction employing both flame and flameless techniques (specific details given in Section 2.4.2)

Calculation of % Recovery

The method employed to calculate % recovery of the respective metals is to calculate the amount of metal in mg/kg from the predicted concentrations taking into account the dilution factors and then to divide by the true amount as stated on the information sheet supplied with the CRM sample.

2.4 Results and Discussion

Section 2.4.1 discusses the application of the data analysis package described in Section 2.3.3 for the quantitative determination of cadmium, lead and copper on carbon-ink screen-printed electrodes. Section 2.4.2 discusses the application of the package to real soil samples in the form of a certified reference material (CRM).

2.4.1 Simultaneous Quantitative Determination of Cadmium, Lead and Copper from DPASV Voltammograms Acquired on Carbon-Ink Screen-Printed Electrodes

Two batches of carbon-ink screen-printed electrodes were compared. The first batch consisted of the same electrodes as fabricated and employed by Cooper (Cooper, 2004). The second batch was fabricated in April 2004 by the author. The first batch will thus be designated C-SPE-O and the second batch C-SPE-N.

2.4.1.1 Univariate Calibration

After acquisition of the voltammograms on both batches as described in Section 2.3.3, voltammograms containing equal concentrations of cadmium, lead and copper were extracted from the respective datasets. Figure 2.5 displays the overlaid raw voltammograms for the three metal ions acquired on batch C-SPE-O (A to E) and the average voltammograms (F). The third measurement for 150 ppb Cd, Pb and Cu is not included due to the insufficient contact between the terminals of the screen-printed electrode employed and the connector. As can be seen, the magnitudes of the peak areas increase with increasing concentration. However, the peak areas for 150 ppb may be cause for concern due to their similarity with the peaks corresponding to 100 ppb (Figure 2.5F). With reference to Figure 2.5D, there is a clear difference between the two voltammograms recorded at 150ppb, especially for cadmium. This will therefore affect the calibration curve significantly. Overall, Figure 2.5 illustrates that there is significant variability between the response of the carbon-ink screen-printed electrodes.
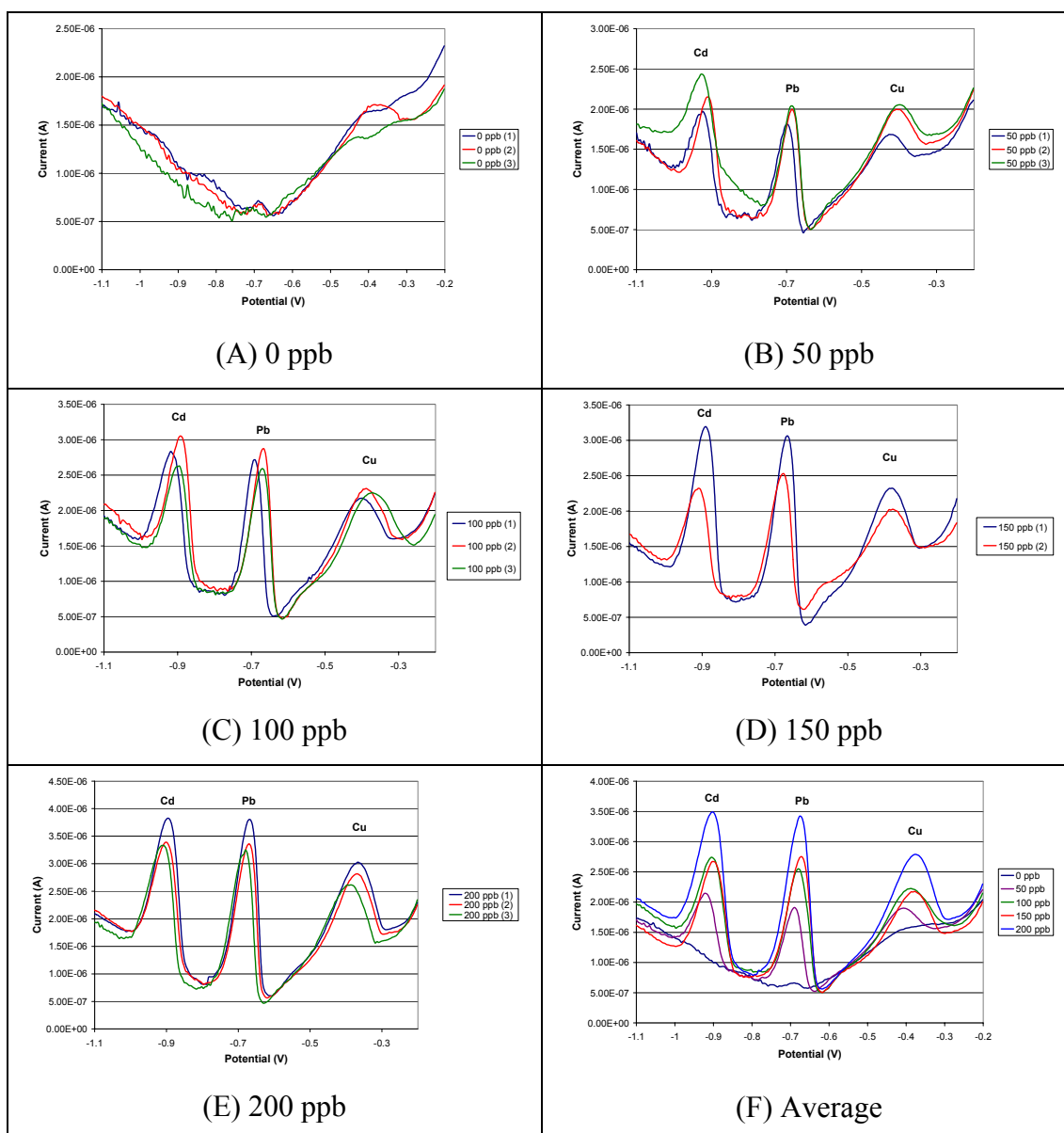
Figure 2.5: Overlaid raw voltammograms for Cd, Pb and Cu in 200 ppm mercuric (II) nitrate, 1% nitric acid, and 0.1M KCl acquired on batch C-SPE-O (A to E) and the overlaid average of the voltammograms (F). Experimental conditions as in Table 2.2.

Figure 2.6 (below) displays the calibration curves for the three metals. Large error bars have been attained for the 150 ppb samples. Lead has given the best coefficient of determination (R^2) resulting in a correlation coefficient of 0.9828. Copper and cadmium have respectively attained correlation coefficients of 0.9782 and 0.9708. The error bars also indicate that each disposable electrode has different characteristics.
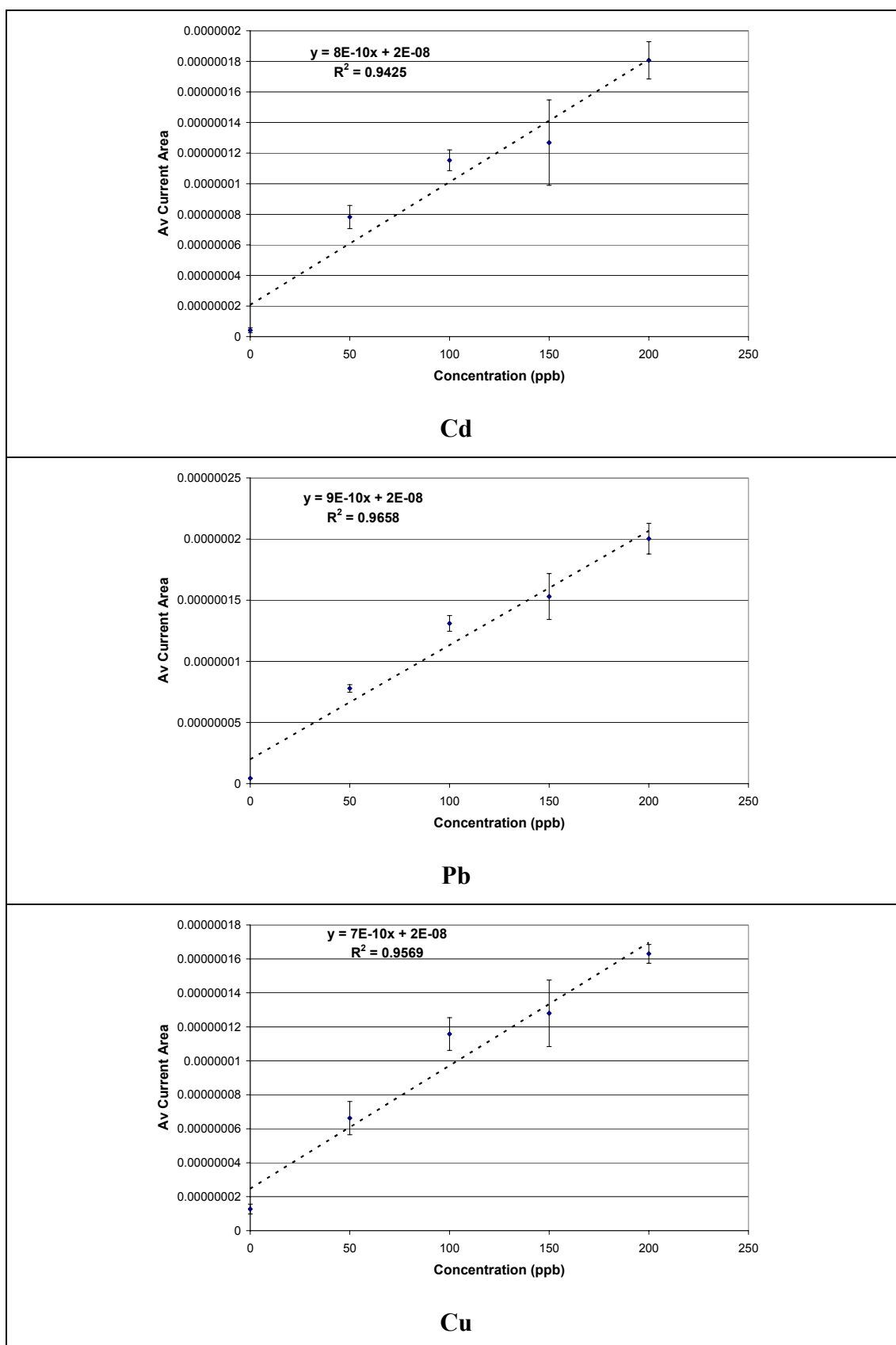
Figure 2.6: Calibration curves for cadmium (Cd), lead (Pb) and copper (Cu) for batch C-SPE-O (n=3). Reagent and experimental conditions as in Figure 2.14 and Table 2.2 respectively

Table 2.4 shows the average predicted concentrations for each metal ion in the three "unknown" samples (Table 2.1) along with the root mean square (RMS) errors calculated. The three samples were also recorded in triplicate.

Table 2.4: Average predicted concentrations of the respective metal ions (in ppb) and corresponding RMS error values for the samples recorded in triplicate on batch C-SPE-O.

Sample	Cd			Pb			Cu		
	Pred	Act	RMS	Pred	Act	RMS	Pred	Act	RMS
UNK1	137	200	66.43	88	80	9.05	144	160	26.53
UNK2	100	100	6.74	18	20	3.13	21	0	22.79
UNK3	-22	0	22.44	134	140	7.51	138	180	42.06

Lead has given the best and more precise predictions compared to the other two metal ions. The main reason is that the lead peak is in the centre of the potential window where there is more of a linear baseline. Both cadmium and copper are more affected by the possible presence of other species, such as hydrogen in the case of cadmium, and mercury in the case of copper. This infers that chemometrics in the form of multivariate calibration is required to reduce the RMS error values.

The calibration curves and average predictions for the three target metal ions for the C-SPE-N batch are displayed in Table 2.5 and Figure 2.7 respectively.

Table 2.5: Average predicted concentrations of the respective metal ions (in ppb) and corresponding RMS error values for the samples recorded in triplicate on batch C-SPE-N

Sample	Cd			Pb			Cu		
	Pred	Act	RMS	Pred	Act	RMS	Pred	Act	RMS
UNK1	232	200	71.92	84	80	11.67	50	160	123.39
UNK2	68	100	36.80	15	20	10.67	43	0	102.33
UNK3	-6	0	6.79	113	140	33.03	123	180	64.84

Cd

Pb

Cu
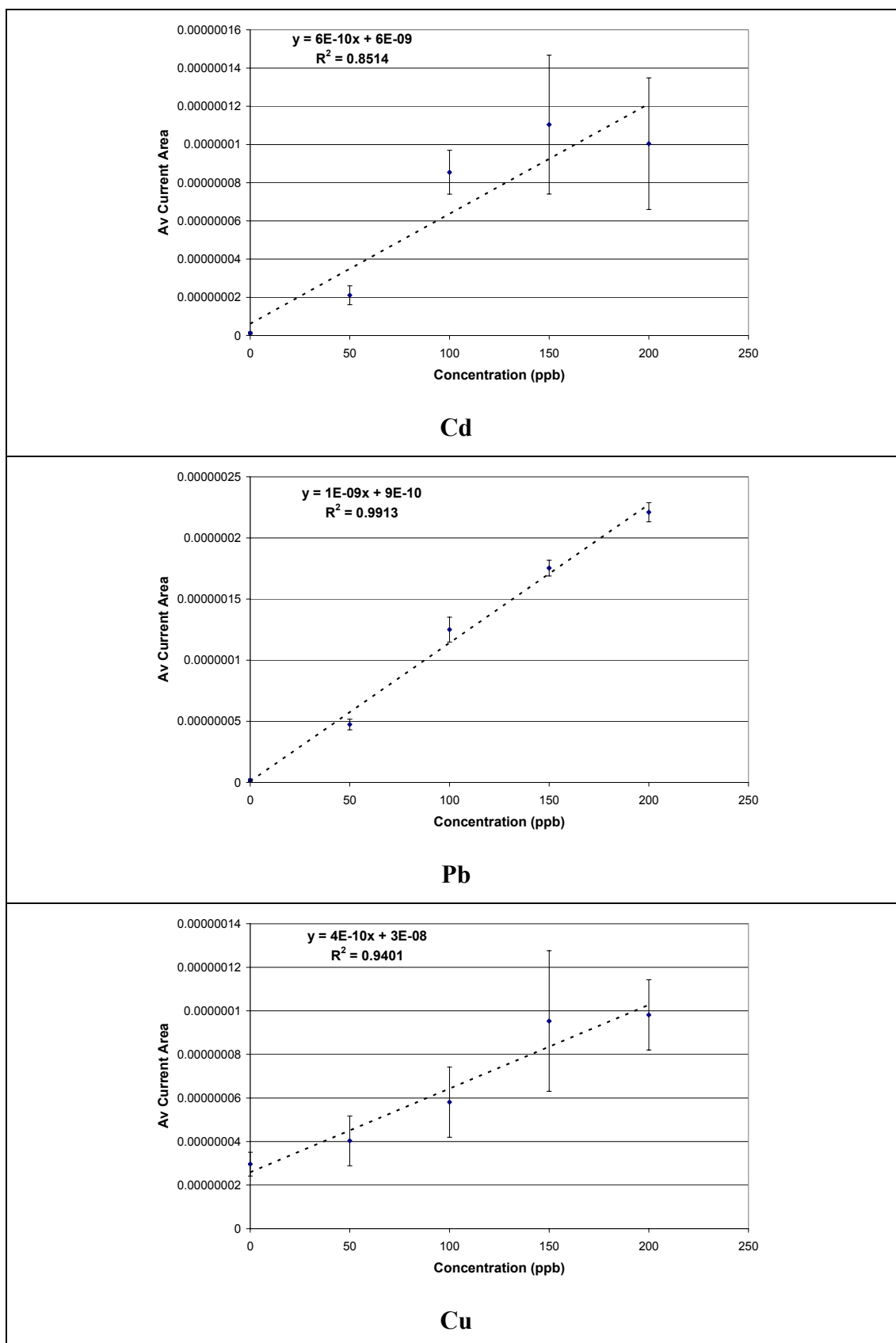
Figure 2.7: Calibration curves obtained for Cd, Pb and Cu on batch C-SPE-N (n=3). Reagent and experimental conditions as in Figure 2.5 and Table 2.2 respectively

As expected, lead (Pb) has given the better calibration curve with a correlation coefficient of 0.9956. In comparing with the C-SPE-O batch of electrodes, the correlation coefficient of lead is better for the C-SPE-N batch of electrodes whilst the opposite can be said for copper and cadmium with correlation coefficients of 0.9696 and 0.9227 respectively. The error bars for batch C-SPE-N are more pronounced than on the C-SPE-O batch. Furthermore, the slopes for cadmium and copper for the C-SPE-N batch are less steep than the C-SPE-O batch indicating a reduction in sensitivity of the new batch for the two metals.

Lead has attained good RMS values in comparison to the other two metals. The RMS values for copper are particularly poor as can be seen in the average predicted concentration values of 50, 43 and 123 ppb for UNK1 (160 ppb), UNK2 (0 ppb) and UNK3 (180 ppb) respectively. The poorer performance of the C-SPE-N batch can be attributed to the physical characteristics of the carbon-ink working electrode (see below). Figure 2.8 (below) displays the overlaid raw voltammograms for the three metals on the C-SPE-N batch (A to E) and the averaged voltammograms (F) which also displays a voltammogram acquired on one electrode from batch C-SPE-O for the same 200 ppb sample for comparison.

There is thus a significant difference in the quality of the voltammograms (Figure 2.5 and Figure 2.8). It is unclear as to why this phenomenon has occurred with the C-SPE-N batch. In particular, the cadmium peak is difficult to interpret due to the high baseline at 50 ppb (Figure 2.8B). It was however, observed that minute bubbles had formed on both the working and counter electrodes. The cause of these bubbles is currently unknown however it is not due to oxygen evolution for this would occur in excess of 1V in aqueous solutions with carbon screen-printed electrodes and a silver/silver chloride reference electrode. This thus led to the voltammograms appearing very noisy. This could imply that the carbon-ink employed is more active than that employed in the C-SPE-O batch of electrodes.

(A) 0 ppb

(B) 50 ppb

(C) 100 ppb

(D) 150 ppb

(E) 200 ppb
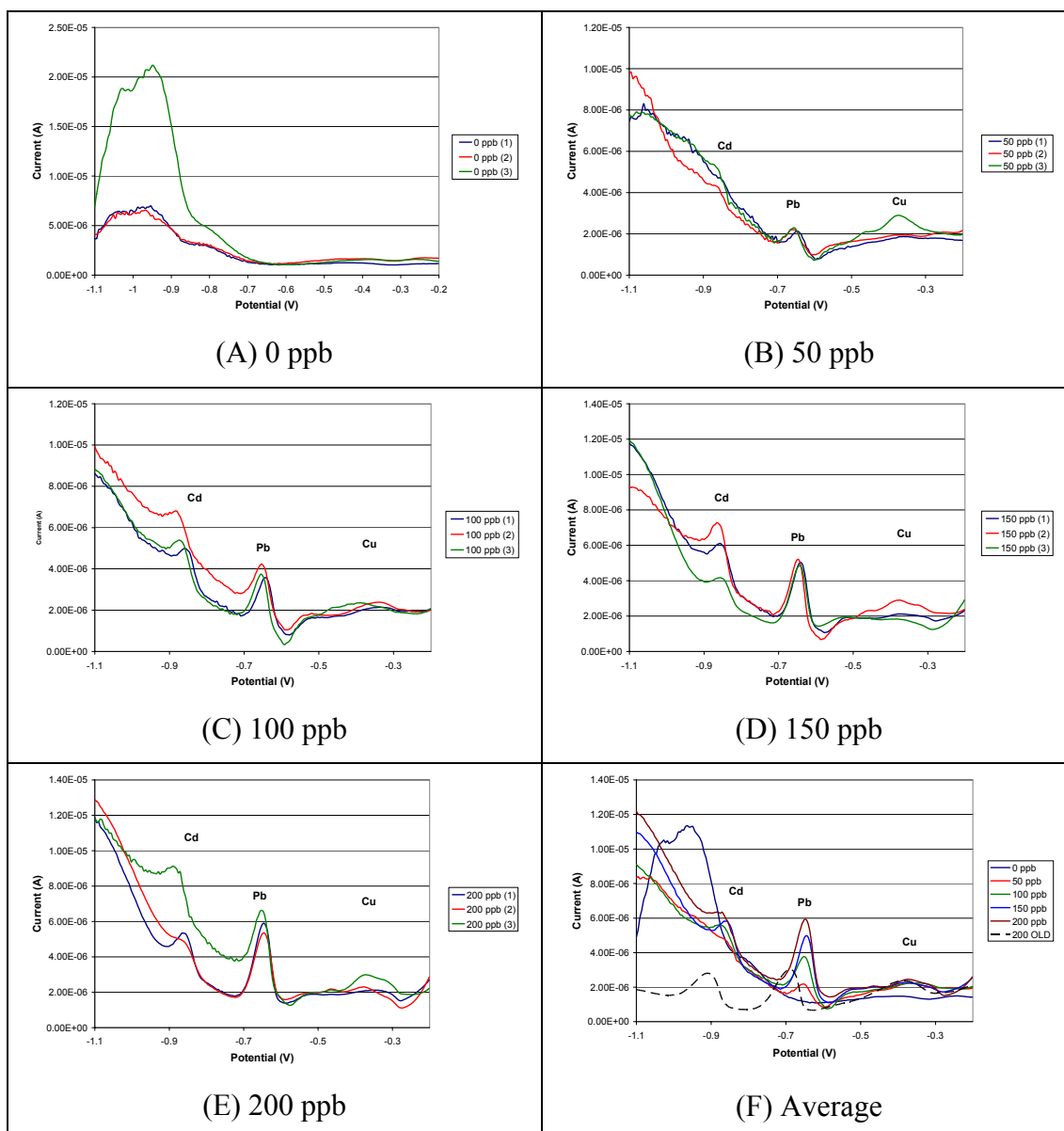
(F) Average

Figure 2.8: Overlaid raw voltammograms for Cd, Pb and Cu acquired on batch C-SPE-N (A to E) and overlaid average of the voltammograms (F). Reagent and experimental conditions as in Figure 2.5 and Table 2.2 respectively.

Another observation was that it was proved difficult to maintain the drop directly over the working area of the C-SPE-N electrode batches suggesting batch-to-batch variability with respect to the physico-chemical properties of the insulation layer ink when compared with the C-SPE-O batch of electrodes. Finally, the blue insulation ink appeared to have a much glossier finish than on the older electrodes and could be as a result of the curing regime employed.

The effect of increasing the deposition time was investigated by the acquisition of three voltammograms at concentrations ranging from 0, 100 and 200 ppb for all three metal ions on the C-SPE-N batch (Figure 2.9).
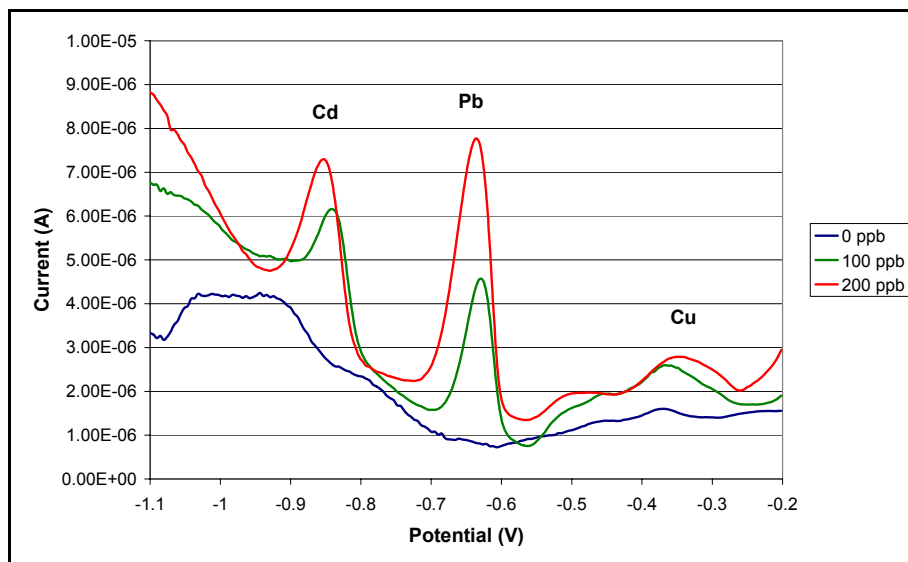
Figure 2.9: Overlaid voltammograms for Cd, Pb and Cu recorded on batch C-SPE-N. Reagent and experimental conditions as in Figure 2.5 and Table 2.2 respectively except for a deposition time of 6 minutes.

An improvement can be seen in the voltammetric peaks, especially for cadmium and lead. However, the voltammogram pertaining to the blank (0 ppb) gives rise to a high baseline, particularly in the region -1.05 to -0.9V. This indicates that there could be an issue with the supporting electrolyte solution interacting with the inks; fouling or degradation of the C-SPE-N batch of electrodes could affect the shelf-life making it considerably less than the C-SPE-O batch of electrodes, which could imply different carbon inks having been employed. A three-point calibration curve (not shown) gave a correlation coefficient of 0.9843, 0.9997, and 0.5467 for cadmium, lead and copper respectively. The poor correlation coefficient for copper can be attributed to the mercury film interfering more prohibitively with the copper due to the longer deposition time. Cooper reported detection limits of 9.1, 1.4 and 32.6 ppb for cadmium, lead and copper respectively (Cooper, 2004).

Finally, the effect of the deposition current on the calculated peak areas of the C-SPE-N batch was investigated. It had been observed that during the acquisition of certain voltammograms, the currents displayed during the deposition step determined the

nature of the acquired voltammogram. For example, if the current exhibited a high negative value, the cadmium peak was more affected by the increase in the baseline, which may be possibly due to the evolution of hydrogen at the negative depositioning potential employed. Table 2.6 displays the peak areas calculated for 200 ppb cadmium on the four C-SPEs whilst Table 2.7 displays the same for 50 ppb lead and 150 ppb cadmium and copper.

Table 2.6: Effect of magnitude of deposition current on peak area on 200 ppb Cd. Reagent and experimental conditions as in Figure 2.5 and Table 2.2 respectively.

C-SPE-N	1	2	3	4
Order (1 = highest magnitude)	3	2	4	1
Cd Peak Area	1.44×10^{-7}	1.41×10^{-7}	1.63×10^{-7}	1.32×10^{-7}

Table 2.7: Effect of magnitude of deposition current on peak areas of 50 ppb Pb, and 150 ppb Cd and Cu. Reagent and experimental conditions as in Figure 2.5 and Table 2.2 respectively.

C-SPE-N	1	2	3	4
Order (1 = highest magnitude)	1	4	3	2
Cd Peak Area	1.39×10^{-7}	1.62×10^{-7}	1.53×10^{-7}	1.42×10^{-7}
Pb Peak Area	4.21×10^{-7}	4.91×10^{-7}	4.76×10^{-7}	4.72×10^{-7}
Cu Peak Area	1.93×10^{-7}	1.57×10^{-7}	1.58×10^{-7}	1.26×10^{-7}

As can be seen in Tables 2.6 and 2.7, the greater the magnitude of the deposition current, the lower the calculated peak areas, with the exception of copper, which can be attributed to the interference of the mercury ions. An additional observation is that on acquiring the validation dataset, the standard containing 180 ppb Cd, Pb and Cu was also recorded on four C-SPE-O batches. It was observed that the magnitude of the deposition current was much lower than that observed for the C-SPE-N batch. For example, the current averaged 17 µA for all four C-SPE-O electrodes, whilst the average observed for the C-SPE-N batch was 50 µA, which thus explains why the voltammetric peaks appear more intense and well-defined with the C-SPE-O batch than the C-SPE-N batch, coupled with the lower baseline in the cadmium region.

An important point to note is that screen-printed electrode manufacturers such as in the blood glucose biosensor industry are aware of this problem and circumvent it by calibrating each batch of electrodes that is produced. The end-user is then supplied

with a calibration chip that loads compensation data into their meter (DiabeteSuffolk, 2006). This therefore implies that the problem is universal. Appendix A4 shows a table comparing a number of blood glucose meters listing operational parameters, physical construct and calibration methods.

Overall, univariate calibrations are not adequate for the datasets acquired on both the C-SPE-O and C-SPE-N batches. In both cases, predictions are poor. RMS error values indicate that batch C-SPE-O is more reliable than batch C-SPE-N.

2.4.1.2 Comparison of Multivariate Calibration Regression Algorithms

Multivariate calibration was applied via the custom-built data analysis package to the datasets containing voltammograms acquired on the C-SPE-O batch with the PCR, NIPALS, SIMPLS and ANN modelling algorithms in order to determine the most efficient when combined with pre-treatment techniques (Ni & Jin, 1999; Donachie *et al.,* 1999; Bessant & Saini, 2000). To reduce computational time a subset was extracted from the training dataset with a three-level three factorial design. Figure 2.10 compares the root mean square (RMS) error values for the prediction of replicate samples of 100 ppb of the target metal ions.

Mean-centring combined with blank subtraction in conjunction with the SIMPLS algorithm led to the most successful predictions. It is clear to see that no pre-treatment resulted in larger RMS error values. However it is recommended that voltammograms pertaining to the blank samples not be subtracted and should thus be included in the model (Herrero & Ortiz, 1998). Furthermore, mean-centring can result in weighting of spectral regions if the peak magnitudes differ substantially that will lead to erroneous modelling (Thissen *et al.,* 2004). As a result, range-scaling was adopted in conjunction with the Savitzky-Golay smoothing function, details of which including advantages have been given in Section 1.5.3 (Savitzky & Golay, 1964).

	A	M	B	RMS
PCR	✔		✔	
		✔	✔	
SIMPLS	✔		✔	
		✔	✔	
NIPALS				
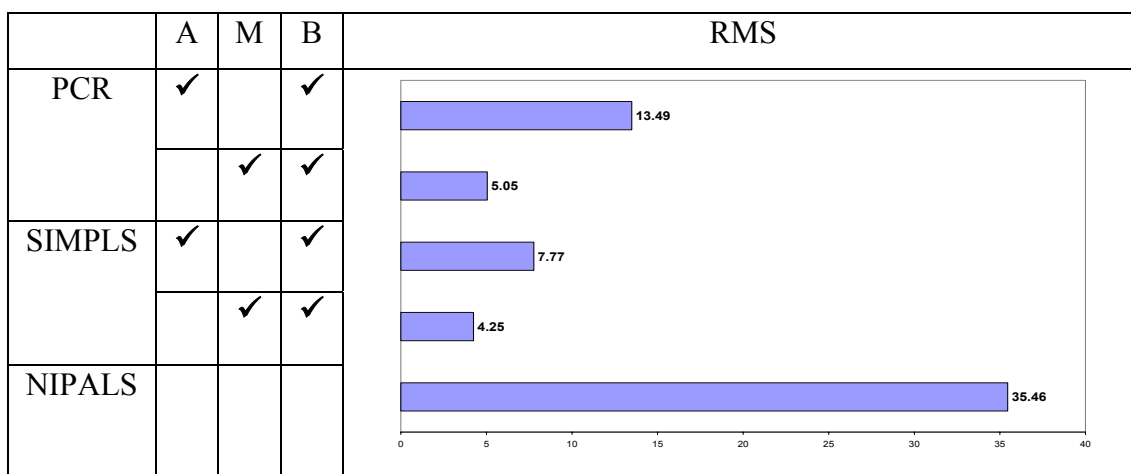
Figure 2.10: RMS Error values for the predictions of replicate samples at 100ppb (n=5) with varying pre-treatment techniques and modelling algorithms.
Key: A denotes auto-scaling; M denotes mean-centring; B denotes blank subtraction.
Experimental design: 3-level (0, 100, 200 ppb), 3-factorial

The literature reports that artificial neural networks (ANNs) are gaining in popularity as an advanced tool when being applied to voltammetric data (Richards *et al.,* 2002). Research has shown that ANNs perform better than PLS with respect to modelling and prediction due to the ability of ANNs to model both linear and non-linear data. Figure 2.11 compares the RMSEP values obtained for SIMPLS and ANNs.

SIMPLS	ANN	PCA	PCs	RMSEP
	✔	✔	50	
	✔	✔	30	
	✔	✔	20	
	✔		N/A	
✔			N/A	
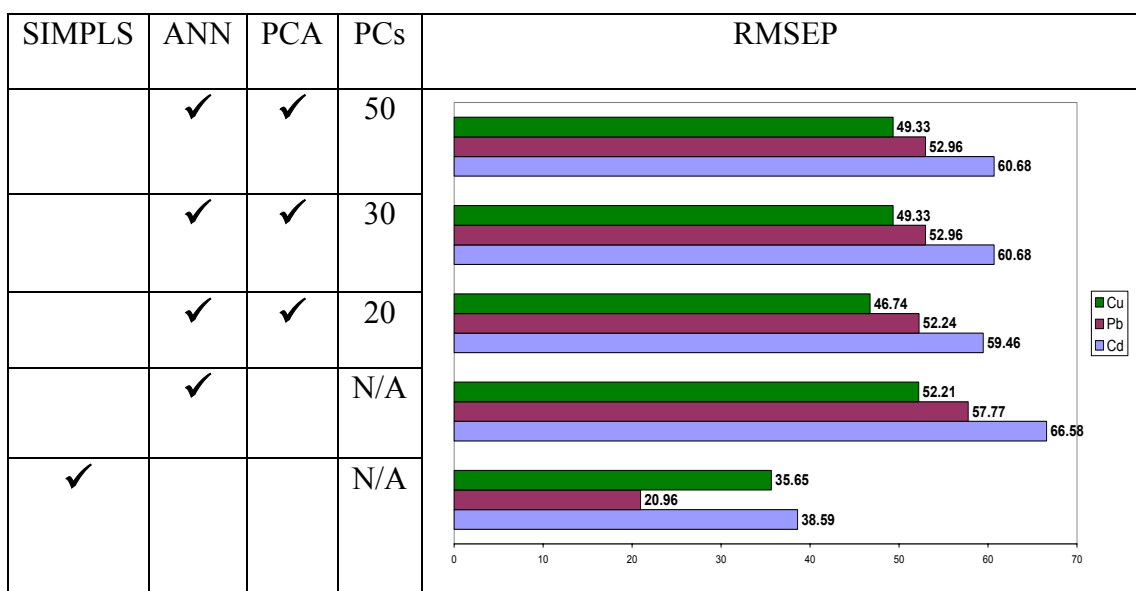
Figure 2.11: RMSEP values for Cd, Pb and Cu recorded on a carbon-ink SPE. Dataset range-scaled and smoothed with Savitzky-Golay function (cubic polynomial and window size of 41). SIMPLS and ANN are the modelling algorithms; PCA denotes data reduction prior to ANN; PCs denotes the number of PCs retained. N/A denotes not applicable

PCA can be performed on data prior to performing ANN (de Carvalho *et al.,* 2000). The scores matrix from the PCA process replaces the original X-data as input matrices for the ANN. The architecture of the neural network (Hagan *et al.,* 1996) was dependent upon the potential scan range, for example, -1.1 to -0.2V, and the number of metal ions that are present (184-17-3). The number of inputs corresponded to the number of variables in each voltammogram. There were 17 neurons in the hidden layer incorporating a sigmoidal transfer function. Three neurons (corresponding to the number of metals) are present in the output layer and a linear transfer function outputs the concentration values of the respective metal ions. The number of neurons was determined by the NNPLS1 algorithm contained with the PLS toolbox (Wise & Gallagher, 1998b). The network was optimised via a back propagation algorithm which reduces the error between the predicted and required outputs (Y_{pred} and Y_{act} respectively) by altering the weights of the links between the neurons within the network (Hagan *et al.,* 1996). As can be observed in Figure 2.11, application of the SIMPLS algorithm resulted in better predictions. Application of OSC (Fearn, 2000) prior to modelling led to an increase in the RMSEP values by an average of 10 ppb.

SIMPLS has thus been shown to be the better multivariate calibration tool for modelling and predicting and was employed for the remainder of the work. Furthermore, as neural networks are time-consuming to train, it is not guaranteed that the above networks were in fact optimised (Richards *et al.,* 2002; Bessant & Saini, 1999; Bessant & Saini, 2000).

2.4.1.3 Multivariate Calibration on Batch C-SPE-O

All predictions were performed after the incorporation of the validation dataset into the model (Martin *et al.,* 2005). In all cases, predictions were performed at the 20th latent variable for all components (Thennadil & Martin, 2005). The leave-one-out cross-validation plot for the training set is shown in Figure 2.12.
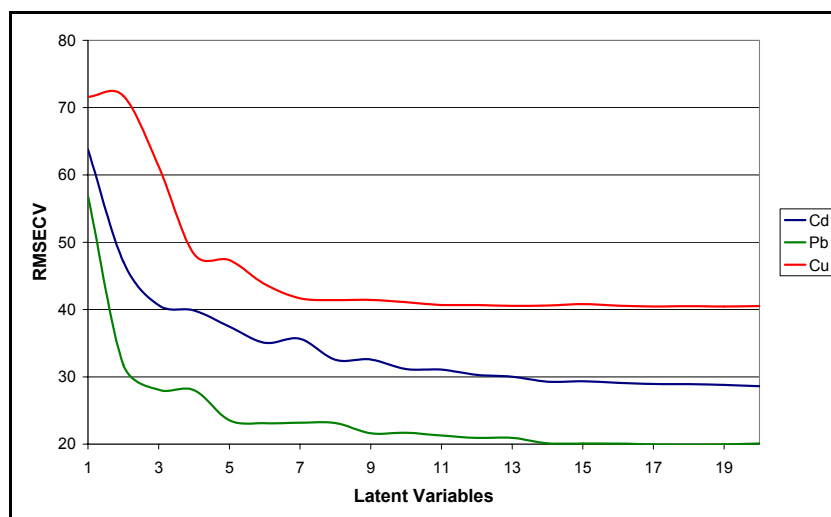
Figure 2.12: LOO CV plot for the training set acquired on the C-SPE-O batch. No data pre-treatment performed. Reagent and instrument conditions as in Figure 2.5 and Table 2.2

The LOO CV plots displayed in Figure 2.12 are far from ideal. The perturbations observed in each plot from the third latent variable to the tenth imply that problems have been encountered within the dataset. The total variance captured by the X- and Y-data set had not reached 100% at the 20th latent variable (99.9414 and 94.5893 respectively). A closer inspection of the training dataset revealed voltammograms in which clear shifts in peak maxima potentials were evident (Figure 2.13).
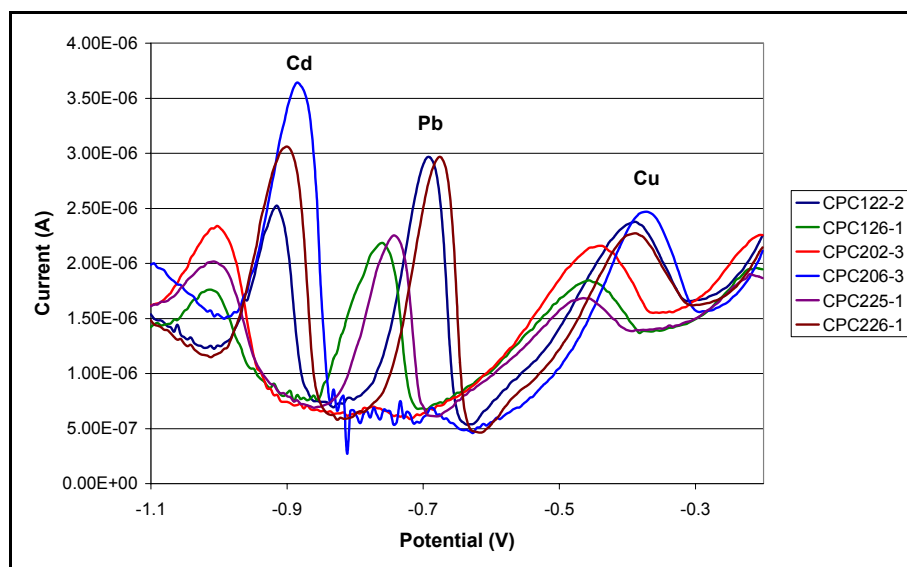
Figure 2.13: Overlaid voltammograms for a number of samples from the training set acquired on batch C-SPE-O. Reagent and experimental conditions as in Figure 2.5 and Table 2.2 respectively. Key: CPC225 implies 200 ppb Cd and Pb; 50 ppb Cu

The shifted voltammograms pose a serious problem for the modelling algorithm. One hypothesis as to the cause of this shifting is degradation of the reference electrode. In order to assist in resolving this problem, a program was constructed in the LabVIEW environment. Its role was to import the appropriate dataset, and to re-align the voltammograms to the more positive region of the voltammogram. This was based on the differences between the index values (potential) corresponding to the three maximum peak currents (corresponding in this case to Cd, Pb and Cu) of a chosen voltammogram deemed to be "normal", and the shifted voltammograms. After the re-alignment, the voltammograms would be cropped at both ends thus reducing the dimensionality to fit a potential window which itself is dependent on the maximum value of the potential shift. Elimination of peak-shifting was described in Section 2.2.1 but the above process is a simplified form akin to the peak alignment of NMR and GCMS data (Forshed *et al.,* 2005; Christensen *et al.,* 2005).

Tables 2.8 to 2.11 display the means, standard deviations and percentage coefficients of variation for the predictions of the concentrations of the synthetic samples as shown in Table 2.1, which is re-produced below for convenience.

Table 2.1: Concentration values for the three "unknown" samples

Sample	Cd	Pb	Cu
UNK1	200	80	160
UNK2	100	20	0
UNK3	0	140	180

Table 2.8: The predicted concentrations of the unknown "real" samples with no data pre-treatment followed by modelling with SIMPLS.

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	163.1	90.7	118.4	15.72	4.46	6.49	9.6	4.9	5.5
UNK2	121.5	23.0	18.6	8.26	6.55	16.30	6.8	28.4	87.8
UNK3	0.0	135.6	154.0	0.00	4.21	8.01	NaN	3.1	5.2

Table 2.9: The predicted concentrations of the unknown "real" samples with range-scaling, SG filtering (polynomial: 3; window: 41), followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	160.5	91.2	131.1	22.34	1.54	3.35	13.9	1.7	2.6
UNK2	129.7	28.0	9.6	8.59	6.20	8.66	6.6	22.1	90.0
UNK3	0.0	138.6	160.7	0.00	1.81	12.58	NaN	1.3	7.8

Table 2.10: The predicted concentrations of the unknown "real" samples with the voltammograms shifted, and no data pre-treatment followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	180.6	101.4	138.6	27.31	3.09	8.55	15.1	3.1	6.2
UNK2	123.6	22.0	35.3	4.79	6.02	31.96	3.9	27.3	90.6
UNK3	0.0	148.9	155.0	0.00	8.10	13.81	NaN	5.4	8.9

Table 2.11: The predicted concentrations of the unknown "real" samples with the voltammograms shifted, with range-scaling, SG filtering (polynomial: 3; window: 41), followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	177.1	101.9	150.5	23.60	2.21	6.17	13.3	2.2	4.1
UNK2	126.5	22.4	34.1	14.67	10.66	30.52	11.6	47.6	89.5
UNK3	0.0	146.7	155.2	0.00	8.55	17.66	NaN	5.8	11.4

It is apparent to see that the models generated have failed to predict the concentration of copper in the UNK2 sample (0 ppb). The closest model being that associated with Table 2.9 at 9.6 ppb. However its precision is still very poor as can be seen by the high %CV value. The predictive capabilities of the models have thus been compromised by the issue of peak shift in the voltammograms, and it may be the case that the virtual instrument (VI) built to shift the voltammograms will need further revising.

The data analysis package permits the application of *weights* to specific regions of the voltammogram. This results in enhancement of the respective voltammetric peaks pertaining to the target analytes and reduces the noise in the remainder of the voltammograms (Martens & Naes, 2001; Forina *et al.,* 2003). The selection of the appropriate weighting parameters is dependent on the quality of the voltammograms

acquired. For example, if a voltammogram contains a low signal-to-noise ratio, the high and low weights are established in which the ratio is increased substantially.

Prior to application of the weights, a condition is first established in which the corresponding sample in the concentration matrix (Y) is scanned, and if the concentration of a particular metal is 0 ppb, then a pre-defined low weight is applied else a pre-defined high weight is applied. Furthermore, the weighted voltammograms are smoothed with the Savitsky-Golay function at a polynomial of 3 (cubic) and a window of 41. To the author's knowledge, no such function has been described in the literature. Table 2.12 details the index ranges in which the high weights are applied to the voltammograms. Figure 2.14 illustrates the effects of weighting.

Table 2.12: The ranges in which the high weight is applied to the respective metal ion.

Metal (C-SPE)	Range: Index (potential)
Cd	21 (-1.005V) → 45 (-0.888V)
Pb	60 (-0.8143V) → 96 (-0.638V)
Cu	140 (-0.424V) → 175 (-0.253V)
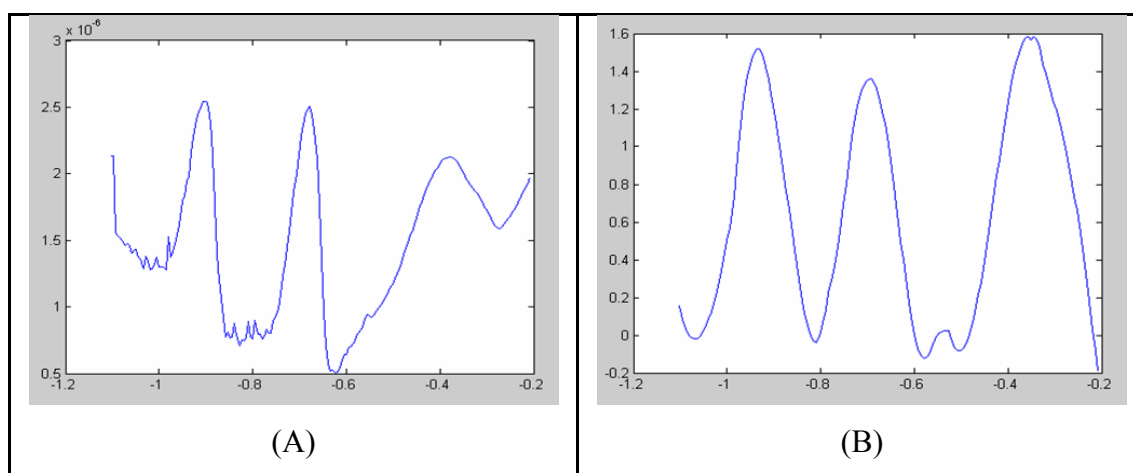
Figure 2.14: Voltammogram corresponding to 100ppb Cd, 100ppb Pb and 200ppb Cu: Raw (A) and weighted (B). High Weight = 2; Low Weight = 0.1

A number of weights were applied in order to improve the overall RMSEP values. The best of the weighted predictions are displayed in Tables 2.13 and 2.14.

Table 2.13: The predicted concentrations of the unknown "real" samples with the voltammograms shifted, weights applied (5.0:0.1; SGp = 3; SGw = 41), and no data pre-treatment followed by modelling with SIMPLS

	MEANS			STD DEV			% CV		
Sample	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	153.1	123.5	151.9	16.23	3.93	19.39	10.6	3.2	12.8
UNK2	119.9	61.3	43.2	17.32	11.87	55.04	14.4	19.4	127.5
UNK3	2.8	151.1	143.7	3.80	9.50	5.68	137.6	6.3	3.9

Table 2.14: The predicted concentrations of the unknown "real" samples with the voltammograms shifted, weights applied (5.0:0.1; SGp = 3; SGw = 41), with range-scaling, SG filtering (polynomial: 3; window: 41), followed by modelling with SIMPLS

	MEANS			STD DEV			% CV		
Sample	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	155.0	121.0	144.6	14.82	4.51	12.73	9.6	3.7	8.8
UNK2	123.7	65.0	45.7	13.60	11.55	54.68	11.0	17.8	119.8
UNK3	0.0	147.6	140.3	0.00	6.92	6.20	NaN	4.7	4.4

Table 2.15 shows the overall RMSEP values for Tables 2.8 to 2.11, 2.13 and 2.14. It is thus clear to see that application of re-aligning the potential peaks in the voltammetric response profiles has improved the overall predictions.

Table 2.15: Overall RMSEP values (in ppb) calculated for Tables 2.15 to 2.20

Table	2.8	2.9	2.10	2.11	2.13	2.14
RMSEP	22.90	20.91	20.56	20.07	32.62	33.75

The calculated overall RMSEP value for the univariate approach (Table 2.4) was 27.98 ppb. Comparing this value to the values in Table 2.15 shows that the multivariate approach has assisted in the calibration of the voltammetric peaks, the most effective being for the range-scaled re-aligned voltammograms (Table 2.11).

Although a separate validation dataset was acquired in order to perform independent test set validations, the data analysis package contains an additional option to split the training set into a separate calibration set and test set. In this instance, the last ten voltammograms in the original training set were set aside as the test set. The SIMPLS model was thus constructed with the calibration set and predictions of the test set were performed. Table 2.16 shows the calculated correlation coefficients for the predicted concentrations and the actual concentrations at varying data pre-treatments.

Table 2.16: Correlation coefficients calculated for the predicted concentrations plotted against actual concentrations. Note: S-G implies Savitzky-Golay smoothing at polynomial 3, window 41

Pre-treatment	Cd	Pb	Cu
Un-treated dataset	0.7500	0.9674	0.7930
Range-scaled and S-G dataset	0.7639	0.9744	0.8634
Dataset with re-aligned voltammograms	0.7769	0.9640	0.8930
Range-scaled and S-G re-aligned dataset	0.7673	0.9698	0.8487
Re-aligned and weighted dataset	0.8335	0.9526	0.7986
Range-scaled and S-G after re-aligned and weighted dataset	0.7782	0.9425	0.7747

In all cases, the concentration of lead was predicted more accurately than the other two metal species. The correlation coefficient for copper is greatly improved after the voltammograms have been re-aligned. Figure 2.15 displays the plot of the predicted concentrations against the actual concentrations for the three metal ions of the dataset comprising the corresponding re-aligned voltammograms.
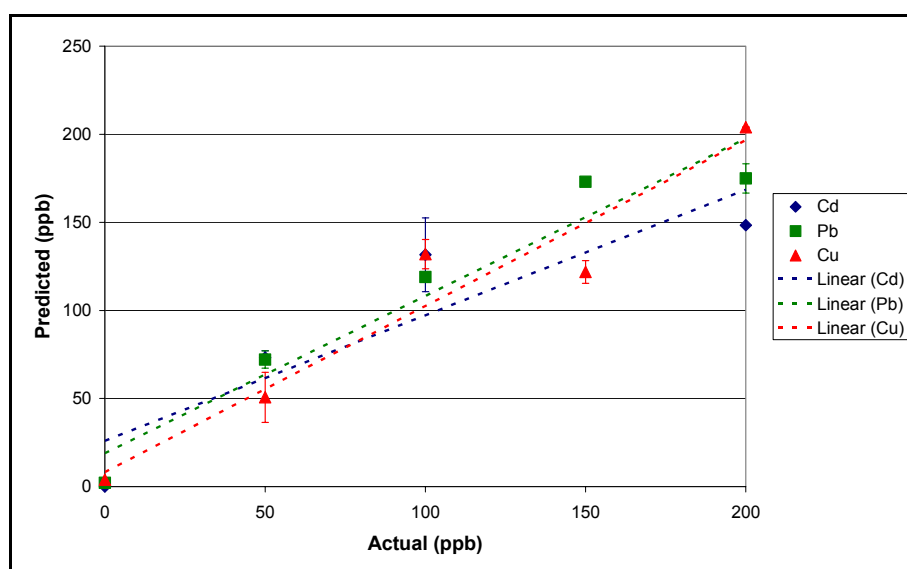
Figure 2.15: Plots of predicted concentration against actual for the three target metal ions from the shifted voltammograms. Reagent and instrumental parameters as in Figure 2.5 and Table 2.2 respectively

The plots in Figure 2.15 show that with the omission of the last ten samples from the original training set, the SIMPLS model generated contains sufficient information to successfully predict samples that were not included in the calibration set. Overall, the application of multivariate calibration with the C-SPE-O batch has led to improved

RMSEP values compared with the univariate approach. The application to the C-SPE-N batch is discussed next.

2.4.1.4 Multivariate Calibration on Batch C-SPE-N

The same experimental design was employed as for batch C-SPE-O except that each sample was measured on four separate occasions on individual electrodes as defined by the randomised design (Section 2.3.3). This leads to a training dataset containing 500 experiments. During the modelling process, two outliers were detected. Figure 2.16 displays the PLS score plot for the first outlier detected.
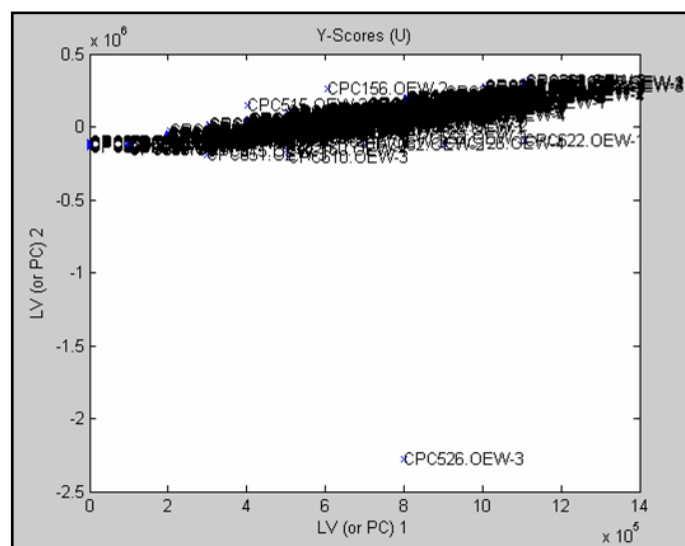
Figure 2.16: The PLS Y-Score (U) plot of LV2 vs LV1 clearly showing the presence of an outlier identified as CPC526-3 (3rd measurement of 50 ppb Cd, 200 ppb Pb, 150 ppb Cu). Reagent and instrumental parameters as in Figure 2.5 and Table 2.2 respectively

Possible causes of the voltammogram of the outlier are a defective electrode or the improper insertion into the connector. The sample was thus omitted from the training dataset. The second sample that was omitted was CPC652.OEW-3 (150 ppb Cd; 50 ppb Pb; 200 ppb Cu; 200 ppm Hg(NO$_3$)$_2$; 1% HNO$_3$; 0.1M KCl). This occurred after the application of the weights detailed in Table 2.17 (PCA scores plot not shown).

Table 2.17: The weighting parameters applied to the dataset after omission of the first outlier detailed in Figures 2.14 and 2.15.

Parameters	Value
High Weight	5.0
Low Weight	0.1
Indices (Matlab)	Cd (17, 45); Pb (65, 95); Cu (111, 163)
Indices (LabVIEW)	Cd (16, 44); Pb (64, 94); Cu (110, 162)
Savitzky-Golay Filter: Polynomial	3.0
Savitzky-Golay Filter: Window	41.0

Tables 2.18 to 2.23 show the means, standard deviations and percentage coefficient of variations for the prediction of the concentrations of the "real" samples. All predictions were performed after the incorporation of the validation dataset into the model. In all cases, predictions were performed at the 20th latent variable for all components as detailed in Section 2.4.1.2. Table 2.1 is displayed for convenience.

Table 2.1: Concentration values for the three "unknown" samples

Sample	Cd	Pb	Cu
UNK1	200	80	160
UNK2	100	20	0
UNK3	0	140	180

Table 2.18: The predicted concentrations of the unknown "real" samples with no data pre-treatment followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	240.5	85.6	166.6	49.76	29.84	33.14	20.7	34.9	19.9
UNK2	148.0	22.1	25.0	9.39	4.93	5.74	6.3	22.3	22.9
UNK3	0.0	146.5	197.6	0.00	8.04	16.53	NaN	5.5	8.4

Table 2.19: The predicted concentrations of the unknown "real" samples with range-scaling, SG filtering (polynomial: 3; window: 41), followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	238.3	91.6	173.7	57.62	40.81	32.45	24.2	44.5	18.7
UNK2	144.6	18.2	25.1	15.51	4.38	4.58	10.7	24.0	18.2
UNK3	0.0	144.8	193.5	0.00	11.83	17.52	NaN	8.2	9.1

Table 2.20: The predicted concentrations of the unknown "real" samples after omission of sample CPC526.OEW-3, and no data pre-treatment followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	236.1	86.3	163.8	43.56	27.99	28.80	18.5	32.4	17.6
UNK2	144.4	18.5	27.0	6.91	3.22	5.92	4.8	17.4	22.0
UNK3	0.0	147.5	196.0	0.00	8.76	15.37	NaN	5.9	7.8

Table 2.21: The predicted concentrations of the unknown "real" samples after omission of sample CPC526.OEW-3, with range-scaling, SG filtering (polynomial: 3; window: 41), followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	233.9	85.7	170.0	55.43	36.72	20.41	23.7	42.9	12.0
UNK2	145.6	19.0	26.6	9.00	1.52	4.59	6.2	8.0	17.2
UNK3	0.0	143.0	192.1	0.00	9.78	16.01	NaN	6.8	8.3

Table 2.22: The predicted concentrations of the unknown "real" samples after omission of sample CPC526.OEW-3, followed by weighting (Table 2.17) with no data pre-treatment followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	224.5	82.9	167.3	33.45	18.40	15.56	14.9	22.2	9.3
UNK2	134.8	29.7	37.1	5.29	4.71	42.80	3.9	15.9	115.5
UNK3	0.0	140.9	192.8	0.00	6.17	15.42	NaN	4.4	8.0

Table 2.23: The predicted concentrations of the unknown "real" samples after omission of sample CPC526.OEW-3, followed by weighting (Table 2.17), with range-scaling, SG filtering (polynomial: 3; window: 41) followed by modelling with SIMPLS

Sample	MEANS			STD DEV			% CV		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	215.1	76.4	161.3	13.98	8.97	11.61	6.5	11.7	7.2
UNK2	150.7	37.3	39.2	7.84	1.28	41.69	5.2	3.4	106.3
UNK3	1.3	142.1	187.4	2.61	8.28	19.54	200.0	5.8	10.4

By comparing the predicted means of the concentrations in Tables 2.18 to 2.23 against Table 2.5, it can be observed that some protocols have performed better for certain metal ions than others. For example, all except Table 2.23 have successfully predicted the cadmium concentration in UNK3 at 0 ppb. The relatively high %CV of 200 for the same sample (for Cd) in Table 2.23 can be considered not to be of great concern due to the actual concentration value being very low (close to zero). If

compared with the lead (Pb) concentration in the same sample (UNK3) the standard deviation (8.28 ppb) is greater but yields a lower %CV (5.8%) due to the relatively greater mean concentration (142.1).

All protocols have failed to predict the real concentration of copper in sample UNK2 at 0 ppb. This can be attributed to the effect of the mercury ions which implies that further pre-treatments are required. Ironically, the closest prediction is seen in Table 2.18 at 25.0 ppb which is prior to the removal of the first outlier. Table 2.24 displays the overall RMSEP values for the respective models created for Tables 2.18 to 2.23.

Table 2.24: Overall RMSEP values (in ppb) calculated for Tables 2.18 to 2.23

Table	2.18	2.19	2.20	2.21	2.22	2.23
RMSEP	23.57	22.65	22.04	21.67	19.74	22.88

The calculated overall RMSEP value for the univariate approach (Table 2.5) was 47.14 ppb. The omission of the first outlier has led to an improvement in the predictions of the respective generated models. Further, it is encouraging to observe the fact that the predictions improved on application of the weights (shown in Table 2.17) to the datasets (in conjunction with omission of the second outlier), which is in contrast to what was observed in Section 2.4.1.3 in which application of weights worsened the predictions.

As had been performed in Section 2.4.1.3, the original training set was divided into a calibration set and a test set which comprised the last ten samples from the training set. Table 2.25 shows the correlation coefficients calculated for the respective models.

Table 2.25: Correlation coefficients calculated for the predicted concentrations plotted against actual concentrations. Note: RS implies range-scaling and Savitzky-Golay smoothing

Pre-treatment	Cd	Pb	Cu
Omission of the first outlier only	0.9752	0.8983	0.9534
Omission of outlier and RS	0.9608	0.9238	0.9546
Omission of outliers and weights applied	0.9599	0.9155	0.9636
Outliers omitted, weights applied, and RS	0.9560	0.9119	0.9538

As can be seen by the correlation coefficients in Table 2.25 the prediction of lead has been the poorest by the four respective models generated. This is in contrast to what

was observed in Section 2.4.1.3 (Table 2.16). The correlation coefficients for cadmium and copper obtained with the C-SPE-N batch of screen-printed electrodes are much improved compared with the C-SPE-O batch (0.7782 and 0.7747 for cadmium and copper respectively after weighting, range-scaling and Savitzky-Golay smoothing). This is most likely as a consequence of the observed voltammetric shifts encountered with the C-SPE-O batch. However, it must be stressed that the calibration set employed for the C-SPE-N batch (500) was considerably larger than the one employed for the C-SPE-O batch (375). Further, due to the randomisation of the training datasets, the last ten samples in the respective test sets will be different.

Figure 2.17 compares four voltammograms originating from two different batches of electrodes (C-SPE-O and C-SPE-N) of two samples; one containing 150 ppb cadmium, lead and copper, and the second containing 150 ppb cadmium, 200 ppb lead and 50 ppb copper.
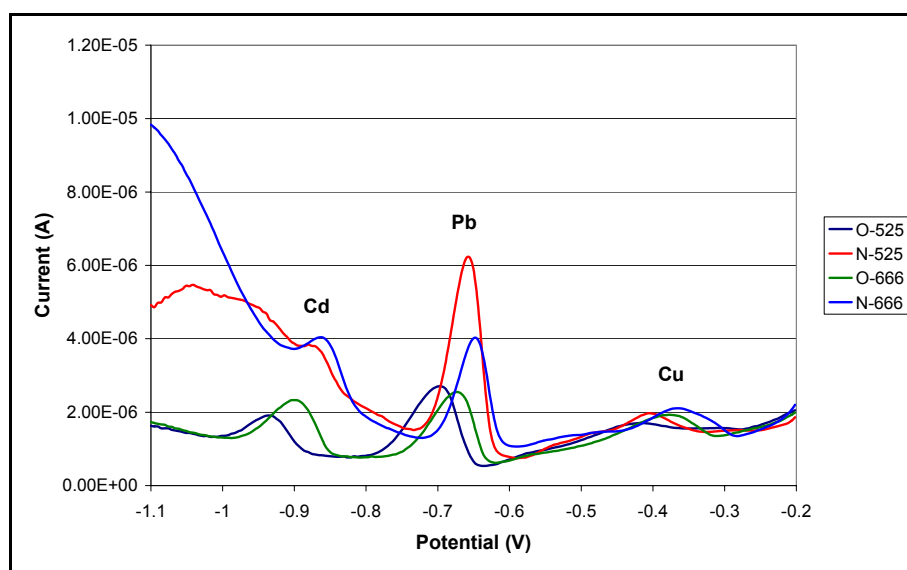
Figure 2.17: Four voltammograms recorded on the two batches: O-625 denotes sample CPC625 (150 ppb Cd; 200 ppb Pb; 50 ppb Cu) recorded on the C-SPE-O batch; N-625 denotes the same as O-625 but for the C-SPE-N batch; O-666 denotes sample CPC666 (150 ppb Cd, Pb and Cu) recorded on the C-SPE-O batch; N-666 denotes the same as O-666 but for the C-SPE-N batch. Reagent and instrumental conditions as in Figure 2.14 and Table 2.3 respectively

As has been previously observed (Figure 2.8), the high background current on the C-SPE-N batch has resulted in the masking of the cadmium (Cd) peak. This phenomenon was observed in many of the plots. Furthermore, it must be stressed that

the samples prepared in the laboratory were employed on both batches of SPEs which thus eliminates human error factors between batches, such that were any errors generated during the preparation of the samples, the effects would be seen on both batches.

The custom-built applications include an option to incorporate the validation set into the model should the RMSEP values be unsatisfactory. Incorporation of the validation set into the model increases the robustness of the model which should lead to the improvement of the efficiency of prediction of the respective model (Martin *et al.,* 2005). This was observed with the C-SPE-O batch. The same cannot be said for the C-SPE-N batch (Figure 2.18).
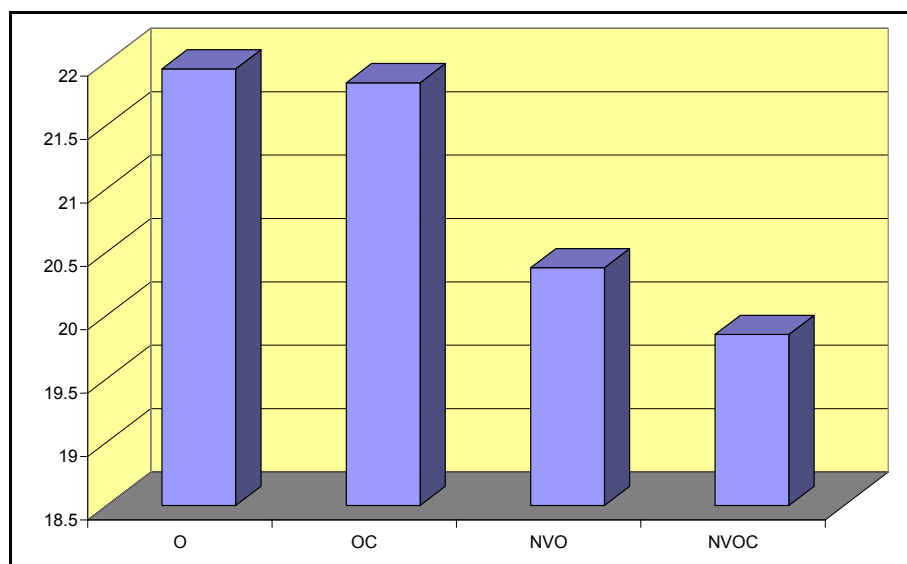
Figure 2.18: Overall RMS error values for the prediction of the samples with unknown concentrations on the C-SPE-N batch. O implies omission of the first outlier; OC implies O in addition to range-scaling and Savitsky-Golay smoothing; NV denotes the validation set was NOT incorporated into the model

The increase in RMSEP values after incorporation of the validation set into the model can be attributed to the high variability in the voltammograms recorded with regards to the high baseline observed from -1.1 to -0.8V. A second hypothesis could be that when the validation set is actually incorporated into the model, the set is merely appended to the end of the training set instead of the combined datasets being randomised. This issue could be addressed in future versions of the data analysis

package. However, this is most unlikely due to better predictions observed after incorporation of the validation set with the C-SPE-O batch.

Overall, the application of weights has improved the predictive ability of the respective model with a calculated RMSEP value of <20 ppb which is an improvement from the optimum value for the C-SPE-O batch at 20.07 ppb.

2.4.2 Application to Real Samples

In order to test the robustness of the multivariate calibration regression models, application to the quantitative determination of target species in real samples is warranted. This also involves validating the findings against a well-established analytical technique such as ICP-MS or AAS. This Section describes such an application (Samek *et al.,* 2001). It is split into several sub-sections. Section 2.4.2.1 will discuss the findings attained for the flame AAS whilst Section 2.4.2.2 for the graphite furnace AAS. Section 2.4.2.3 will discuss the findings via standard addition of the DPASV voltammograms whilst Section 2.4.2.4 will discuss the findings via the multivariate calibration regression models developed in Section 2.4.1. Finally Section 2.4.2.5 will compare all the results with appropriate conclusions drawn.

2.4.2.1 Flame AAS (F-AAS)

Table 2.26 details the hardware parameters and conditions employed for the F-AAS (Section 1.4.4.1) and Table 2.27 lists the parameters employed for the target analytes.

Table 2.26: Hardware and computational parameters for the F-AAS instrument

Atomiser	Standard Burner	Delay Time (s)	15
Measurement Mode	Working Curve	Signal Record	Direct
Flame	Acetylene-Air (2125 - 2400°C	Signal Mode	Background Correct
Oxidant Pressure (KPa)	160	Replicates	Std = 3 Sample = 3
Oxidant Flow (L/min)	15	Calculation	Integration
Calculation Time (s)	5	Slicing Height	10% (P. W. Only)

Table 2.27: Specific instrumental parameters for the target elements

Element	Current (mA)	Wavelength (nm)	Slit (nm)	Fuel Flow (L/min)	Burner Height (mm)
Cd	7.5	228.8	1.3	1.5	5.0
Pb	7.5	283.3	1.3	1.7	7.5
Cu	7.5	324.8	1.3	1.7	7.5

The mode of operation for the F-AAS was to warm up the lamp; position it for the optimum intensity through the flame; self-calibration; samples measured with a cleaning step in between via the auto-sampler.

Calibration standards were prepared by diluting a 1000 ppm commercial stock standard (Fisher Scientific, Leicester, UK) to 1 ppm by taking 1 ml of the 1000 ppm and adding 999ml of RO water. The individual standards were then prepared by diluting 0, 1, 2, 3, 4 and 5ml of the 1ppm standard to 25 ml hence preparing standards with the following respective concentrations of 0, 40, 80, 120, 160 and 200 ppb. Furthermore, 1ml of the filtered CRM extract (prepared as detailed in Section 2.3.4) was diluted 25-fold prior to analysis. However, this dilution was too great yielding % recoveries in excess of 150% for the majority of the target analytes. Standards were thus prepared again for the three metal ions in addition to employing a five-fold dilution of the filtered CRM extract. Figures 2.19, 2.20 and 2.21 display the calibration curves for cadmium, lead and copper respectively.
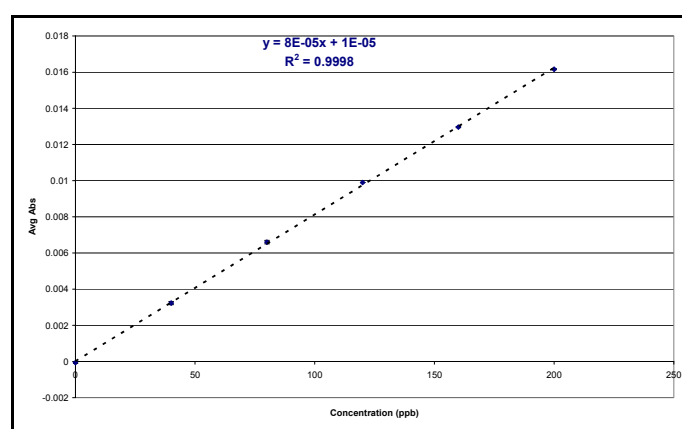
Figure 2.19: Calibration curve for cadmium performed via F-AAS.

Cd concentrations: 0, 40, 80, 120, 160 and 200 ppb
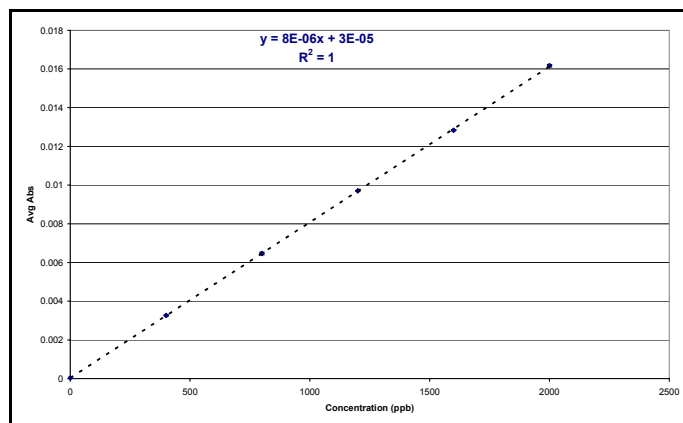
Figure 2.20: Calibration curve for lead performed via F-AAS.

Pb concentrations: 0, 400, 800, 1200, 1600, 2000 ppb.
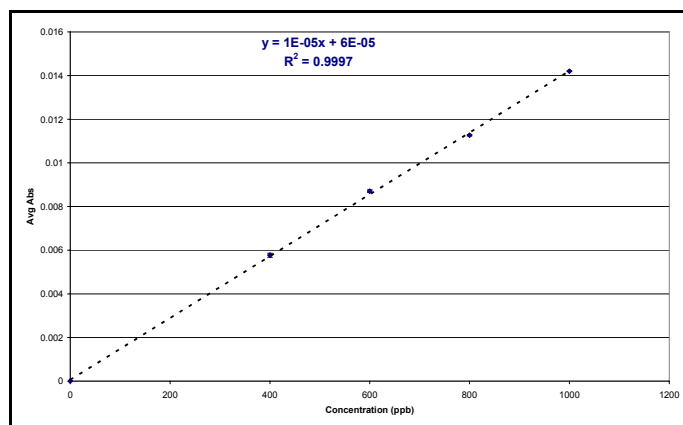
Figure 2.21: Calibration curve for copper performed via F-AAS.

Cu concentrations: 0, 200, 400, 600, 800, 1000 ppb

By employing a smaller dilution factor coupled with a wider concentration range, the absorbance values obtained were 5- to 10-fold more distinct from the noise leading to more reliable concentrations calculated resulting in greater confidence in the calculation of % recovery. Table 2.28 lists the % recovery for all three target elements for the CRM extract.

Table 2.28: Calculated % recovery for Cd, Pb and Cu via F-AAS of the filtered CRM extract after application of a dilution factor of 5.

Element	% Recovery from CRM
Cd	31.73%
Pb	9.71%
Cu	27.83%

2.4.2.2 Graphite Furnace AAS (GF-AAS)

The GF-AAS technique is generally more sensitive than the flame technique due to the smaller sample volumes employed coupled with elevated temperatures (Rouessac & Rouessac, 2000). Table 2.29 lists the hardware parameters and conditions employed for the GF-AAS and Table 2.30 lists the parameters for the specific target elements.

Table 2.29: Hardware and computational parameters for the GF-AAS

Cuvette	Pyro (electrothermal)	**Signal Mode**	Background Correct
Measurement Mode	Working Curve	**Replicates**	Std = 3 Sample = 3
Temperature Control	Optimum	**Calculation**	Peak Height
Signal Record	Direct	**Slicing Height**	10% (P. W. Only)

Table 2.30: Specific instrumental parameters for the target elements

Element	Current (mA)	Wavelength (nm)	Slit (nm)	Ashing and duration °C (s)	Atomising and duration °C (s)	Cleaning and duration °C (s)	Sample Volume (μl)
Cd	7.5	228.8	1.3	300 (30)	1500 (10)	1800 (4)	20
Pb	7.5	283.3	1.3	400 (30)	2000 (10)	2400 (4)	20
Cu	7.5	324.8	1.3	600 (30)	2700 (10)	2800 (4)	20

The acquisition time is longer than with the flame due to gradual increases in temperature between the stages listed in Table 2.30 in the following order: ashing, atomising and cleaning. Prior to the ashing is a drying stage which for all three target elements ranges between 80 to 120°C for over 30 seconds. Figures 2.22, 2.23 and 2.24 display the calibration curves for cadmium, lead and copper acquired on the GF-AAS respectively.
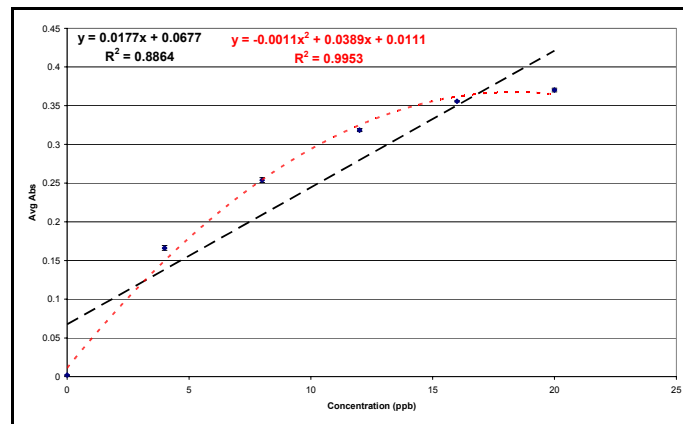
Figure 2.22: Calibration curve for cadmium performed via GF-AAS.

Cd concentrations: 0, 4, 8, 12, 16 and 20 ppb
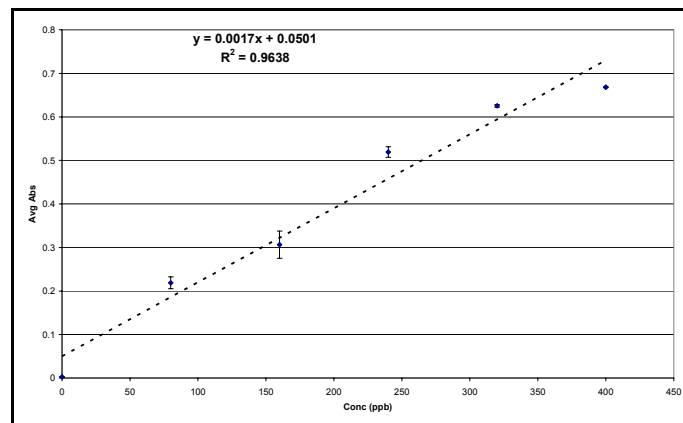
Figure 2.23: Calibration curve for lead performed via GF-AAS.

Pb concentrations: 0, 80, 160, 240, 320, and 400 ppb
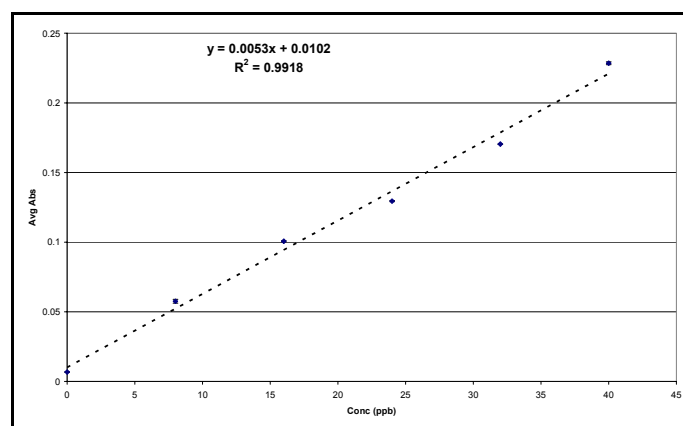
Figure 2.24: Calibration curve for copper performed via GF-AAS.

Cu concentrations: 0, 8, 16, 24, 32 and 40 ppb

The calibration curve for cadmium (Figure 2.22) is non-linear and closely follows the quadratic equation as shown in red. The correlation coefficients for lead and copper

are better. The error bar for the 160 ppb lead standard indicates that there was considerable variation when that standard was measured. The curve also possesses non-linear tendencies.

In order to ensure that the filtered CRM extract was sufficient, four dilution factors were employed: 5, 10, 25 and 50. Table 2.31 shows the % recoveries for each target element at each dilution factor.

Table 2.31: Calculated % recovery for Cd, Pb and Cu via GF-AAS of the filtered CRM extract after application of the specified dilution factors

Element	Dilution Factor			
	5	10	25	50
Cd	6.29%	10.20%	13.00%	15.86%
Pb	1.38%	0.76%	0.27%	29.53%
Cu	14.42%[*]	20.62%	42.37%	71.89%

Note that the cadmium calculation was via the straight line regression ($R^2 = 0.8864$). The % recovery for copper at a dilution factor of 5 (denoted by *) must be omitted due to the absorbance values exceeding the calibration range. Overall, one would have expected to observe the % recovery to be similar taking the dilution factors into account. The dilution factor of 50 was made from a 1:2 dilution of the 25 factor. The % recovery selected was based on the respective absorbance values being in the centre of the absorbance values related to the respective concentration ranges. In this instance, the % recovery for cadmium is thus 15.86 (50); lead is thus 29.53 (50); copper is thus 20.62 (10).

2.4.2.3 Standard Addition

Figure 2.25 displays the overlaid voltammograms acquired on the Autolab pertaining to the prepared CRM samples with increasing concentrations (100, 150 and 200 ppb) of cadmium, lead and copper.
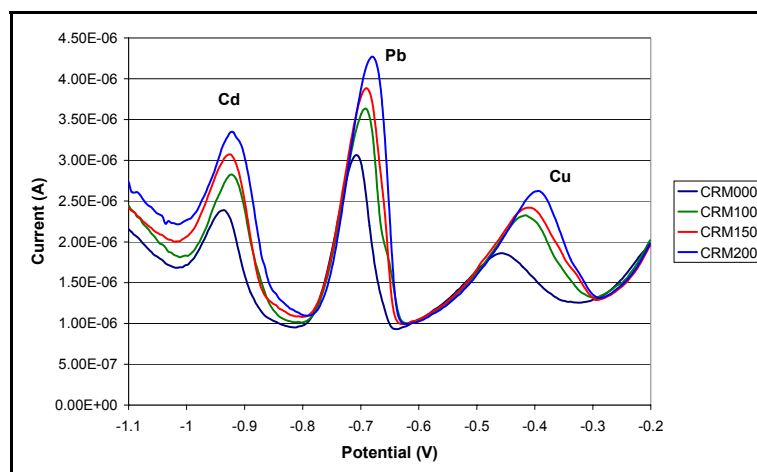
Figure 2.25: Overlaid voltammograms of Cd, Pb and Cu in the filtered CRM extract. Instrumental conditions: Autolab PSTAT10; deposition time: 165s; deposition potential: -1.1V; scan rate: 10mV/s)

It is evident to see that the extraction procedure employed has been successful in extracting the target metal ions. Figure 2.26 displays the standard addition curves for the three target elements which were recorded in triplicate.
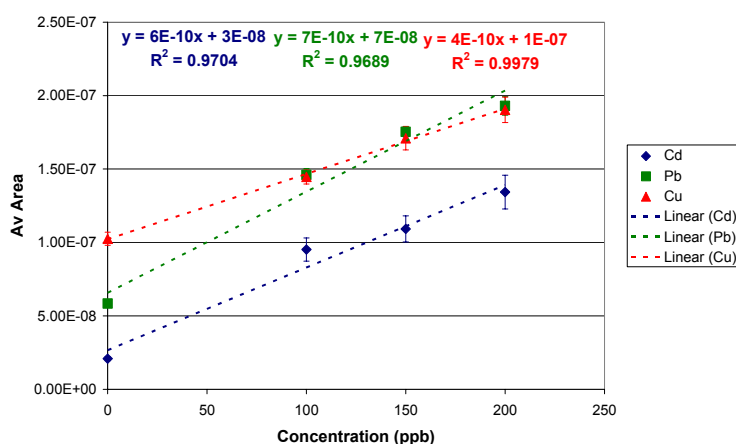
Figure 2.26: Standard addition performed after calculation of the peak areas obtained in Figure 2.25.

As is clear to see in Figure 2.26, the correlations coefficients are very good with values of 0.9851, 0.9843 and 0.9989 for cadmium, lead and copper respectively. The error bars gradually increase as the concentrations do, particularly for copper which is to be expected due to its proximity to mercury as reflected in the increase in %CV from 5.22 to 12.18%. By comparison of the coefficients of determination (R^2), one can deduce that addition of copper is more linear than the other two elements. This

implies that a determination of the actual amount of each metal analyte in the CRM sample is possible. Table 2.32 shows the % recovery calculated from the amounts in the CRM.

Table 2.32: % recovery calculated for Cd, Pb and Cu via standard addition of the filtered CRM extract after application of a dilution factor of 5

Element	% Recovery from CRM
Cd	22.40%
Pb	21.51%
Cu	74.67%

It can be observed that the % recovery calculated for copper in Table 2.32 conforms to that calculated for copper at a 50-fold dilution (Table 2.31). Prior to any further comparisons and inferences, Section 2.4.2.4 discusses the application of SIMPLS models developed in Section 2.4.1.

2.4.2.4 Multivariate Calibration Models

Tables 2.33 to 2.38 show the calculated % recoveries taking into account the appropriate dilution factors via the respective SIMPLS models for the latent variables suggested by the minimum RMSEP values and the F-test, in addition to the predictions of all three target metal ions via the 20th latent variable (LV20).

Table 2.33: Determination of the % recovery of Cd, Pb and Cu present in the CRM via the SIMPLS model constructed from range-scaled voltammograms (C-SPE-O)

Target Ion	Min RMSEP	F-test	LV20
Cd	22.63%	27.58%	21.40%
Pb	19.88%	22.45%	19.09%
Cu	11.87%	17.56%	9.66%

Table 2.34: Determination of the % recovery of Cd, Pb and Cu present in the CRM via the SIMPLS model constructed from shifted and range-scaled voltammograms (C-SPE-O)

Target Ion	Min RMSEP	F-test	LV20
Cd	23.12%	28.51%	20.14%
Pb	18.04%	21.54%	18.04%
Cu	13.10%	15.67%	10.79%

Table 2.35: Determination of the % recovery of Cd, Pb and Cu present in the CRM via the SIMPLS model constructed from shifted and mean-centred voltammograms (C-SPE-O)

Target Ion	Min RMSEP	F-test	LV20
Cd	15.58%	23.88%	17.67%
Pb	15.76%	18.19%	16.31%
Cu	9.82%	8.69%	10.35%

Table 2.36: Determination of the % recovery of Cd, Pb and Cu present in the CRM via the SIMPLS model constructed from shifted and weighted voltammograms (C-SPE-O)

Target Ion	Min RMSEP	F-test	LV20
Cd	25.21%	30.26%	25.21%
Pb	14.69%	17.05%	15.28%
Cu	14.09%	17.66%	14.77%

Table 2.37: Determination of the % recovery of Cd, Pb and Cu present in the CRM via the SIMPLS model constructed from weighted voltammograms after omission of an outlier (C-SPE-N)

Target Ion	Min RMSEP	F-test	LV20
Cd	44.10%	7.10%	44.10%
Pb	14.06%	7.98%	15.34%
Cu	17.84%	22.52%	17.84%

Table 2.38: Determination of the % recovery of Cd, Pb and Cu present in the CRM via the SIMPLS model constructed from weighted and range-scaled voltammograms after omission of an outlier (C-SPE-N)

Target Ion	Min RMSEP	F-test	LV20
Cd	32.71%	3.04%	32.71%
Pb	20.76%	14.09%	18.10%
Cu	21.99%	19.95%	21.99%

Tables 2.33 to 2.38 have shown varying amounts in the calculated % recovery via the respective SIMPLS regression models with the majority ranging between 15 and 35%. None of the models have predicted a copper extraction of ~70 – 75% as seen in Tables 2.31 and 2.32. This implies that the models have accounted for the presence of mercury in the solution compared to the standard addition method. The next Section will compare the entire calculated % recovery for the target analytes.

2.4.2.5 Comparison of Calculated % Recovery via AAS, Standard Addition and Multivariate Calibration Models

Table 2.39 compares the "best" calculated % recovery from the above sections leading to the actual determined % recovery for each respective element.

Table 2.39: The best agreements deduced from the comparisons of the above sections.
Key: (1) SIMPLS model constructed from re-aligned and mean-centred voltammograms (C-SPE-O); (2) SIMPLS model constructed from re-aligned and range-scaled voltammograms (C-SPE-O);
(3) SIMPLS model constructed from weighted and range-scaled voltammograms after omission of an outlier (C-SPE-N)

Element	GF-AAS	Standard Addition	Chemometrics
Cd	15.86%	22.54%	15.58%(1)
Pb	29.53%	21.51%	21.54% (2)
Cu	20.62%	74.67%*	19.95% (3)

It is encouraging to observe that the calculated values for the multivariate calibration models agree with the values for the GF-AAS, especially for cadmium and copper. The high value for the calculated % recovery for copper can be attributed to the presence of the mercury in addition to any other impurities (organic) that are present in the soil and was not removed during the digestion procedure. Furthermore, the effect of the background current in relation to the decrease and increase in baseline for cadmium and copper respectively contribute to their respective poor recoveries via chemometrics.

2.5 Conclusions

The custom-built data analysis package constructed in both the Matlab and LabVIEW environments have been successfully applied to the quantitative determination of cadmium, lead and copper. The package imports the voltammograms, processes them and exports the results for future processing on other dedicated analysis packages. However, a future task will be to store the data and results generated into one large compressed file or a database, instead of the multitude of text files that are currently generated. The database would hold such information as the predicted data, information on the captured variance generated from the construction of the calibration models, and the RMS error values.

The data analysis package developed can construct a range of multivariate calibration regression models such as PCR, NIPALS, SIMPLS and ANNs (only in Matlab) in addition to performing a wide range of data pre-treatment techniques such as range-scaling, mean-centring, weighting and smoothing. Outliers can be omitted via preliminary analysis with PCA. In this study, the better algorithm to employ is SIMPLS as illustrated by the RMSEP value of 4.25 ppb when compared against NIPALS and PCR (Figure 2.10), and 21 – 38 ppb when compared against ANNs at 46 to 67 ppb (Figure 2.11). Although ANNs is a more powerful algorithm, it is most beneficial for very large and non-linear data sets (Richards *et al.,* 2002). The datasets employed here are linear at the current concentration range. The combination of outlier omission, weighting, range-scaling and Savitzky-Golay smoothing has seen a vast improvement in the root mean square error of prediction (RMSEP) values for the target metal ions (for example 19.74 ppb for sample omission and weighting of the voltammograms on the C-SPE-N batch) compared to the univariate approach (47.14 ppb).

It had been reported that the C-SPE-O batch of carbon-ink screen-printed electrodes perform much better than the C-SPE-N batch, even though some voltammograms acquired with the older batch were subjected to a shift in the negative potential, as much as -0.15V in some instances. However, the SIMPLS algorithm was still able to model the voltammograms after pre-treatment with a custom-built re-alignment

module. It was also observed that the deposition currents on the C-SPE-O batch are lower in magnitude than the C-SPE-N batch. This implies that the carbon inks employed in the manufacture of both batches are likely to possess different characteristics.

Graphically, one can clearly see the potential peak maxima obtained on the C-SPE-O batch are greater than those of the C-SPE-N batch. Attempts were made to improve the responsiveness of the C-SPE-N electrode batch by increasing the deposition time, applying a conditioning potential and high temperature electrode treatment, as suggested by the ink manufacturer. Although improvements were observed in peak height and area with cadmium and copper, the opposite was seen for lead.

The % recovery as determined by GF-AAS for cadmium, lead and copper is 15.86%, 29.53% and 20.62% respectively. Very good agreements are observed with the multivariate calibration models with 15.85%, 21.54% and 19.95% respectively. The % recovery calculated via the standard addition of lead is 21.51%. This is likely to be attributed to the lead peak coinciding with a plateau in the baseline whereas for cadmium and copper, the baselines were steep leading to complications with respect to peak area and height determination which leads to the increase in the % recovery. Furthermore, a wider concentration range could have been employed in order to determine more accurately the % recovery of the respective target metal analytes. The % recovery as determined by F-AAS was deemed unreliable due to the absorbance values being close to the noise.

During and after the development of the custom-built data analysis package, application of data pre-treatment techniques, modelling, validation and prediction to voltammograms acquired on gold-ink screen-printed electrodes for the quantitative determination of arsenic, mercury and copper was performed. The outcome is thus discussed in Chapter 3.

CHAPTER 3:

QUANTITATIVE DETERMINATION OF ARSENIC, MERCURY AND COPPER ON GOLD-INK SCREEN-PRINTED ELECTRODES

3.1 Overview

This chapter focuses on the application of multivariate calibration to DPASV voltammograms acquired on novel gold-ink screen-printed electrodes. A general introduction will discuss and compare mercury and gold working electrodes in relation to the quantitative analysis of arsenic. References to the application of multivariate calibration techniques to the acquired voltammograms will also be made. The materials and methods employed will be detailed followed by an in-depth discussion of the results obtained which include the qualitative comparison of a number of batches of disposable gold-ink screen-printed electrodes, application of the custom-built data analysis package described in Chapter 2 to account for the overlapping peaks attained in the respective voltammograms, and application to a real soil sample. Finally, an overall conclusion is drawn.

3.2 Introduction

Chapter 2 discussed the successful application of multivariate calibration to voltammograms acquired on carbon-ink screen-printed electrodes for the determination of cadmium, lead and copper. The main advantages of these electrodes are the relatively low cost in fabrication, their disposability leading to the omission of laborious cleaning steps, and the application of small sample volumes which therefore minimises reagent wastes. The major disadvantage with regards to the carbon-ink screen-printed electrodes is the use of a mercury salt in the form of mercuric (II) nitrate at a relatively high concentration compared to the target metal analytes which co-deposit on to the working electrode by amalgamating with the reduced mercury film.

3.2.1 Mercury Working Electrodes

In recent years, environmental agencies have grown particularly concerned with the increasing amounts of arsenic and mercury in the environment (Cooper, 2004). Legislation has thus been established in which these elements must be quantitatively determined preferably in the sub-ppb range. Arsenic and mercury are quantitatively determined by hydride generation AAS and cold-vapour AAS respectively (Rouessac & Rouessac, 2000). However, growing interest in field-based methods makes these analytical techniques impractical in terms of portability.

The acquisition of arsenic and mercury on carbon-ink screen-printed electrodes is not feasible. Mercury cannot be determined because of the use of the mercury salt as mentioned above to form a thin mercury film on the carbon working electrode. Further, the oxidation potential of arsenic is more positive than that of mercury implying that an amalgam will not be formed. However, this does not imply that arsenic cannot be determined with mercury electrodes. In the presence of copper and employing differential pulse cathodic stripping voltammetry (DPCSV), arsenic can in fact be quantitatively determined on a mercury electrode (Sadana, 1983; Zima & van den Berg, 1994; Li & Smart, 1996).

DPASV cannot be employed in conjunction with mercury working electrodes for arsenic determination because of reasons given in the previous paragraph. However, it can be employed in conjunction with gold working electrodes because the surface nature of the gold metal allows direct depositioning of the metals such as arsenic and copper (Matsumoto *et al.,* 1994). In turn, due to the inter-metallic compounds formed between copper and arsenic, the electrochemical method of cathodic stripping voltammetry (CSV) can instead be employed (Sadana, 1983). Employing mercury working electrodes as with a hanging mercury drop electrode (HMDE) meant that rigorous and laborious cleaning processes could be omitted since a non-fouled mercury electrode surface is recreated with each successive droplet formation. Gold electrodes were prone to extreme fouling mostly due to the formation of an oxide layer which was difficult to remove. Furthermore, they suffer from poor precision, a limited sensitivity and serious "memory" effects (Zima & van den Berg, 1994; Li & Smart, 1996). However, with the advent of novel gold-ink screen-printed electrodes

initially employed by Cooper (Cooper, 2004) and discussed further in Section 3.4 these effects have been overcome.

The presence of copper in samples can lead to the formation of inter-metallic compounds in the presence of arsenic on mercury working electrodes (Sadana, 1983). Although it was admitted by the authors that the role of the copper was not fully understood, the following reaction was proposed:

$$Cu_3As_2 + 12H^+ + 3Hg + 12e^- \rightarrow 2AsH_3 + 3H_2 + 3Hg(Cu)$$
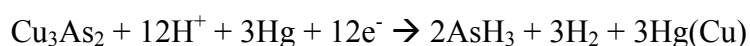
The formation of copper arsenide was postulated to occur during the depositioning step:

$$2As^{3+} + 3Hg(Cu) + 6e^- \rightarrow Cu_3As_2 + 3Hg$$
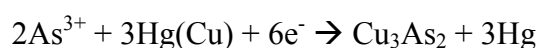
It was further postulated that the copper that is released from the inter-metallic compound with arsenic dissolves in to the mercury which leads to the formation of a voltammetric peak thus rendering arsenic quantification problematic. It was however observed that there are many stoichiometric forms of the inter-metallic compound some of which are actually unknown. This was postulated from the voltammetric profiles attained. Furthermore, hydrogen is also released. A copper (II) concentration of $5 \pm 1\mu g/ml$ gave well-defined voltammetric peaks relating to the presence of arsenic. It was therefore advised that this amount be added to a given ground water sample provided that the natural copper concentration did not exceed $6\mu g/ml$. If so then a simple dilution to $5\mu g/ml$ would suffice. In this way, the arsenic could be determined via standard addition in which arsenic would be added without the need to add more copper.

This method of arsenic determination with DPCSV with mercury electrodes has been employed for many years. A common application is the determination of arsenic (III) in sea water (Zima & van den Berg, 1994). The authors reported that pre-concentrating in the presence of copper led to a non-linear response. They proposed complexing arsenic (III) with pyrrolidine dithiocarbamate (PDC) due to the complex being stable in acidic media. The optimum depositioning potential was determined at

-0.3V (vs Ag/AgCl) with varying times depending on the concentration of the arsenic (III); one minute if greater than 100mM or three minutes if less. The stripping peak was thus observed at -0.2V. Copper was found to interfere with the determination of the arsenic but not cadmium or lead.

3.2.2 Gold Working Electrodes

With the onus on performing faster, more reliable determinations with minimal impact on the environment by reducing waste solvent production, alternative electrochemical procedures have been evaluated. Alternative working electrodes are thus required. Gold electrodes have been employed for many years alongside glassy carbon, mercury film and platinum electrodes (Rouessac & Rouessac, 2000). These electrodes also possess varying potential ranges in aqueous solutions as shown in Table 3.1.

Table 3.1: The potential windows, in aqueous solutions, of four types of working electrodes in 1M H_2SO_4 electrolyte

Type	Potential window (vs Ag/AgCl)
Glassy carbon	-0.6V $\rightarrow$ +1.3V
Gold	-0.3V $\rightarrow$ +1.0V
Mercury	-1.3V $\rightarrow$ +0.3V
Platinum	0.0V $\rightarrow$ +1.2V

The potential ranges are also dependent on the solvent composition, electrolyte and the reference electrode used (Skoog *et al.,* 1996). The potential range for a mercury electrode in 1M KCl against a saturated calomel electrode (SCE) is -1.8V to +0.2V. In aqueous solutions, these ranges are defined by the simultaneous oxidation and reduction of water to give oxygen at positive potentials and hydrogen at negative potentials respectively.

Gold disk microelectrodes (GDM) were fabricated in-house to determine the amount of disulfiram (DSF) – a fungicide – by *Differential Pulse Adsorptive Stripping Voltammetry* (DP-AdSV) (Agüí *et al.,* 2002). A microelectrode was used to reduce the deposition time. The GDM was pre-treated daily by polishing with diamond powder for one minute before placing into a buffer solution of $H_2PO_4^-/HPO_4^{2-}$ (at pH

6.0) and successively applying a range of negative potentials from -0.4 to -1.5V at intervals of -0.2 to -0.3V for thirty seconds each. The DSF adsorbed well to the GDM. Compared to conventional macro-electrodes, the GDM also offers better signal-to-noise ratios corresponding to a detection limit of 6.3×10^{-8} mol l^{-1} and greater precision (RSD of 1.3%).

Gold working electrodes tend to suffer from "memory effects" as a result of incomplete removal of copper from the electrode surface in addition to poor repeatability and low sensitivity. The procedures required for regenerating the surface of the gold working electrode tends to be time-consuming (Feeney & Kounaves, 2002). However, this is deemed a necessity if quantitative results are required. Feeney and Kounaves developed a portable micro-array of gold electrodes which was employed for both field and *in-situ* analysis. SWASV was the voltammetric method employed. A limit of detection of 0.1 ppb was attained for arsenic. It was observed that copper suppressed the arsenic signal significantly. For example, the signal corresponding to 100 ppb As (III) was suppressed by 50% when 50 ppb Cu (II) was present as observed with mercury working electrodes. The authors concluded by encouraging the development of on-site monitoring and screening devices.

Developing portable and robust devices has become a major goal for researchers over the recent years. One example is the development of a portable instrument that employed a flow cell in which a gold film was renewed on a platinum wire electrode (Huang & Dasgupta, 1999). The platinum electrode is cleaned by application of a potential at +0.95V (vs Ag/AgCl) for 5 seconds. As the cell is flushed with a gold (III) solution (50 ppm) a potential of -0.1V is applied for 9 seconds. Finally, a potential of +0.6V is applied which removes any impurities which will have co-deposited with the gold film. Quantitative determinations of arsenic (III) and total arsenic were performed with limits of detection in the region of 0.5 ppb attained at a deposition potential of 80 seconds. However the authors also noticed that in the presence of copper (~20 fold), the arsenic signal was suppressed (~50%). The arsenic signal decreased further as the concentration of copper was increased. The authors therefore attributed this phenomenon to the formation of the inter-metallic compound Cu_3As.

Multivariate calibration can take into account the effects of inter-metallic compound formation. Section 2.2 has discussed the applications of a number of techniques to voltammetric datasets employing CLS, MLR, PCR, PLS, and ANNs. PLS regression was employed to determine the amount of arsenic (As) in the presence of copper by stripping voltammetry on gold electrodes (Jagner *et al.,* 1994). The authors were able to model such effects as the inter-metallic compound formation of arsenic with copper, stripping peak overlap with tin (IV), and at concentrations of arsenic below 1µg/L, and the background effect of gold working electrode surface. Although the arsenic concentration in "natural waters" is in the "sub-µg/L" level, interference from copper (II) can make the accurate determination of arsenic difficult. The authors, however, showed that PLS could overcome this. They also stress that "at higher concentration ranges, the calibration design must provide ample variations in the concentrations of the analytes". Table 3.2 details very good agreement attained when the predicted concentrations of arsenic (V) in synthetic samples where the copper (II) and tin (IV) concentrations were treated as unknowns.

Table 3.2: Results from PLS calibration predictions of arsenic (V) in the presence of 100 µg l⁻¹ copper (II) and 500 µg l⁻¹ tin (IV) treated as unknowns

True concentration of As ($\mu g\ l^{-1}$)	Number of Samples	Predicted concentration of As ($\mu g\ l^{-1}$)
10	6	11.8 ± 1.3 (11%)
30	6	30.0 ± 1.3 (4.2%)
50	4	49.2 ± 0.9 (1.8%)
70	6	69.6 ± 1.2 (1.7%)
90	6	88.9 ± 1.9 (2.1%)

Table 3.1 listed the potential window of a number of working electrodes. In some cases, these electrodes can be employed simultaneously such as in "electronic tongues" (Winquist *et al.,* 1997; Winquist *et al.,* 2000; Winquist *et al.,* 2005). These are employed for the analyses of liquids just as "electronic noses" are employed for the analyses and identification of gases. They both contain an array of sensors with specific but differing characteristics which are tailored for identifying specific components in the liquids.

Electronic tongues initially developed were based on potentiometry which thus limited their application to charged species in addition to being sensitive to noise (Winquist *et al.,* 1997). However, advancements in electronics led to the application of voltammetry which broadened the scope of selectivity. The advantages of voltammetry over potentiometry were the ability to apply a potential in addition to varying the scan rate, pulse amplitude and the mode of stripping. The data generated from these sensors was plentiful and thus warranted the application of multivariate analysis in the form of principal component analysis (PCA) for data reduction and artificial neural networks (ANNs) for modelling (Winquist *et al.,* 2000). A major application for electronic tongues is in on-line monitoring. However, such devices must be robust and be able to withstand harsh conditions such as strongly acidic and alkali environments (Winquist *et al.,* 2005).

3.2.3 Disposable Gold-Ink Working Electrodes

The majority of working electrodes still suffer from the cumbersome task of intense cleaning and polishing, such as with the gold disk electrode especially if the cleaning cycle is not equally efficient leading to variance in the measurements. The need for disposable electrodes is paramount for a more rapid turn-around of data generation. Gold-ink screen-printed electrodes have recently been successfully employed in this laboratory to determine and quantify the amounts of arsenic and mercury in real and synthetic samples (Cooper, 2004). Optimal parameters obtained have been a deposition potential of 0.0V, a deposition time of 30 seconds and a 4M HCl electrolyte. However, the voltammetric peaks obtained tended to be rather broad. Furthermore, arsenic speciation factors could not be accounted for in one measurement, and thus reduction of the less toxic arsenic (V) was performed with L-cysteine following a set method and determined under differing voltammetric conditions than for arsenic (III). In addition, if in the presence of copper, the quantification of mercury could not be accounted for due to the two elements possessing very similar electrode potentials and thus leading to severe voltammetric peak overlap.

The application of multivariate calibration techniques is thus a necessity in this instance. To the author's knowledge, applications of multivariate calibration to gold-

ink screen-printed electrodes for the simultaneous determination of arsenic, mercury and copper have not been reported in the literature. This chapter will therefore discuss the application of multivariate calibration on voltammograms acquired on disposable gold-ink screen-printed electrodes. The deconvolution of the overlapping mercury and copper peaks is thus deemed of paramount importance.

It must be noted that gold-ink screen-printed electrodes have been in use for many years for example in the electronics industry (Hoffman & Nakayama, 1968). However, the manufacture of these electrodes is costly due to the high curing temperatures (~ 750°C) and the use of a ceramic substrate. The inks employed in this study are a new concept in that they can be cured at much lower temperatures (~ 150°C) and so can be printed on to plastic sheets making them low-cost and therefore amenable to mass production.

In addition to the evaluation of these inks, the protocols developed and discussed in Chapter 2 will also be applied with the aim of quantifying the presence of the three target metals: arsenic, mercury and copper.

3.3 Materials and Methods

The gold-ink screen-printed electrodes were primarily employed to detect arsenic (III), mercury (II) and copper (II). Background information on two of these three target metal ions is given followed by details of the methods and materials employed.

3.3.1 The Target Metal Analytes

The physical attributes of arsenic and mercury are presented in Sections 3.3.1.1 and 3.3.1.2 respectively.

3.3.1.1 Arsenic (As)

Arsenic was one of the earliest elements to have been discovered, even before the elemental discovery of nitrogen (in 1772) and phosphorus (in 1669) dating back to the

fifth century BC (Greenwood & Earnshaw, 1990). Early recorded isolations of the element were made by Albertus Magnus (AD 1193 – 1280) who heated *orpiment* (As_2S_3) with soap. Arsenic is in fact classed as a semi-metal (or metalloid). It is known to chemists as a *chalcogen* due to it being found in combination with sulphur (S), selenium (Se) and tellurium (Te), rather than as oxides or in silicates. Its abundance in the Earth's crust is on average 1.8 ppm.

On an industrial scale, arsenic (As) is obtained by the smelting of *arsenopyrite* (FeAsS) or *loellingite* ($FeAs_2$) in the absence of air at a temperature of 700°C. The arsenic gas that is produced is then condensed. It is mainly used in alloys with copper and lead. Organo-metallic compounds of arsenic are used considerably in the agricultural sector, mainly as pesticides and herbicides to destroy pests and weeds respectively. Three examples are *monosodium methylarsonate* ($NaMeHAsO_3$), *disodiummethylarsonate* (Na_2MeAsO_3) and *cacodylic acid* (also known as *dimethylarsenic acid*, $Me_2AsO(OH)$). Arsenic acid ($AsO(OH)_3$) is used for defoliation of cotton bolls and as a wood preservative whilst sodium arsenite is used in sheep and cattle dips. In recent years, however, their use is being decreased due to the overall toxic effects of the arsenic-containing compounds and more efficient replacements are being sought.

Though arsenic is toxic, it can be difficult to diagnose (Relfe, 2003). This is due to the majority of the arsenic ions leaving the body after around three days of exposure with only small amounts being stored in the brain, bones, organs and tissue, which accumulate over time. This can lead to a range of cancers and conditions such as diabetes. Exposure itself can lead to stomach cramps, vomiting, diarrhea, nausea and a decrease in the production of red and white blood cells. Inorganic arsenic can be absorbed through the gastrointestinal tract and be distributed to the majority of the vital organs including the skin (ATSDR (Agency for Toxic Substances and Disease Registry), 1989). Humans are more susceptible to arsenic poisoning than animals which only suffer a depression in the immune system.

Recent reports have shown that arsenic is becoming a huge problem in ground water especially in Third World countries like Vietnam, Nepal, India and Bangladesh, but also in the US (Pearce, 2003). For example, of 3000 tube wells measured in Bihar

(India), 40% of them contained arsenic levels in excess of the World Health Organisation (WHO) limit which is 0.01 mg per litre. However, the Bangladesh Standards for Testing Institution (BSTI, 1989) have set the limit to 0.05 mg per litre (World Health Organisation (WHO), 2000). Researchers in West Bengal have been reported to be working on a "homeopathic rememdy" for arsenic poisoning based on arsenic oxide which had seen the reduction of liver toxicity in mice after having been fed the antidote which they call *Arsenicum Album* (Bhattacharya, 2003).

3.3.1.2 Mercury (Hg)

The abundance in the Earth's crust is on average 0.08 ppm (Greenwood & Earnshaw, 1990). The main ore is *cinnabar* (HgS) and is converted to the liquid metal by heating in air at 600°C. The metal is volatile thus easily contaminates the atmosphere. The effects of mercury in the environment have already been explained such as the Minamata Bay disaster in Japan (Hosohara *et al.,* 1961). Mercury was also used to make the felt on hats smooth; however the workers were exposed to the mercury vapour and consequent damage to their central nervous systems resulted – a 'mad hatter'.

The most toxic form of mercury is methyl mercury, $MeHg^+$ and this was used in fungicides - the seed actually being coated with the $MeHg^+$ (Ferguson, 1990). However, when it breaks down, the mercury will get into the soils and hence into the food chain. $MeHg^+$ pesticides have now been banned. The decreasing order of toxicity of mercury salts are thus: Organic > Inorganic (divalent) > Inorganic (monovalent). Methyl mercury can enter the central nervous system rapidly leading to changes in behaviour and neural disorders (ATSDR (Agency for Toxic Substances and Disease Registry), 1989; Goyer, 1991). Ingesting the elemental mercury usually causes no ill-effects (Goldwater, 1972). Ingesting inorganic salts can lead to renal failure, gastrointestinal problems, and even death if the dose is up to 4g (Goyer, 1991). Inhalation of the vapour can irritate the respiratory tract, affect the kidneys, severely affect the central nervous system or death if huge amounts are inhaled (ATSDR (Agency for Toxic Substances and Disease Registry), 1989).

Mercury is used in the electrical instruments industry as a conducting medium and in highly efficient lights due to the vapour (Greenwood & Earnshaw, 1990). Mercury electrodes are used in the electrolysis of brine. Long term exposure to mercury also causes loosening of teeth. The main intake of mercury comes from fish (2.61μg). The methylation of mercury in rivers and lakes is important at low pHs due to more of the Hg^{2+} ions being extracted from sediments via ion exchange processes at high hydrogen ion concentrations. The low pH favours the formation of more methyl mercury since more of the Hg^{2+} is extracted.

Mercury has thus entered the environment mainly by human activities, from broken thermometers to pharmaceutical properties but sewage effluent is a major cause of pollution with up to 10 times the level of mercury that is found in natural waters (Manahan, 1994).

3.3.2 Preparation of Gold-Ink Screen-Printed Electrodes

The gold-ink screen-printed electrodes were printed by a third-party organisation (Du Pont (UK) Ltd., Bristol, UK) according to a design supplied by Cranfield University. The manufacturers prepared a number of batches of these novel and proprietary inks for the purposes of comparing different ink formulations and properties. Specific codes relating to the inks and substrates employed will be given when appropriate in Section 3.4. A description of the manufacture of the gold-ink screen-printed electrodes is given in Table 3.3 (Cooper, 2004). These inks were printed on to a polyester sheet with a *77T*, 45° mesh screen at the manufacturer's premises in Bristol (UK).

Table 3.3: Description of the manufacture of a batch of gold-ink screen-printed electrodes by Du Pont as employed by Cooper

Day	Procedure
1	The polyester sheets (*HT5*) of thickness 125μm are heat-stabilized. The three carbon tracks are printed using carbon-ink (7105) on the sheet in groups of eight. These are left to air-dry at room temperature overnight.
2	The silver/silver chloride ink (5874) is printed on the appropriate carbon track ensuring that the ink covers the track completely. This is the reference electrode. This is left to air-dry at room temperature overnight.
3	The gold ink (7106-E93335-92D) is applied to the working electrode in the same manner as the reference electrode. It is critical that the gold ink covers the working electrode completely and does not touch the reference electrode and the counter electrode. These are again left to air-dry overnight.
4	The clear insulating ink (20μm thickness) is applied to the sheets. The sheets are again left to air-dry overnight.
5	The sheets are then *cured* in a fan-assisted box-oven for 15 minutes at a temperature of 130°C.

Figure 3.1 shows a single polyester sheet with a group of eight gold-ink screen-printed electrodes.
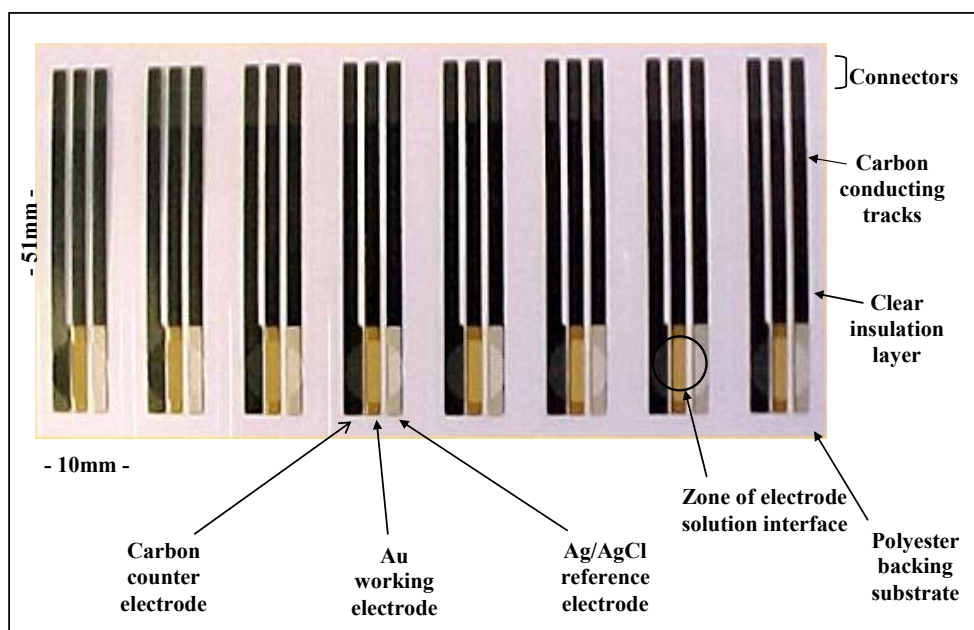
Figure 3.1: Gold-ink screen-printed electrodes. (Reprinted with kind permission from Cooper (Cooper, 2004))

Section 3.4.1 discusses two studies performed in this thesis in which a number of low-temperature curable ink formulations are compared.

3.3.3 Sample Preparation

All glassware had been soaked overnight in a 10% v/v solution of nitric acid. The glassware was rinsed with HPLC water (Fisher, UK) and dried in an oven for 1-2 hours at 80°C.

25% (w/v) solution of sodium hydroxide (NaOH) – (AnalaR, BDH, UK) – was prepared in HPLC water (Fisher, UK). This involved dissolving 25.51g of NaOH (Assay = 98%) in HPLC water and then making up to 100ml.

0.13218g of arsenic trioxide (As_2O_3) – (SPEXCertiprep, UK) – was dissolved in 1ml of the above 25% NaOH solution and then immediately acidified with 2ml of 32% v/v hydrochloric acid (HCl) – (Analytical Grade, Fisher, UK) – with a specific gravity (SG) of 1.16. The solution was made up to 100ml with HPLC water, resulting in a final concentration of 1000ppm.

0.2678g of hydrazinium chloride ($NH_2NH_2.2HCl$) – (Fisher, UK) – was dissolved in HPLC water and made up to 50ml resulting in a final concentration of 50mM. This was employed to prevent the oxidation of arsenic (III) to the electro-inactive arsenic (V) species (Sadana, 1983; Svancara *et al.,* 2002; Cooper, 2004).

The newly made up 1000ppm As (III) stock standard solution along with the commercially available stock standard solutions also at 1000ppm of Cu and Hg (BDH, UK) were diluted to 10ppm with HPLC grade water.

The molar concentration of the HCl was calculated to be 10.17M taking into account the specific gravity (SG), formula weight (~36.5 g/mol) and assay (32% v/v).

The respective metal standards were diluted to their appropriate concentrations (details given below) with 1.96ml of HCl and 100µl of hydrazinium chloride and

made up to a final volume of 5ml with HPLC water. This resulted in each sample containing 4M HCl and 1mM hydrazinium chloride.

Training Set

A randomised five level (0, 200, 400, 600, and 800 ppb), three factorial design was employed such that 125 samples were prepared. The design was generated by a custom-built in-house program (in the LabVIEW environment).

Validation Set

A randomised three level (300, 500, and 700 ppb), three factorial design was employed such that 27 samples were prepared. Again, the design was generated by a custom-built in-house program and incorporated into the training set design as shown in Figure 3.2.
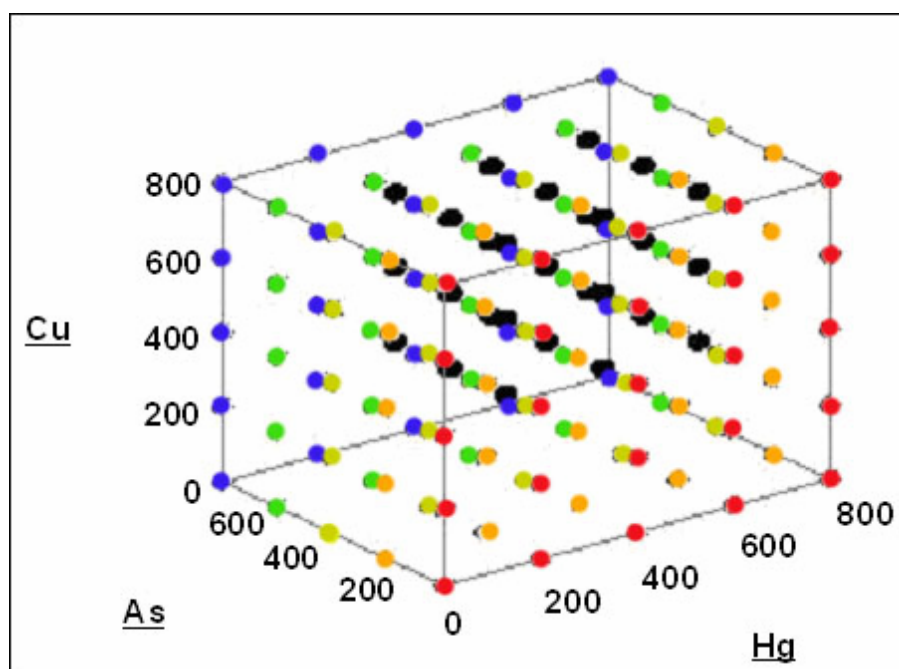
Figure 3.2: Overall experimental design: Training set (coloured with respect to As concentration) consists of five levels and three factors resulting in the preparation of 125 samples containing arsenic (As), mercury (Hg) and copper (Cu) at concentrations ranging from 0, 200, 400, 600 and 800 ppb; Validation set (●) consists of three levels and three factors resulting in the preparation of 27 samples containing As, Hg and Cu at concentrations ranging from 300, 500 and 700 ppb

Unknown Set

In addition to the training and validation sets detailed above, five "unknown" samples were prepared in the same manner in order to determine the strengths and efficiencies of the models generated. The concentrations of the respective metal ions are given in Table 3.4.

Table 3.4: Concentrations (in ppb) of the metal ions in the "unknown" samples

Sample	As	Hg	Cu
UNK1	600	400	200
UNK2	400	0	400
UNK3	0	600	0
UNK4	0	600	200
UNK5	500	200	200

3.3.4 Data Acquisition

The voltammograms were acquired on a multi-channel Autolab PGSTAT10 potentiostat (Eco Chemie, Holland) connected to a PC running under Windows 98 (Microsoft, USA), and controlled by the General Purpose Electrochemical Software (GPES) program (Version 4.6, Eco Chemie, Holland). A 4-way screen-printed electrode (SPE) connector was constructed in-house (by the author) in order to facilitate multi-electrode data acquisition. The instrumental set up is displayed in Figure 3.3.
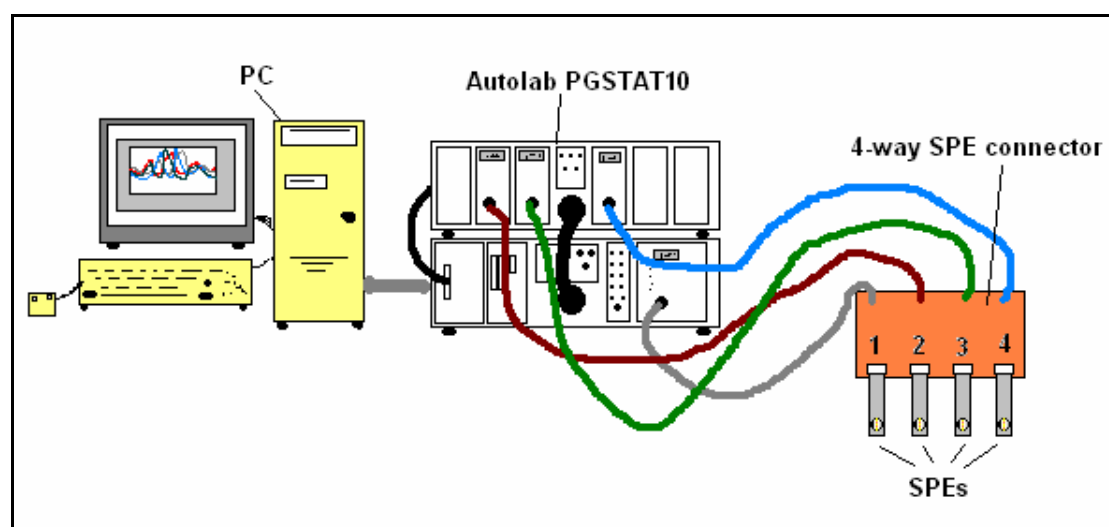
Figure 3.3: Instrumental set up of the multi-channel Autolab PGSTAT10 potentiostat

Due to only a limited number of electrodes having been supplied by Du Pont, only three measurements per sample were simultaneously performed. The electrodes were disposed of after each measurement. Table 3.5 details the experimental conditions employed during the acquisition.

Table 3.5: Instrumental parameters employed for the acquisition of the voltammograms

Parameter	Value
Starting Potential (V)	0.0
End Potential (V)	0.5
Deposition Potential (V)	0.0
Deposition Time (s)	30
Potential Pulse (mV)	50
Pulse Width (ms)	50
Period (ms)	500
Scan Rate (mV/s)	10

3.3.5 Data Processing

Data pre-treatment, modelling, validation and predictions were performed via the custom-built data analysis package described in Section 2.3.3. In summary, the package performs the following tasks:

- Import datasets consisting of DPASV voltammograms
- Perform a range of data pre-treatment techniques such as PCA, weighting, mean-centring, range-scaling and Savitzky-Golay smoothing
- Construct multivariate calibration regression models via one of a range of modelling algorithms (NIPALS, SIMPLS, PCR and ANNs)
- Perform leave-one-out cross validation and independent test set validation
- Predict the unknown concentrations of the target components in a given sample

The package can thus perform the same tasks as any given commercial product but allows the user more freedom in the manipulation of the data.

3.3.6 Preparation of the CRM Soil Sample Extract

The filtered soil extract from the certified reference material (CRM) sample employed in Section 2.4.2 was also analysed for arsenic (III). This involved transferring 1ml of the extract to a clean 15ml plastic centrifuge tube followed by the addition of:

- 1.96ml of concentrated hydrochloric acid (10.17M)
- 100μl of 50mM hydrazinium chloride
- HPLC water up to 5ml ensuring that the solution contained:
    - 4M HCl
    - 1mM hydrazinium chloride

Another two samples were prepared in the same manner but with equal amounts of arsenic (III) and copper (II) added. These two solutions contained an additional 400 and 800 ppb of both arsenic and copper. Voltammograms were thus acquired on an Autolab PSTAT10 potentiostat (Eco Chemie, Holland). Mercury (II) was not added due to the overlap with the copper (II) peak (Section 3.4).

3.3.7 Determination of Arsenic (V) on Gold-Ink Screen-Printed Electrodes

The samples were prepared in accordance with Cooper (Cooper, 2004). The objective is to reduce any electro-inactive arsenic (V) species in the CRM extract to the electro-active arsenic (III).

- A 20% solution of L-cysteine was prepared by dissolving 1g into 5ml HPLC grade water
- On dissolution, 1.25ml of concentrated hydrochloric acid (32%; 1.16SG) was added followed by dilution to 20ml with HPLC grade water.
- 0.1ml of this solution was added to the CRM samples prepared in Section 3.3.6 and heated in a water bath for ten minutes at 75°C
- Acquisition of the voltammograms was achieved on an Autolab PSTAT10 (Eco Chemie, Holland) under the following instrumental parameters:
    - Deposition potential: -0.3V
    - Deposition time: 30s

- o   Range: -0.3V to +0.5V

- o   Scan rate: 10 mV/s

- o   Interval: 0.5s

- o   Step potential: 5mV

- o   Pulse amplitude: 50mV

- o   Pulse Duration: 50ms

All potentials were recorded in triplicate (n = 3) unless otherwise stated and all stated potentials are relative to Ag/AgCl reference electrodes (+0.197V vs SHE).

3.4 Results and Discussion

Due to the novelty of the low-temperature curable ink material, a number of different batches of gold-ink screen-printed electrodes were fabricated and supplied by Du Pont. These batches differed from one another with respect to the formulation of the inks employed to the different curing methods. Section 3.4.1 discusses electrode composition and electrochemical performance. Section 3.4.2 will discuss the application of the multivariate calibration regression models to datasets consisting of voltammograms acquired on the gold-ink screen-printed electrodes for the quantitative determination of arsenic, mercury, and copper. Finally, Section 3.4.3 briefly discusses the determination of arsenic in the same CRM soil sample that was employed in Section 2.4.2.

3.4.1 Comparison of Select Batches of Gold-Ink Screen-Printed Electrodes

Due to propriety issues, the formulations of the actual inks employed in this study are not known by the author. Therefore any distinctions made between the various inks employed in this study are made via codes. Two sets of comparisons were performed. The first involved batches that varied slightly in their fabrication in relation to the batch employed by Cooper (Cooper, 2004) (Section 3.4.1.1). The second comparison involved batches that were fabricated with alternative ink formulations coupled to different curing techniques (Section 3.4.1.2).

3.4.1.1 Comparative Study 1: Comparison of gold electrodes with previous gold electrode study of Cooper

Four sets of gold-ink screen-printed electrodes were supplied by Du Pont in September 2004. Table 3.6 lists the inks employed in each set including the encapsulant which is the compound employed to agglomerate the gold nano-particles within the ink.

Table 3.6: Composition of the four batches of gold-ink screen-printed electrodes as represented by codes

Batch	Carbon	Ag/AgCl	Au	Encapsulant
A	7105	5874	BQ331	5036
B	7105	5874	BQ331	5036
C	7105	5874	BQ331	5018
D	E100735-155	5874	BQ331	E017257-1

An internal discussion within Du Pont led to the hypothesis that the solvent employed was interacting with the encapsulant (5036). The screen-printed electrodes (SPEs) were dried in an oven at 130°C. Batch A was prepared in the same manner as those employed by Cooper (Cooper, 2004) (Batch N). Batch B involved "a dwell time of 30 minutes" before the encapsulant was dried. In Batch C, a "UV cure dielectric" was employed (5018). Batch D employed an alternative and more active carbon along with a different encapsulant. Further details are not available due for propriety reasons.

For the purposes of this comparative study, individually prepared standard concentrations of arsenic (As), mercury (Hg) and copper (Cu) at concentrations of 0, 100, 200, 400, 600 and 800 ppb were employed. The arsenic standards included 1mM hydrazinium chloride whereas the mercury and copper standards did not. The corresponding voltammograms of the standards were recorded on all five batches (A, B, C, D and N). Good peak responses are observed for batches A, B, C and D with the exception of Batch N which displays a severe reduction in response for mercury and copper, but a flattened peak for arsenic (Figure 3.4). This implies that the Batch N electrodes have suffered degradation over a period of time. The same observation was evident when the mercury standard also included 0.1M KCl (not shown).
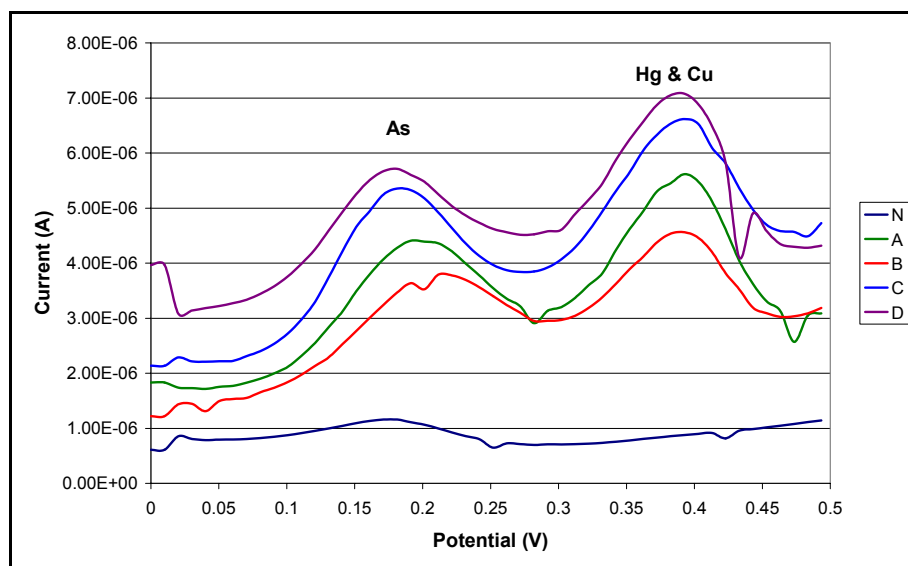
Figure 3.4: Voltammograms recorded for 600 ppb As, Hg, and Cu in 4M HCl, 1mM hydrazinium chloride. Deposition time: 30s; Deposition potential: 0.0V; Scan Rate: 20mV/s
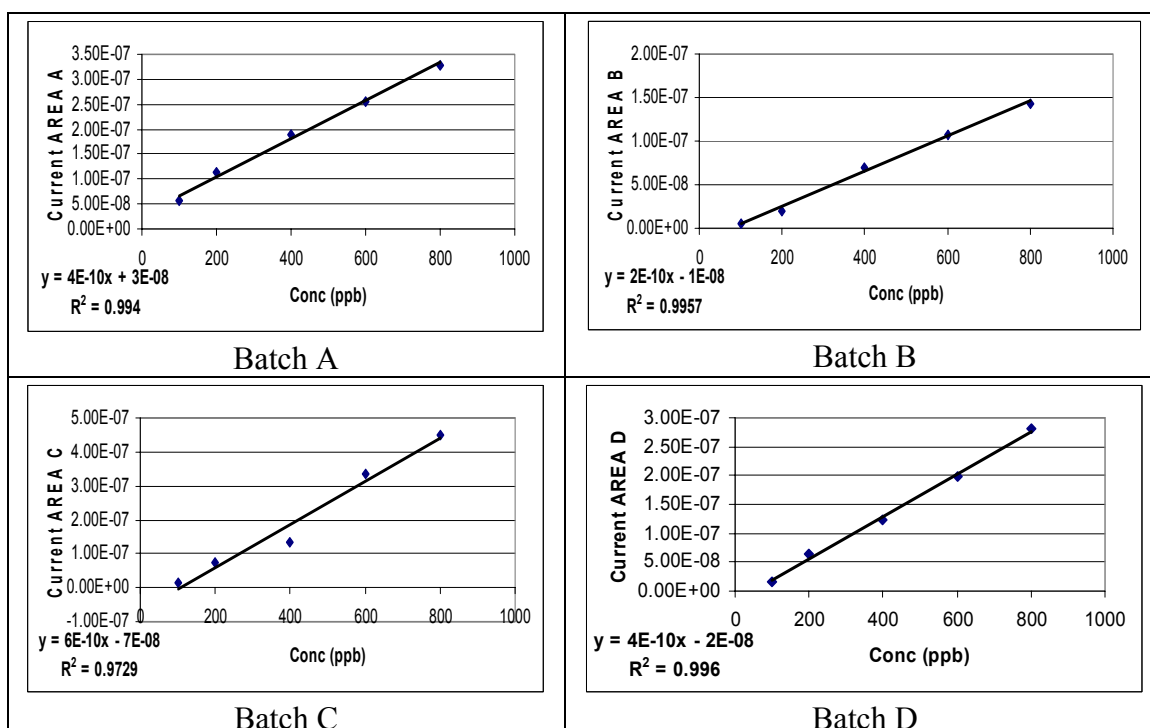
Figure 3.5: Calibration curves calculated from peak areas for As (III) on the four batches: A, B, C and D. Experimental parameters as in Figure 3.4

It was therefore not possible to calibrate mercury and copper on the Batch N electrodes. Figure 3.5 (above) illustrates the calibration curve for arsenic recorded on the other 4 electrode groups.

By referring to the calculated coefficient of determinations (R^2) for batches A, B, C and D in Figure 3.5, the best calibration has been obtained for the Batch D electrodes. As there were a limited number of electrodes in each batch for this study, the individual calibration curves for mercury and copper were not recorded. Instead, the metal ions were mixed in the same proportions. For example, standards of 0, 200, 400, 600 and 800ppb containing a mixture of arsenic (III) and mercury (II), and also arsenic (III) and copper (II) were prepared. Tables 3.7 and 3.8 summarise the correlation coefficients obtained on all four batches for the calibration curves of arsenic and mercury, and arsenic and copper respectively.

Table 3.7: Correlation coefficients for As (III) and Hg (II) for Batches A to D. JC refers to correlation coefficients obtained by Cooper (Cooper, 2004) under the same conditions. Experimental parameters as in Figure 3.4

Batch	As (III)	Hg (II)
A	0.9586	0.9806
B	0.9112	0.9778
C	0.7366	0.9472
D	0.9989	0.9957
JC (N)	0.9912	0.9942

Table 3.8: Correlation coefficients for As (III) and Cu (II) for Batches A to D. Experimental parameters as in Figure 3.4

Batch	As (III)	Cu (II)
A	0.9987	0.9021
B	0.9549	0.7515
C	0.8560	0.2369
D	0.9915	0.8645
JC (N)	Not measured	Not measured

In Table 3.7, the Batch D electrodes have produced the better response, even slightly better than those obtained by Cooper under the same conditions. In contrast, Batch C has produced the poorest correlation coefficients. In Table 3.8, the Batch A electrodes have produced the better correlation followed by Batch D. The Batch C electrodes have again performed poorly. Cooper had not acquired voltammograms for standards containing only arsenic (III) and copper (II) ions.

Cadmium (Cd) and lead (Pb) were also analysed on the four test batches in addition to Batch N. Cooper had reported the possibility of a stripping peak at -0.25V (Cooper,

2004). However, no distinct peaks were observed except for a significant peak at +0.05V which increased as the concentration of the cadmium and lead increased. Figure 3.6 displays the overlaid voltammograms obtained for a solution containing 400 ppb cadmium, lead, arsenic, mercury and copper in 4M HCl and 1mM hydrazinium chloride recorded on all but Batch B.
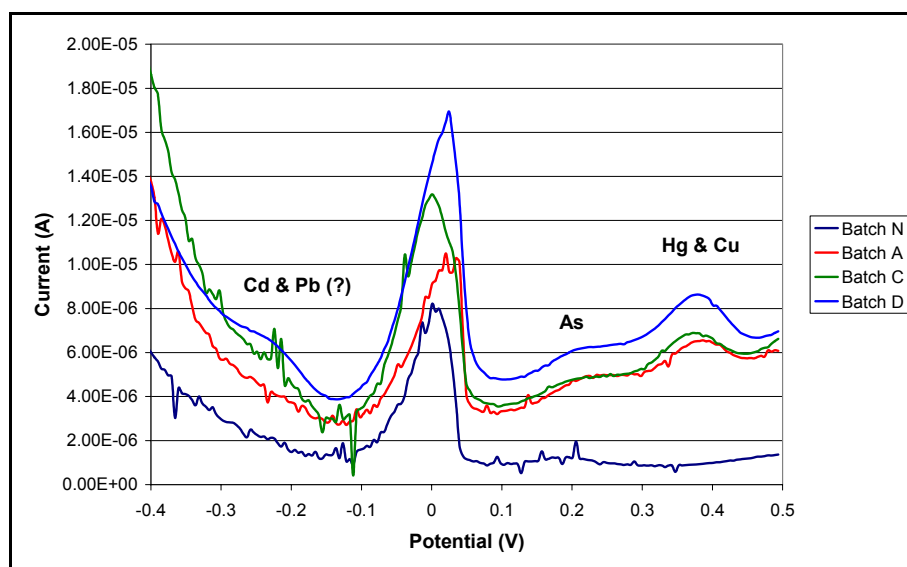
Figure 3.6: Voltammograms for 400 ppb Cd, Pb, As, Hg and Cu in 4M HCl and 1mM hydrazinium chloride recorded on electrodes of Batch N, A, C and D.
Deposition potential: -0.6V; Deposition time: 30s; Scan Rate: 10mV/s; Interval time: 0.5s

As can be seen in Figure 3.6, the Batch D electrodes have produced the better response, with a clear definition of the arsenic (III), mercury (II), and copper (II) peaks. Furthermore, the effects of noise are less pronounced in the Batch D voltammogram compared to the other batches. There is also a noticeable shoulder at ~-0.25V indicating peak activity which complies with the findings of Cooper. Further investigations however would be warranted to determine whether the peaks at +0.05V do pertain to cadmium (II) and/or lead (II), however this was not the scope of this work. Figure 3.7 compares the peak areas obtained for the five metal ions for Batches A and D.
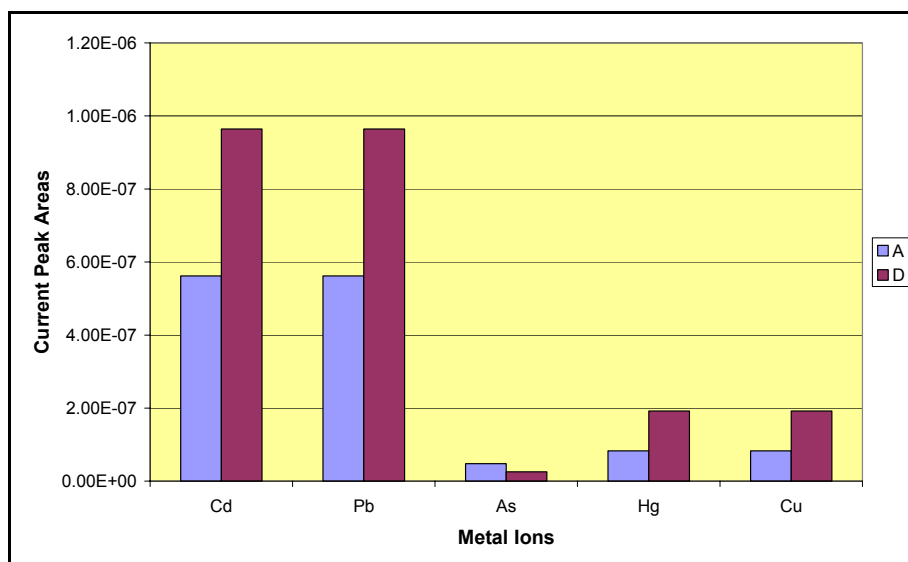
Figure 3.7: Comparison of peak areas calculated in Figure 3.6 for Batches A and D

Figure 3.7 illustrates that the response of Batch D is almost twice than that for Batch A for all the metals except for arsenic.

To conclude, the Batch D electrodes offer the best response followed by Batch A with Batch C the poorest. Unfortunately, this conclusion cannot be statistically supported due to only a limited number of electrodes (sensors) available to work with (~ 30). It has however been observed that there is a finite shelf-life as shown by Group N.

3.4.1.2 Comparative Study 2: Comparison of gold electrodes with different ink formulations to the previous gold electrode study

A major objective of the collaborative study between Cranfield University and Du Pont was to establish an improved low-temperature curable gold ink formulation. One aspect of the study was to ascertain the performance of the novel curable inks over time (12 months in the dark at ambient temperature of 20°C). Figure 3.8 compares the voltammogram acquired on a Batch D electrode of a standard containing 800 ppb arsenic (III) and mercury (II) and 200 ppb copper (II) with a voltammogram of the same standard acquired on a solid-state gold electrode.
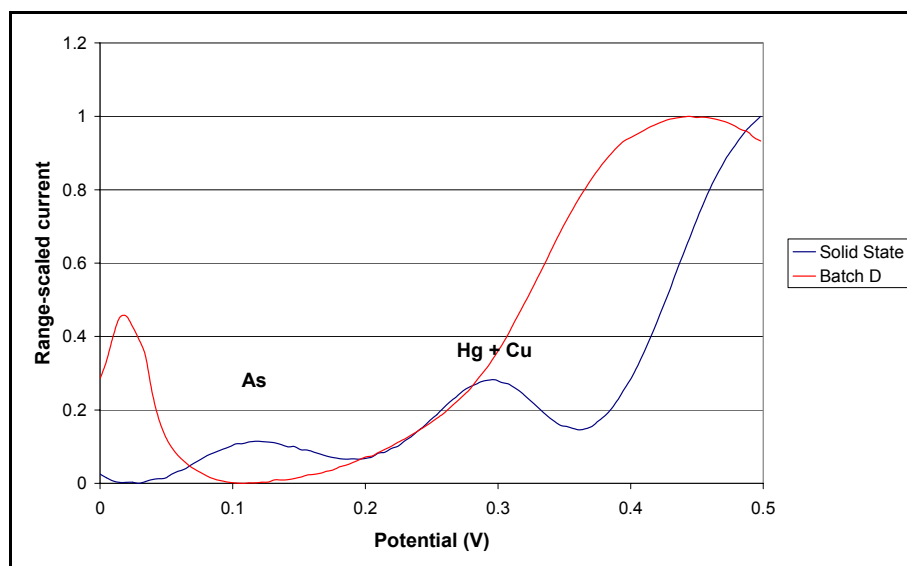
Figure 3.8: Voltammograms of 800 ppb As (III), 800 ppb Hg (II) and 200 ppb Cu (II) in 4M HCl and 1mM hydrazinium chloride acquired on solid-state gold electrode (blue) and Batch D screen-printed electrode (red) after a storage in the dark at 20°C for 12 months. (Deposition potential: 0.0V; Deposition Time: 30s; Scan Rate: 10mV/s).

There are no distinctive peaks observed on the Batch D screen-printed electrode in contrast to the solid-state electrode in which the presence of the three metal ions with a broad peak at ~ +0.15V for arsenic (III) and +0.32V for mercury and copper. This further supports the fact that further ink development by Du Pont would be warranted. Application of a conditioning potential of +1.0V for up to thirty seconds saw no improvements in responses (figure not shown). Some of the sheets were placed in the oven at differing lengths of time from 30 minutes up to 4 hours, the objective being to investigate if electrode performance had been compromised by the adsorption of components from the atmosphere to the gold-ink working electrodes, which could be removed by heating, but again no improvements were observed.

Another four batches of screen-printed electrodes were supplied by Du Pont (July 2005) with different ink formulations such that new carbon-ink, reference-ink and encapsulant formulations were employed; only the gold-ink formulation remained unchanged. Again, the exact formulations of these inks are not known by the author due to propriety issues. Table 3.9 details the composition of the new batches.

Table 3.9: Composition of the four new batches of gold-ink screen-printed electrodes. Codes refer to labels corresponding to the appropriate inks

Batch	Carbon	Ag/AgCl	Au	Encapsulant
1	BQ225	BQ164	BQ331	BQ425[1]
2[2]	BQ225	BQ164	BQ331	BQ425
3	BQ221	BQ164	BQ331	BQ425
4	BQ225	BQ164	BQ331	BQ411

The screen-printed electrodes were prepared using "HT5 grade PET" (poly(ethylene terephthalate)). A "static box oven" was used at 130°C. The same experimental procedures were employed as detailed in Section 3.4.1.1. Figure 3.9 displays the averages of the overlaid voltammograms for each respective standard recorded on Batches 1 to 4.
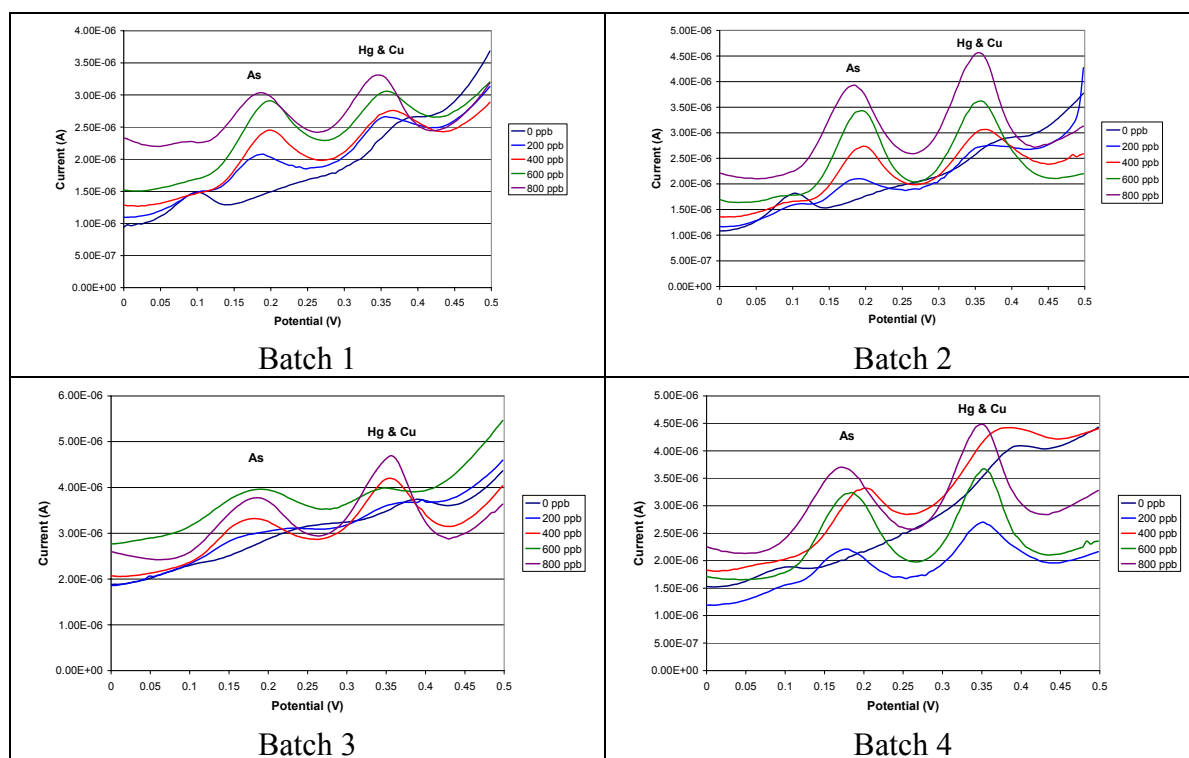
Figure 3.9: The averaged voltammograms acquired on Batches 1 to 4 corresponding to the standard concentrations of As, Hg and Cu (0, 200, 400, 600 and 800 ppb in 4M HCl and 1mM hydrazinium chloride). Experimental conditions as in Figure 3.8

It is uncertain as to what the flattened peaks at +0.1V and +0.4V pertain to, as observed in the voltammograms corresponding to the blanks (Bx-00, where x = 1, 2, 3

---

[1] UV-curing at 600mJ/cm^2
[2] Re-drying using a 1meter/min 130°C IR belt profile.

or 4). It is postulated that these impurities arise from the acid, the hydrazinium chloride or inks used in the electrode manufacture.

Table 3.10 shows the correlation coefficients acquired from the average of the peak areas obtained for each metal in the four batches.

Table 3.10: Correlation coefficients for the calibration curves recorded for each metal in each batch. Average of the calculated peak areas plotted against concentration.

Batch	As	Hg
1	0.8615	0.9598
2	0.9929	0.9703
3	0.9688	0.8180
4	0.9716	0.9737

As seen in Table 3.10, the better batches, based on the correlation coefficient data, appear to be 2 and 4. It was therefore concluded that Batch 2 offered the better response and thus should be employed for the quantitative determination of the three target elements. Based on the correlation coefficients listed in Table 3.10, Batch 4 could also be employed.

Four further batches of electrodes were thus prepared by Du Pont and supplied in November 2005. The electrodes had been prepared employing 125µm HT5 grade PET (poly(ethylene terephthalate). These in effect were split into two sub-sections in which one was dried in a box-oven at 130°C for 15 minutes (O) whilst the other was dried on an "IR combi belt drier at 130°C" moving at a rate of 1m/min (I). Table 3.11 lists the codes of the inks and encapsulants employed.

Table 3.11: The codes and methods of preparation of the batches of gold-ink screen-printed electrodes supplied by Du Pont in November 2005. Quantity refers to the number of sheets per batch each containing 8 screen-printed electrodes

ID	C	Ag/AgCl	Au	Cure	Encapsulation	Quantity
AO	BQ225	BQ164	BQ331	Oven	BQ411	30
AI	BQ225	BQ164	BQ331	IR	BQ411	75
BO	BQ242	BQ164	BQ331	Oven	BQ411	20
BI	BQ242	BQ164	BQ331	IR	BQ411	30

By comparing Table 3.11 with Table 3.9, the AO and AI batches above resemble Batch 4 in Table 3.9. The BO and BI batches have employed a different carbon ink but the same reference and gold inks in addition to the same encapsulant. As there were 75 sheets of eight electrodes, the AI batch was employed to acquire the voltammograms for the training, validation and unknown datasets (Section 3.4.2). However, by calculating and summing the standard errors of each point, it was found that Batch BI yielded the greatest repeatability (Figure 3.10).
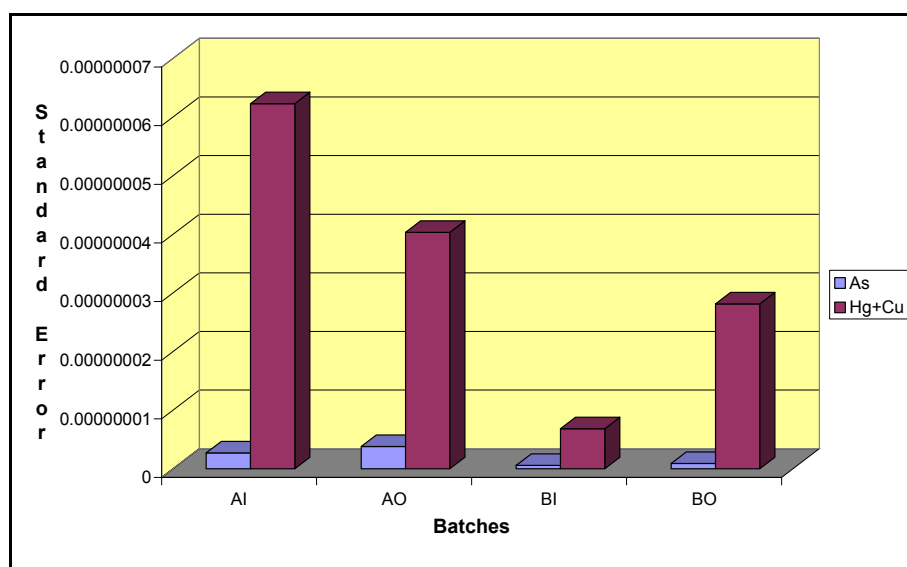
Figure 3.10: The calculated standard errors for each batch in terms of the areas of the respective metal ion peaks acquired from a standard containing 500 ppb As (III), Hg (II) and Cu (II) in 4M HCl and 1mM hydrazinium chloride (n = 3).

Deposition potential: 0.0V; Deposition time: 30s; Scan rate: 10mV/s

From Figure 3.10 the B batches are more reproducible than the A batches as supported by Table 3.12.

Table 3.12: Comparison of the relative standard deviations of the four electrode batches

BATCH	As (III)	Hg (II) and Cu (II)
AI	6.34%	47.39%
AO	15.47%	59.18%
BI	1.89%	18.60%
BO	2.63%	35.19%

Although the reproducibility for arsenic in the AI batch is very good (6.34%), that for mercury and copper was relatively poor in comparison (47.39%). Based on this

observation, and due to electrode availability the AI batch was employed for further studies.

3.4.2 Quantitative Determination of Arsenic, Mercury and Copper by Multivariate Calibration

Having identified the current optimum electrode manufacturing process, the aims of this study were to quantitatively determine the concentrations of arsenic (As), mercury (Hg) and copper (Cu) on gold-ink screen-printed electrodes via differential pulse anodic stripping voltammetry (DPASV) and multivariate calibration. The main objective had been to account for the overlapping mercury and copper peaks.

Similar to the findings in Section 2.4.1 the SIMPLS algorithm was identified to be the most efficient modelling approach compared against the NIPALS algorithm, PCR and ANNs. For example, the RMSEP for copper at the 7th latent variable was 81.22, 76.06, and 76.02 ppb for PCR, NIPALS and SIMPLS respectively after range-scaling and Savitzky-Golay smoothing. Figure 3.11 compares the RMSEP values for the quantitative determination of arsenic, mercury, and copper via models constructed via the SIMPLS and ANNs algorithms.

SIMPLS	ANN	PCA	PCs	RMS
	✔	✔	25	Cu 98.07; Hg 559.78; As 324.89
	✔	✔	20	Cu 97.15; Hg 561.48; As 317.59
	✔		N/A	Cu 91.54; Hg 149.54; As 211.75
✔			N/A	Cu 76.02; Hg 147.24; As 185.33
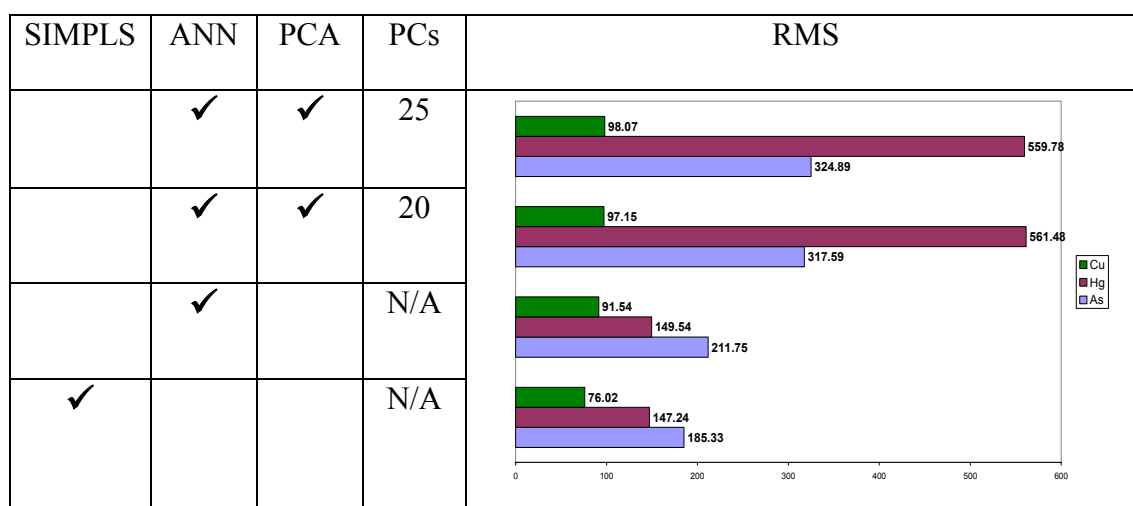
Figure 3.11: RMS error values for As, Hg and Cu recorded on a gold-ink SPE (Batch N). Dataset range-scaled and smoothed with Savitzky-Golay function (cubic polynomial and window size of 41). SIMPLS and ANN are the modelling algorithms; PCA denotes data reduction prior to ANN; PCs denotes the number of PCs retained. N/A denotes not applicable

The architecture of the neural network (Hagan *et al.,* 1996) was dependent upon the potential scan range, for example, 0.0 to 0.5V, and the number of metal ions that are present (98-15-3). There were thus 15 neurons in the hidden layer incorporating a sigmoidal transfer function. Three neurons are present in the output layer and a linear transfer function outputs the concentration values of the respective metal ions. The number of neurons was determined by the NNPLS1 algorithm contained with the PLS Toolbox (Wise & Gallagher, 1998b). The network was optimised via a back propagation algorithm which reduces the error between the predicted and required outputs by alteration of the weights of the links between the neurons within the network (Hagan *et al.,* 1996). As can be observed in Figure 3.11, application of the SIMPLS algorithm resulted in better predictions.

Direct modelling of the entire training set with the SIMPLS algorithm was not possible due to the presence of outliers. This was apparent when leave-one-out cross-validation was performed (Figure 3.12).
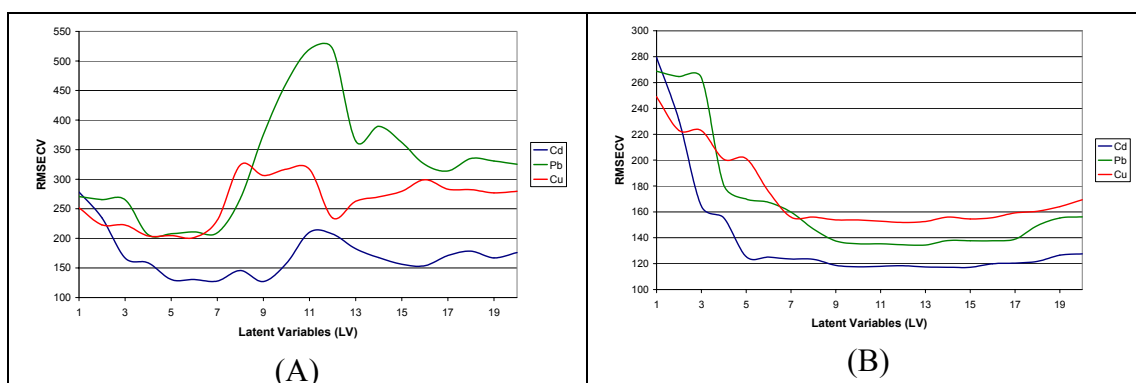
Figure 3.12: Application of leave-one-out cross-validation on the training set prior to removal of outliers (A) and after (B)

This was confirmed by observing the SIMPLS X-scores (Figure 3.13) where the presence of outliers is obvious. Confirmation is given by generally plotting the first latent variable of the X-score (T) against the first latent variable of the Y-score (U); this reveals the presence of strong outliers (Eriksson *et al.,* 2000).
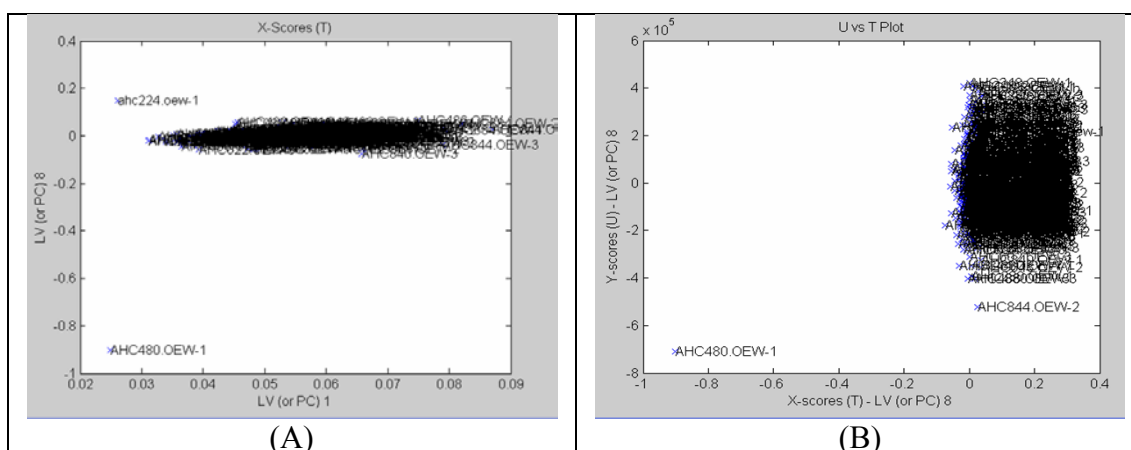
Figure 3.13: SIMPLS X-score plot (A) and U vs T plot (B) revealing the presence of outliers

After removal of the four outliers illustrated in Figure 3.13, another four were revealed and removed (not shown). Prior to performing independent test set validation, the validation set was subjected to the SIMPLS algorithm which revealed the presence of an outlier in the X-score (T) plot (not shown). The cause of the outliers was due to a faulty connection of the respective screen-printed electrode to the analytical system.

After removal of the outlier from the validation set, the set was integrated with the training set and then re-modelled (Martin *et al.,* 2005). Table 3.13 compares the variance captured at the first latent variable in both the X and Y block under different data pre-treatments. It shows that the application of mean-centring has resulted in fewer captured variance. This is attributable to the fact that mean-centring results in weighting of spectral regions if the peak magnitudes differ substantially, in this instance the increasing background current (Thissen *et al.,* 2004).

Table 3.13: Captured variance in the first latent variable by modelling with the SIMPLS algorithm after applications of various data pre-treatments

Pre-treatment	X-block	Y-block
None	84.93%	76.12%
Range-scaling	94.70%	71.03%
Mean-centring	69.55%	13.64%

With reference to Figure 3.12B the recommended numbers of latent variables to employ are 4, 5 or 6 for mercury, arsenic and copper respectively. However inclusion of more latent variables in to the model ensures that non-linear and inter-metallic

effects are also modelled (Thennadil & Martin, 2005). A number of models were constructed with the SIMPLS algorithm after relevant pre-treatment of the datasets. Figure 3.14 displays the appropriate RMSEP plots calculated from the prediction of the prepared "unknown" samples listed in Table 3.4.

Table 3.4: Concentrations (in ppb) of the metal ions in the "unknown" samples

Sample	As	Hg	Cu
UNK1	600	400	200
UNK2	400	0	400
UNK3	0	600	0
UNK4	0	600	200
UNK5	500	200	200

	MR	FT	L20	RMSEP
WTD			✔	196.96
		✔		127.20
	✔			196.96
M-C			✔	104.80
		✔		136.40
	✔			106.19
R-S			✔	103.37
		✔		129.54
	✔			98.50
RAW			✔	106.02
		✔		133.05
	✔			113.33
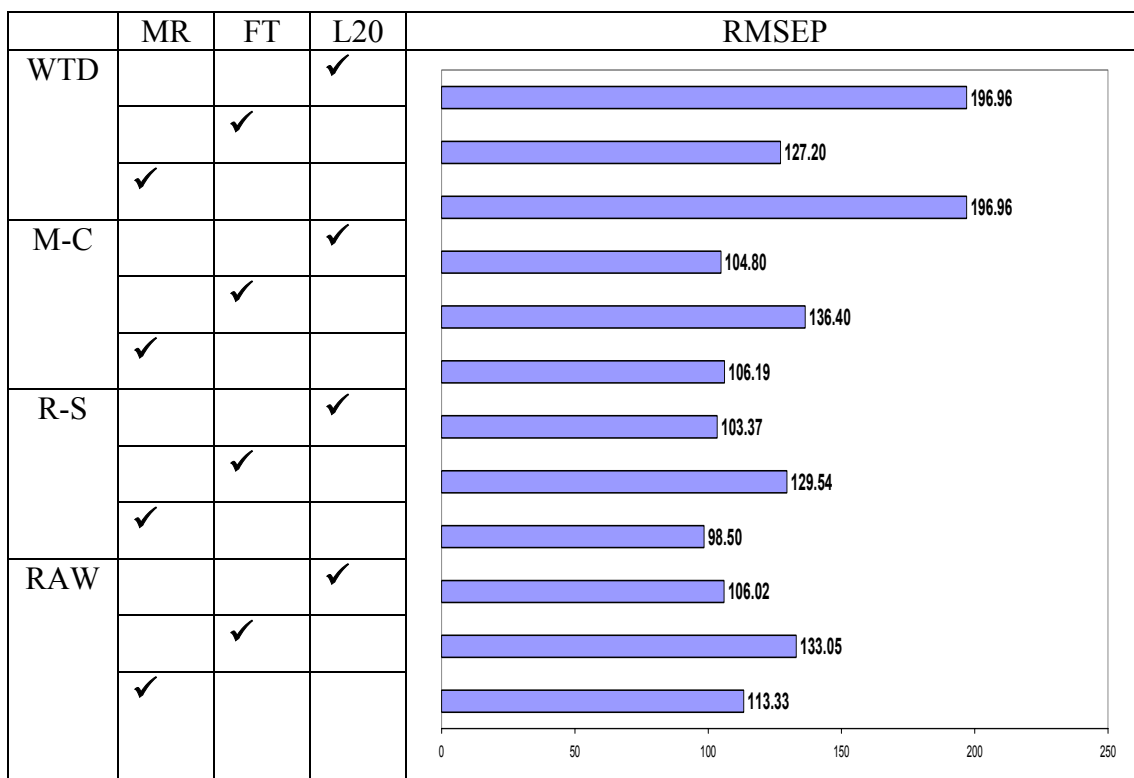
Figure 3.14: The RMSEP plots for the prediction of the "unknown" dataset in Table 3.4 with respect to the different data pre-treatments employed. Note that all appropriate outliers removed Key: RAW implies no scaling; R-S implies range-scaling and Savitzky-Golay smoothing (polynomial: 3; window: 21); M-C implies mean-centring and Savitzky-Golay smoothing (polynomial: 3; window: 21); WTD implies the application of weights (High: 5.0; Low: -2.0; Savitzky-Golay smoothing (polynomial: 3; window: 21); MR and FT implies predictions at the latent variables suggested by minimum RMSEP and F-test respectively; L20 implies predictions for all metals at the 20th latent variable

As can be seen in Figure 3.14, the range-scaled data (R-S) had produced the best predicted values with an overall RMSEP of just fewer than 99 ppb for the optimum latent variables suggested via the minimum RMSEP values for each respective metal ion. In most instances, the suggested optimum latent variables via the F-test (Section 1.5.5.2) have not proved efficient at predicting the unknown samples, except for the weighted variables. Further pre-treatment of the weighted datasets do not lead to an overall improvement of the predictability of the generated models (not shown). This can be deemed discouraging in respect of having attained lower RMSEP values in Section 2.4.1.4. Figure 3.15 displays the voltammogram of a standard containing 600 ppb arsenic, 800 ppb mercury and 600 ppb copper prior to and after the application of the weighting function (previously described in Section 2.4.1.3). Table 3.14 shows the index value ranges where the high weights were applied.
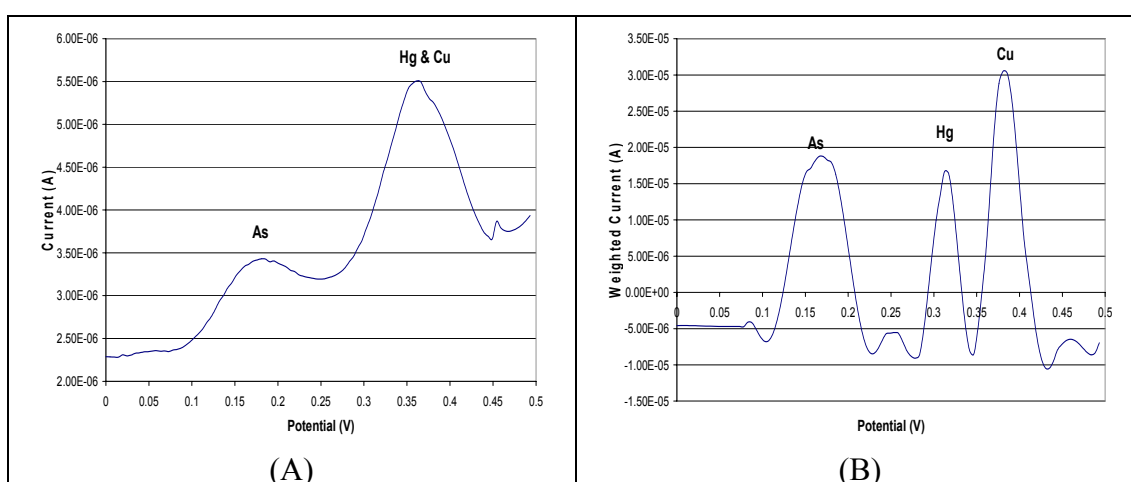
Figure 3.15: The effect of weights applied to the training dataset illustrated by a standard sample

(600 ppb As and Cu; 800 ppb Hg in 4M HCl and 1mM hydrazinium chloride).

Deposition potential: 0.0V; Deposition time: 30s; Scan rate: 10mV/s

Key: A: Raw; B: Weighted (High: 5.0; Low: -2; Savitzky-Golay (SG) polynomial: 5; window: 21)

Table 3.14: The ranges in which the high weight is applied to the respective metal ion

Metal (Au-SPE)	Range: Index (potential)
As	28 (0.151V) → 42 (0.220V)
Hg	63 (0.322V) → 68 (0.347V)
Cu	75 (0.381V) → 84 (0.425V)

In addition to obtaining three distinctive voltammetric peaks after the application of the weights, a linear baseline is obtained. This is due to the application of the low weight to the variables outside the ranges listed in Table 3.14. However, an

improvement in the RMSEP values was not attained. This warranted closer inspection
of the training dataset. In particular, voltammograms pertaining to samples containing
equal concentration values ranging from 0 to 800 ppb of the three metal analytes were
extracted and plotted (Figure 3.16).
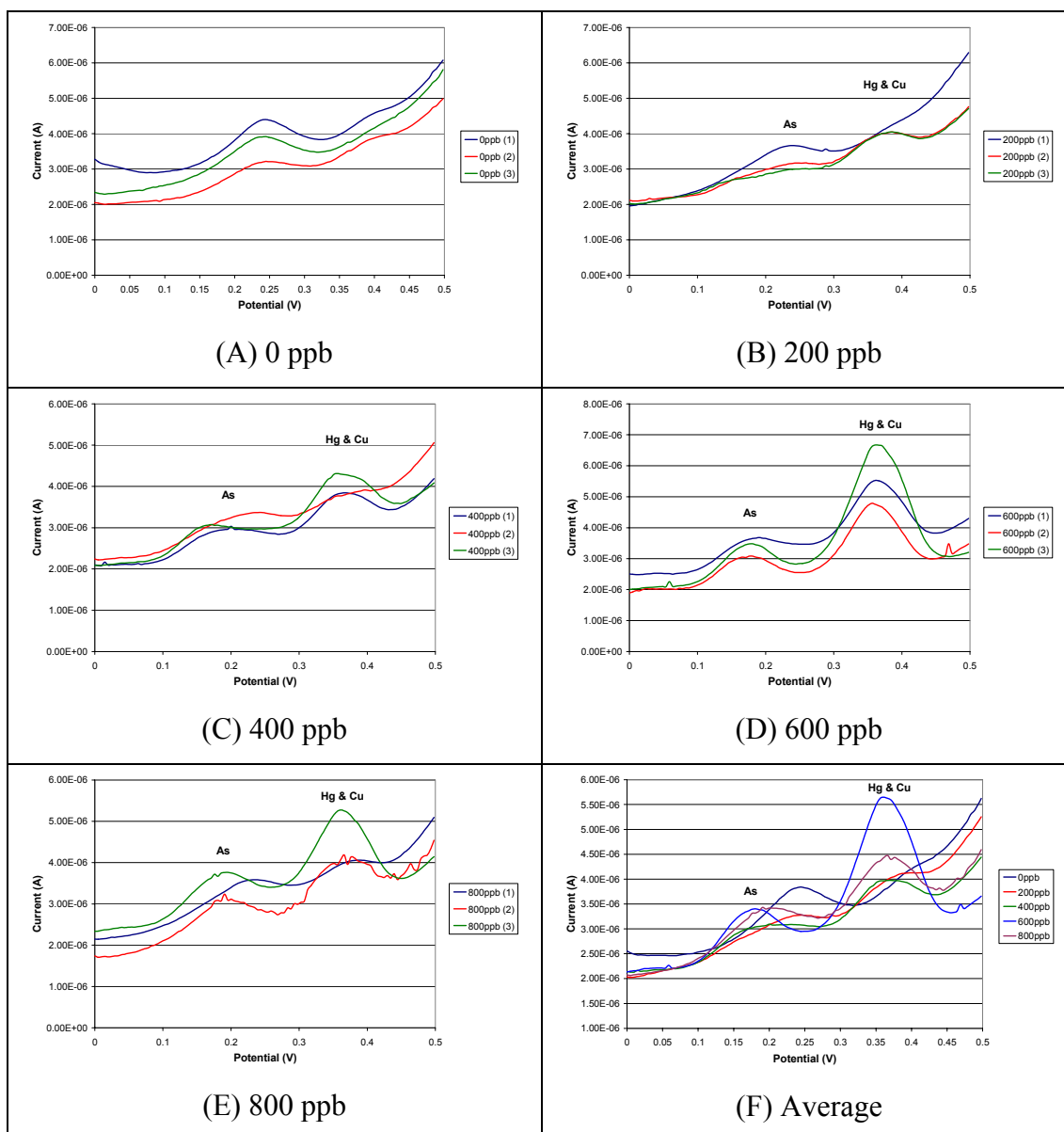
Figure 3.16: The raw voltammograms for equal concentrations of arsenic (III), mercury (II) and copper (II) in 4M HCl and 1mM hydrazinium chloride (A to E) and the average of the voltammograms (F).
Deposition potential: 0.0V; Deposition time: 30s; Scan rate: 10mV/s

A significant peak is evident at 600 ppb (Figure 3.16F) which leads to the postulation that these discrepancies in the data results in the high RMSEP values and will also make successful univariate calibrations difficult unless the sample is omitted. It had been observed in Figure 3.10 that a very high standard error had been observed for mercury and copper for the AI batch compared to the other electrode batches therefore problems could have been foreseen. Furthermore, the averaged voltammogram attributed to the blank (0 ppb) contains a broad peak at ~+0.25V which is clearly observed in Figure 3.16A. Again, as it was not known what the inks contained, it was therefore difficult to postulate as to the cause of this peak. However, it was appeared to be suppressed in the presence of metal ions such as arsenic. It is thus feasible that the model had failed to account for the blank due to the peak which resulted in the attainment of the high RMSEP values. Figure 3.16 has shown that there is significant variation between the individual gold-ink screen-printed electrodes especially in Figures 3.16D and 3.16E for 600 and 800 ppb respectively.

Another postulation is that at differing concentrations of mercury (II), the working electrode's response varies. As it is not known what comprises the inks, especially with regards to the gold, a reaction or series of reactions could be taking place which inhibits the successful deposition of the mercury ions on to the gold working electrode, a phenomenon that becomes more pronounced at higher concentrations (e.g. > 400 ppb). As the work here is novel, no other problems have been cited thus far in the literature. However attempts have been made by other researchers to create their own low-temperature curing gold inks (Huang *et al.,* 2003).

A calibration curve for the mixture of mercury and copper was acquired (not shown). The standard containing 600 ppb mercury (II) contributed to the very poor curve obtained (R^2 = 0.2134); omission of the standard led to a considerable improvement (R^2 = 0.9929). This further re-enforces the point that mercury interaction with the gold working electrode may well involve a number of complex phenomenae, particularly when higher concentrations of mercury are present (e.g. > 600 ppb). Future work should consider such further investigations.

Figure 3.14 illustrated that the RMSEP values were marginally better by 5 – 10 ppb for the range-scaling than the mean-centring. This can be attributed to better defined peaks observed in the loading plots (Figure 3.17).
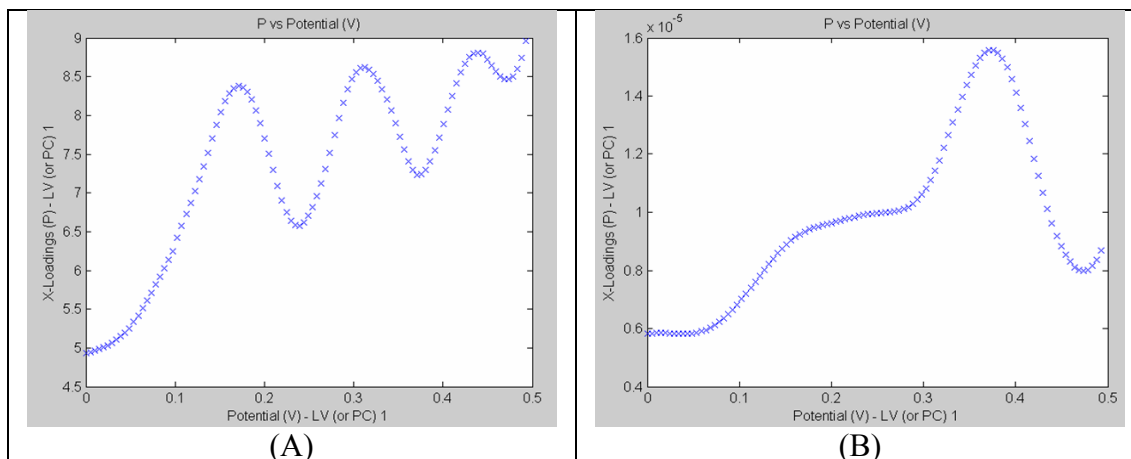
Figure 3.17: Comparison of the X-loadings (LV1) against the potential for SIMPLS models generated after data pre-treatment of the training set involving range-scaling (A) and mean-centring (B).

In the absence of a validation set, the custom-built data analysis package is capable of splitting the training set into a separate calibration set and a test set, as explained in Section 2.4.1.3. Table 3.15 shows the correlation coefficients obtained when the predicted concentration values were plotted against the actual values for a number of pre-treated datasets. The last ten samples in the training set were employed as a test set. Only the optimum latent variables suggested via the minimum RMSEP values were employed.

Table 3.15: Correlation coefficients for the plots of predicted concentration versus actual concentration with respect to the different pre-treatment techniques employed

Description of pre-treatment techniques employed	Correlation coefficient		
	As	Hg	Cu
Before weighting of variables	0.9619	0.8131	0.6348
Mean-centred[3]	0.9669	0.8220	0.6941
Range-scaled[3]	0.9697	0.8275	0.6273
After Weighting	0.9250	0.8572	0.8844
Weighting + Mean-centred[3]	0.9101	0.8563	0.9018
Weighting + Range-scaled[3]	0.9206	0.8410	0.9291

---

[3] And Savitzky-Golay smoothing with a polynomial of 3 and a window of 21

Table 3.15 has shown that application of the weights to the datasets in conjunction with range-scaling and Savitzky-Golay smoothing has indeed improved the predictive capability of the respective models especially for copper as we see the correlation coefficient rise from 0.6273 to 0.9291. However, a slight loss in correlation for the arsenic ion is observed after the weighting has been applied. In addition, prior to application of the weights, the range-scaling slightly improved the correlation for arsenic. Figure 3.18 compares the plots of predicted against actual concentrations for the range-scaling of the datasets without and with the application of weighting.

(A)

(B)
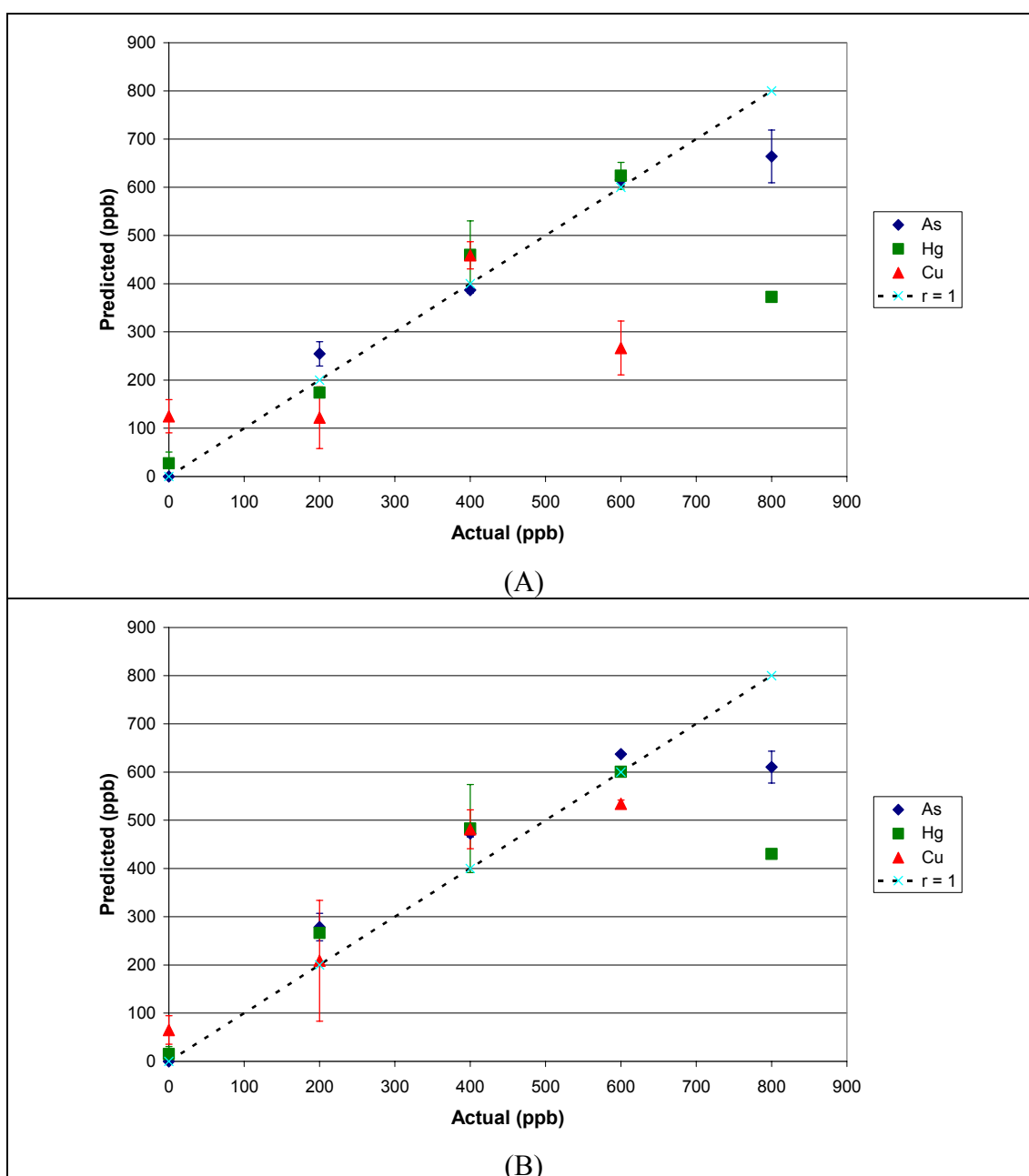
Figure 3.18: Predicted versus actual plots for As (III), Hg (II) and Cu (II) with range-scaling but no weighting (A) and range-scaling after weighting (B).

It is clear to see in Figure 3.18A that the prediction of copper (at 250 ppb) at the actual concentration of 600 ppb has contributed to the attainment of the correlation coefficient of 0.6273. Application of the weights has thus improved the prediction of the copper (520 ppb) at the actual concentration of 600 ppb; even with the prediction at 200 ppb being widely dispersed (Figure 3.16B).

3.4.3 Application to Real Samples

Standard addition, multivariate calibration and HG-AAS will be discussed in relation to the determination of arsenic in the CRM extract.

3.4.3.1 Standard Addition

Voltammograms were acquired on both the AI and AO batches of gold-ink screen-printed electrodes (Figure 3.19).
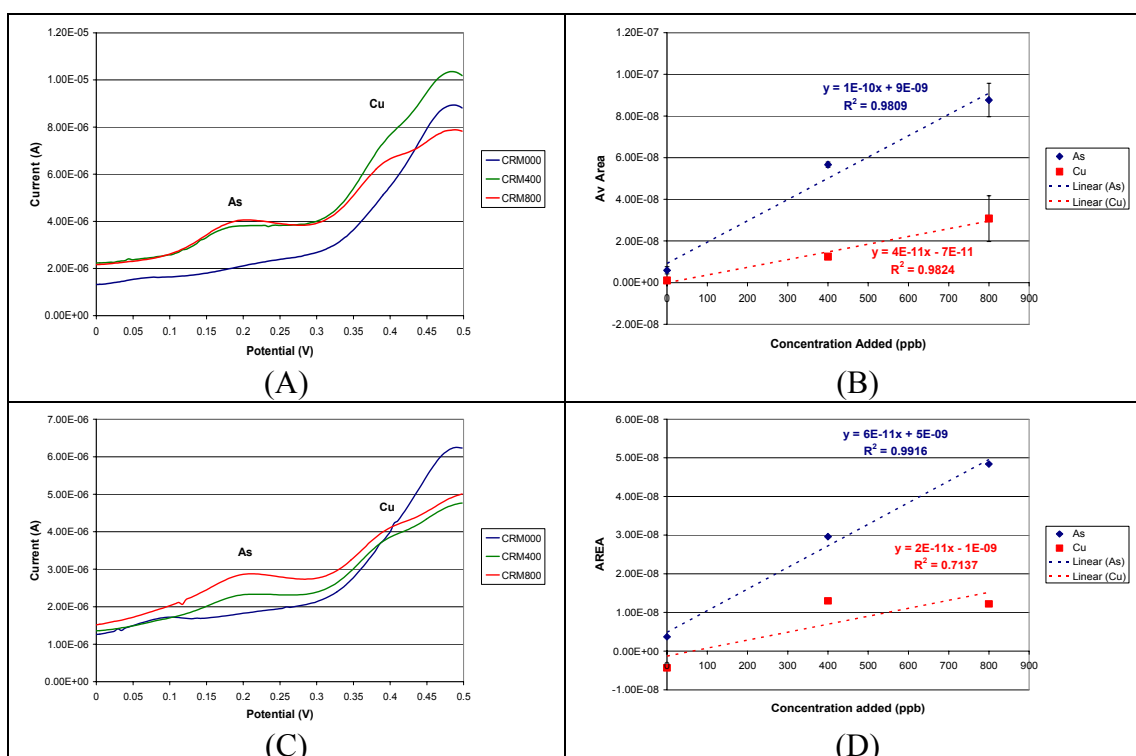
Figure 3.19: Overlaid voltammograms (A) and (C) and standard addition plots (B) and (D) acquired on gold-ink screen-printed electrodes. Batches: AI (A and B); AO (C and D). Instrumental conditions: deposition potential: 0.0V; deposition time: 30s; scan rate: 10 mV/s.

The voltammograms pertaining to the CRM extract only (CRM000) in Figure 3.19A and 3.19C for the AI and AO batches respectively closely resemble voltammograms pertaining to a blank solution, particularly as one can observe the characteristic flattened peak at ~ +0.1V (refer to Figure 3.16). This implies that there is no arsenic in the extract. With regards to the copper, it is difficult to quantitatively ascertain whether the copper was extracted due to the steep increase in current from +0.3V up to +0.5V. Copper was indeed extracted (Section 2.4.2) however it is plausible to suggest that coupled with other interfering ions in the CRM and the 4M HCl, the copper peak is suppressed. A number of proposed hypotheses are the formation of copper (I) chloride (Vvedenskii & Grushevskaya, 2003; Matsumoto *et al.,* 1994), and inter-metallic bonding with arsenic and any other trace metals present in the CRM (Sadana, 1983; Zima & van den Berg, 1994; Copeland *et al.,* 1974; Ben-Bassat & Azrad, 1978).

It must be noted that there is no error data reported in Figure 3.19D since only one measurement per sample was performed because of the limited availability of the AO batch. Addition of arsenic to the CRM extract leads to a positive response at +0.15 to +0.20V but not so for copper. This increase is more pronounced on the AO batch leading to a better coefficient of determination (R^2) value. However, the calibration curve for copper is much worse on the AO batch. When extrapolating to Y=0, it is clear to see from both Figures 3.19B and 3.19D that the predicted concentration of copper is zero which would imply that copper had not been extracted from the CRM.

According to the standard addition, the procedure recovered ~90% of the arsenic present in the CRM for both the AI and AO batches. The true amount of arsenic in the CRM was 5.64 mg/kg. Other methods do exist for extracting arsenic from soils (not discussed here) but the only way to confirm whether any arsenic is present in the CRM solution is to measure it on a standard laboratory instrument such as ICP-MS.

3.4.3.2 Multivariate Calibration

Applying the SIMPLS model constructed in Section 3.4.2, in conjunction with range-scaling and Savitzky-Golay smoothing with a polynomial of 3 and a window size of 21, indicates that there is no arsenic and copper present in the CRM. This does not

concur with the values supplied with the CRM (5.64 mg/kg arsenic and 18.8mg/kg copper respectively). Values of 0 ppb (and thus 0 mg/kg) were attained throughout. The same was observed when the SIMPLS model originating from un-scaled data (after outlier removal) is applied. This implies that the method of standard addition is unreliable here possibly due to the matrix effects and analyte interactions in the CRM. This suggests that the extraction method is inefficient with respect to arsenic extraction. However there are too many variables to consider at present, and one must be aware that the relatively high RMSEP values had been observed in Section 3.4.2.

3.4.3.3 Hydride Generation AAS (HG-AAS)

The CRM extract employed in Section 2.4.2 was diluted ten-fold (0.5ml → 5ml). Standards were prepared and recorded on the AAS also employed in Section 2.4.2 in conjunction with the hydride generation (HG) unit. Table 3.16 details the hardware parameters whilst Table 3.17 lists the parameters used for arsenic determination.

Table 3.16: Hardware and computational parameters for the HG-AAS instrument

Atomiser	Standard Burner	Delay Time (s)	25
Measurement Mode	Working Curve	Signal Record	Direct
Flame	Acetylene-Air (2125 - 2400°C	Signal Mode	Background Correct
Oxidant Pressure (KPa)	160	Replicates	Std = 2 Sample = 1
Oxidant Flow (L/min)	15	Calculation	Peak Height
Calculation Time (s)	5	Slicing Height	10% (P. W. Only)

Table 3.17: Parameters for arsenic as employed on the HG-AAS

Element	Current (mA)	Wavelength (nm)	Slit (nm)	Fuel Flow (L/min)	Burner Height (mm)
As	10	193.7	1.3	1.5	7.5

As seen in Table 3.16, the standards were measured in duplicate whilst the sample was measured only once due to being limited in volume. Figure 3.20 displays the calibration curve of atomic absorbance at increasing concentrations of arsenic.

y = 0.0006x + 0.0025
R^2 = 0.9987
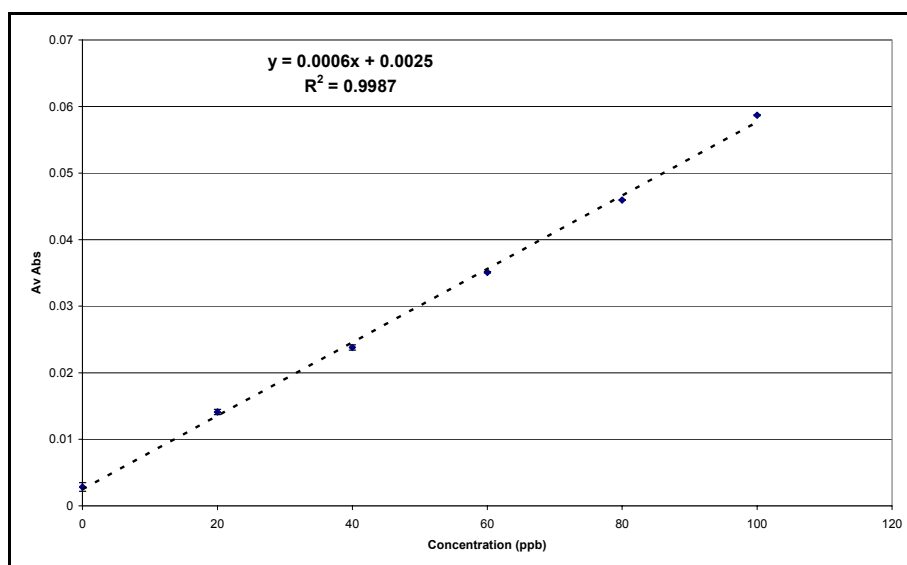
Figure 3.20: Calibration curve for the detection of arsenic via HG-AAS. Instrumental parameters described in Tables 3.16 and 3.17

An excellent correlation was thus attained. The % recovery of the CRM extract was calculated to be 11.76%. The value of 11.76% implies that arsenic was extracted from the CRM soil sample. HG-AAS is classed as a highly sensitive technique. However, the previous section stated that the multivariate calibration regression models predicted no arsenic present. This implies that the gold-ink working electrodes were not sensitive enough to detect the arsenic. However, one can argue that arsenic was only present in the CRM extract as arsenic (V) which would therefore result in no peak attained (Figure 3.19) at the experimental conditions employed for DPASV (Table 3.5) which is the optimum for the detection of arsenic (III). This implies that the procedure described in Section 3.3.6 which was developed by Svancara (Svancara *et al.,* 2002) must be employed to reduce any arsenic (V) species to arsenic (III). Section 3.4.3.4 discusses the findings.

3.4.3.4 Arsenic (V) on Gold-Ink Screen-Printed Electrodes

A qualitative determination of arsenic (V) was performed on both the AO and AI batches of gold-ink screen-printed electrodes. Figure 3.21 displays the overlaid voltammograms acquired on both batches.
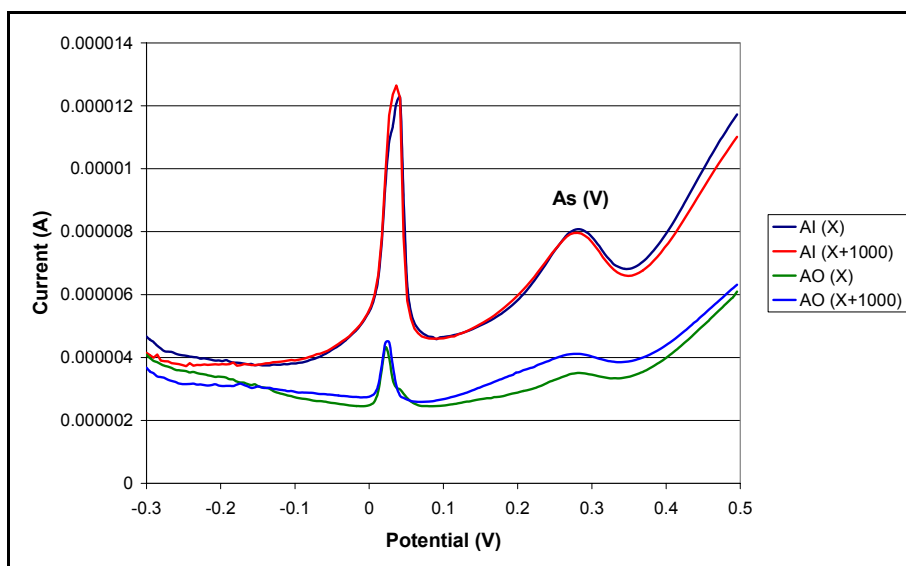
Figure 3.21: Overlaid voltammograms for the CRM extract for the determination of arsenic (V) on both the AI and AO batches of gold-ink screen-printed electrodes.
Key: X refers to the unknown concentration of As (V) in the CRM; X + 1000 refers to the addition of 1000 ppb As (V) standard solution.
Instrumental parameters: Deposition potential: -0.3V; Deposition time: 30s; Scan rate: 10mV/s

The profiles for the AI batch in Figure 3.21 overlay exactly over one another implying that the addition of the arsenic (V) standard has not been detected. In contrast a response has been detected on the AO batch. This is reflected in Figure 3.22.
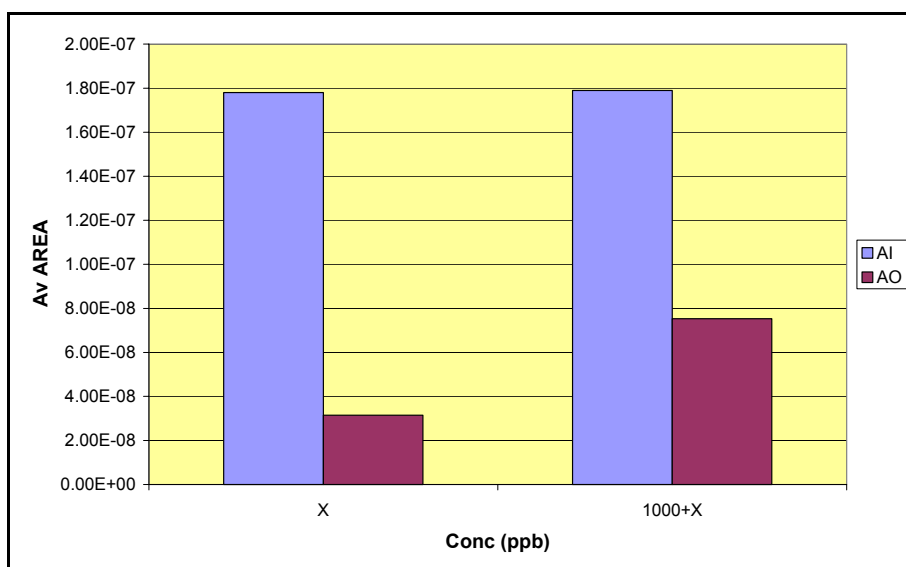
Figure 3.22: Average Areas calculated on both batches for the determination of arsenic (V) at +0.40V for the CRM extract. Instrumental conditions as in Figure 3.21

The AO batch is distinctly more responsive to the presence of the reduced arsenic addition than the AI batch. However, there are a number of parameters to consider. There are the presence of other metals in the CRM which will have also been extracted, in particular copper which we know to form inter-metallic complexes with arsenic and many other metals such as nickel and zinc (Herrero & Ortiz, 1997). In this respect, it is possible that the arsenic (III) generated by the reduction with the L-cysteine has formed an inter-metallic bond with copper ions present in the CRM. Cooper had also observed one broad peak for As (V) at ~+0.3V (Cooper, 2004). Furthermore, it is possible that not all of the arsenic (V) is reduced to arsenic (III) during the reduction step.

Figure 3.23 displays the PCA score and loadings plot for the voltammograms acquired on the AI and AO batches.
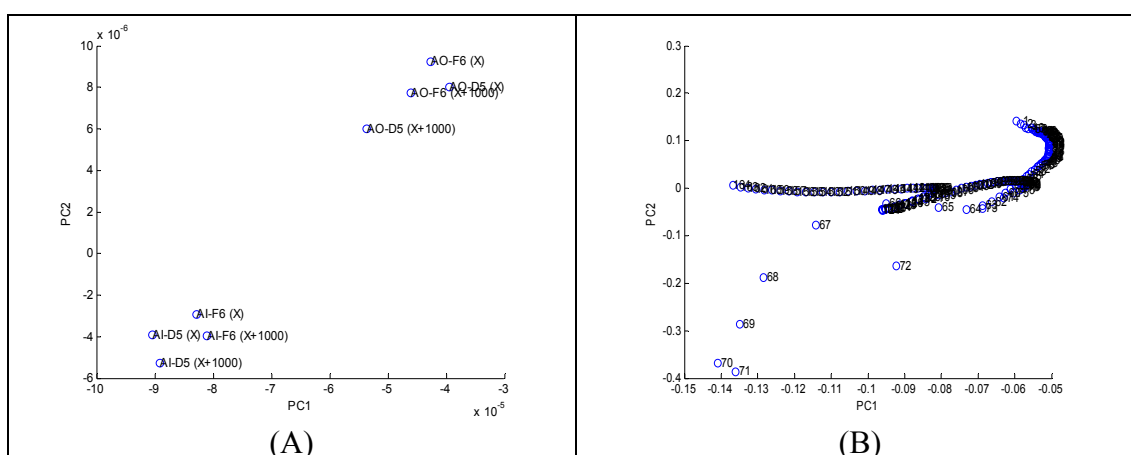
Figure 3.23: PCA scores (A) and loadings (B) of the voltammograms acquired on both the AI and AO batches of gold-ink screen-printed electrodes

As can be seen the AI and AO batches have been clearly classified into separate groups (Figure 3.23A). The corresponding loadings plot indicates that the sharp peak at 0.0V leads to the distinction between the two batches (Figure 3.23B). The PCA loadings plot illustrates that the greatest variance is observed in PC2. The intense peak at 0.0V can be observed in Figure 3.21 and can possibly be attributed to the chloride ions in the electrolyte equilibrating with the chloride ions on the reference electrode, or reacting with the gold nanoparticles in the ink formulation since no peak was observed when a sulphuric acid electrolyte was employed (Cooper, 2004). By looking

at Table 3.11 the only difference between the two batches is the curing employed (box oven for AO and IR belt for AI).

In short, due to the presence of the other metal ions in the CRM, in particular copper, it will therefore be difficult to quantify the presence of arsenic in the CRM. The solution to this is to employ specific masking agents that will complex, for example with copper and prevent it from interfering with the deposition of arsenic on to the working electrode. Examples of masking agents for copper are 1,10-phenanthroline, picolinic acid, and L-cysteine which has been shown to stop the complete reduction of arsenic (V) to arsenic (III) due to the masking of the copper ions present (Ek & Hulden, 1987; de Campos *et al.,* 2002). A typical masking agent for zinc is sodium or potassium cyanide (Skoog *et al.,* 1996).

3.5 Conclusions

The comparative studies detailed in Section 3.4.1 in conjunction with the application of one of the batches (AI) in Section 3.4.2 indicate that further research is still required by Du Pont in producing more stable, robust and reproducible low-temperature curable inks for their screen-printed electrodes. In addition, these electrodes should possess longer shelf-lives, just like their carbon counterparts. The most promising batches tested were Batch 2 in Section 3.4.1.2 closely followed by Batch 4. The batches employed in the quantitative study were derived from Batch 4, due to electrode supply issues, even though Batch 2 yielded slightly superior voltammograms in terms of peak areas and repeatability.

The most efficient pre-treatment technique has been range-scaling coupled with Savitzky-Golay smoothing with a polynomial of 3 and a window size of 21. However, the current SIMPLS models that were generated are deemed inadequate and inefficient to successfully predict the concentrations of arsenic, mercury and copper in unknown samples. Complete separation of the mercury and copper peaks has not yet been achieved. More work is thus required in varying the weights in order to optimise the parameters and thus to lower the RMSEP values, for at this instance they are too high (ranging from 98 to 130 ppb for range-scaled and Savitzky-Golay smoothed data).

The extraction procedure described in Section 2.3.4 has not been successful in relation to the determination of arsenic on a gold-ink screen-printed electrode under the same conditions. This can be attributed to a number of causes such as the irreproducibility of the gold-ink electrodes themselves, the complex characteristics of the arsenic ion (speciation), and the other metallic ions present in the given CRM. Though standard addition gave a recovery of ~90% the multivariate approach predicted 0% analyte present. Observation of the respective voltammograms leads one to conclude that the standard addition method is inappropriate in this regard due to the presence of the other metal analyte ions. Furthermore, additional concentrations should have been employed in addition to 400 and 800 ppb.

HG-AAS in contrast has shown that arsenic was indeed extracted from the CRM soil sample. However, judging by the % recovery (11.76%) the process described in Section 2.3.4 was not efficient for arsenic. DPASV measurements performed on two batches of gold-ink screen-printed electrodes for the qualitative determination of arsenic (V) have indeed shown that arsenic was extracted from the CRM soil sample. The AO batch has been shown to be more responsive than the AI batch and a PCA score plot clearly distinguished between the two batches mainly due to a significant voltammetric peak at 0.0V.

In relation to the objectives of the ARTDEMO project the work described in this chapter can lead to the development of a rapid screening tool for the determination of metal species – thus a peak at a certain potential suggests the presence of copper, mercury and/or arsenic present which necessitates re-running the sample on carbon-ink screen-printed electrodes where arsenic will not interfere with copper. This could be done simultaneously on the multiplex potentiostat developed in-house (Section 4.4 and Appendix A5). Therefore these electrodes are useful for site characterising and mapping i.e. field based tool that enables better field-based management of a site.

Electrochemical methods offer a more cost-effective solution to the determination of metal ions in a sample. Anodic and cathodic stripping voltammetry (ASV and CSV respectively) are heavily employed in this field. However, certain procedures involve the use of mercury-film electrodes and it is therefore desirable to find a safer and more efficient means of determining the metal ions in the absence of mercury. This chapter has thus illustrated the potential of applying multivariate calibration techniques to voltammograms acquired on gold-ink screen-printed electrodes.

Furthermore, the application to field analysis necessitates the need to further develop analytical techniques in portable devices leading to the screening and quantification of target analytes. Chapter 4 presents such a device.

CHAPTER 4:

DEVELOPMENT OF A PORTABLE FIELD-DEVICE INCORPORATING MULTIVARIATE CALIBRATION REGRESSION MODELS

4.1 Overview

This chapter discusses the advances made in the development of portable field devices and its importance as applied to on-line and at-line measurements. Electro-analytical flow systems will be discussed with the focus being on the on-line detection of heavy metals. The chemometric tool of instrument standardisation will be introduced with examples of its importance in multivariate data analysis. The development of an in-house custom-built personal digital assistant (PDA) program designed to process multivariate data will be presented in conjunction with acquisition of DPASV voltammograms on carbon-ink and gold-ink screen-printed electrodes via a custom-built multi-channel controller. Finally, preliminary work on the development of an automation system is also presented.

4.2 Introduction

Chapter 2 discussed the application of multivariate calibration to DPASV voltammograms acquired on carbon-ink screen-printed electrodes, whilst Chapter 3 discussed the same but on voltammograms acquired on gold-ink screen-printed electrodes. In each case, the voltammograms were acquired on laboratory-based instruments which require a dedicated bench space, a high performance personal computer containing the software application to control the instrument in addition to other applications for data manipulation, a qualified technician to operate them and power from a 220-240V alternating current (AC) mains input.

Recent years have seen the development of so called "laboratory-on-a-chip", the onus being to fabricate portable devices for application in the field in which the major components of the main sensor are miniaturised on to a chip. An ideal example is the

manufacture of on-line sensors which acquire data in real-time and transmits it to a central computer where it is manipulated prior to storage on a dedicated database which has been one of the major objectives of the ARTDEMO project. If the aim is to determine the quality of river water at a given site, these sensors can thus be in the form of flow cells (electro-analytical) or electronic tongues (Winquist *et al.,* 2005).

If conditions prevent the acquisition of on-line measurements such as the harshness of the environment or the need to chemically pre-treat the sample, then at-line measurements are preferred. These involve the development of portable field instruments in which all the major components are contained within a small purpose-built device, which is powered, for example by a 9V direct current (DC) battery. The device itself can be stand-alone or controlled by a personal digital assistant (PDA). The PDA will contain the appropriate software application to acquire, manipulate and store the data. Furthermore, should any abnormal levels of target analytes be observed, certified action schemes can be put into immediate effect, as has been adopted by one of the ARTDEMO project partners: *The Gothenburg Region Association of Local Authorities* (GR), Gothenburg, Sweden (Lundh *et al.,* 2003).

The following sections will discuss in greater detail the application of electro-analytical flow cells leading to the benefits of miniaturisation in addition to the importance of instrument standardisation which can relate the measurements performed on a field-based instrument to measurements performed on a laboratory-based instrument.

4.2.1 Electro-analytical Flow Systems

Recent advances have been reviewed, spanning the last several years, in the field of electro-analytical flow measurements, emphasising the important role it will play in the development of, for example, miniaturised measuring devices, mechanisation and automation of analytical processes (Trojanowicz *et al.,* 2003). The need to reduce costs, and at the same time attain better detection levels, is paramount. These devices will be beneficial for on-line or at-line monitoring of ground water sites and rivers especially with regards to the ARTDEMO project.

4.2.1.1 Flow Injection Systems

The concentration of lead in blood by *hydrodynamic* voltammetry in a flow injection system with a wall-jet detector has been reported (Jaenicke *et al.,* 1998). The wall-jet detector had to be configured in such a way that there was no breakage in the jet of electrolyte striking the working electrode; the jet diameter was negligible compared with the diameter of the electrode, and that the reference and counter electrodes were sufficiently remote from the working electrode so that the boundary layer was not disturbed. Emphasising the long-term effects of lead in the environment and body, the authors still utilised a mercury film electrode (MFE). However, the carbon surface was modified with a Nafion®- membrane (polytetrafluoroethylenesulphonate) in order to prevent proteins present in the blood from fouling the electrode. The MFE offered a higher surface-to-volume ratio compared to the mercury drop which resulted in a faster diffusion of ions to the surface of the electrode and thus sharper stripping peaks. The authors concluded that the fabrication of disposable screen-printed electrodes could be used in portable instruments, eventually leading to the preparation of "a solid matrix blood-lead sensor".

As well as applications to the medical sector, such as the determination of urea in human serum samples (Walcerz *et al.,* 1998), flow analysis coupled with voltammetric detection has been suggested as a tool for on-line monitoring in the environment and industrial processes (Trojanowicz *et al.,* 2003). Furthermore, the use of dual electrodes or even four parallel electrodes were discussed, the latter being polarised at different potentials in order to determine the isomers of *resveratrol* which is a natural phenolic compound in wine and grape juice (Zhu *et al.,* 2000). Detection at four potentials allows peak purity to be attained when dealing with complex matrices due to "comparing ratios at different energies for both standards and samples". It was also reported that metallic copper was employed as a working electrode in capillary electrophoresis (CE) voltammetry for the determination of amino acids (Trojanowicz *et al.,* 2003). This leads to the development of portable integrated microchips "for application in the field".

DPASV can be combined with a flow system to determine the concentration of heavy metals in alga samples (Fernandez-Bobes *et al.,* 1998). An HMDE was employed as

a working electrode resulting in high reproducibility. De-oxygenation of the solution, which can result in poor sample throughput, was not required. The pre-concentration time was 60 seconds and the amount of sample that passed through the cell was dependent on the flow rate and the loop volume. Matrix effects were eliminated by the use of bismuth as an internal standard. The concentrations of zinc, cadmium, lead, and copper in the alga samples, as determined by the FI-DPASV compared favourably to ICP-MS data.

Dual pulse staircase voltammetry (DPSV) has been employed in conjunction with multivariate calibration to quantitatively determine the amounts of glucose, fructose and ethanol in given samples (Richards *et al.,* 2003). The calibration sets were developed via an in-house custom-built automated device employing a number of pumps to prepare a mixture of analytes as determined by the experimental design and to employ a flow cell in which the electrochemical measurements were performed. The best multivariate calibration models were developed by ANNs, which were optimised further by genetic algorithms. Lower error values were attained compared to manually prepared samples coupled with manual data collection.

4.2.1.2 On-line Detection of Heavy Metals

Other systems include a cylindrical microcell containing a working microsensor unit and a reference and counter microsensor unit for the on-line voltammetric detection of heavy metals (Keller & Buffle, 2000a). The microcell had contained integrated micro-channels to enable the solution to "flow-through". The same authors then coupled it to a "supported liquid membrane (SLM) pre-concentration technique" which itself is mostly used for the "industrial separation and recovery of target elements: gases, amino acids, organic acids, anions, complexes and cations" (Keller & Buffle, 2000b).

The SLM contains a microporous hydrophobic membrane which is impregnated with a liposoluble complex-forming agent dissolved in an organic solvent. When the membrane is placed in a sample containing the metal ion such as lead or cadmium, the metal ion reacts with the complex-forming agent at the membrane/sample interface and migrates through the membrane towards the membrane/strip interface where the metal ion combines with a hydrophilic ligand present in the aqueous strip solution.

ASV is then performed on the mercury-coated iridium microelectrode. Iridium was used due to its low solubility in mercury. The authors state that mercury electrodes are more reliable than solid-state electrodes, however they also reported that the mercury had to be plated on to the iridium surface by pumping a degassed mercury (II) solution into the hollow-fibre SLM after its assembly followed by application of a potential of -1.1V vs the Ti/IrO$_2$/MES (pH 6) reference electrode for 15-30 minutes (MES: morpholino ethane sulphonic acid). The overall system was thus able to analyse for trace metals without the need for sample pre-treatment.

A "high-throughput fast-scan ASV in a microflow system" has also been developed (Zhou *et al.*, 1997). Due to the low dead volume of the flow-injection system coupled with the miniaturised "flow-onto thin-layer electrochemical cell" a dramatic reduction in sample consumption was obtained along with an increase in sample throughput. Only a few seconds are required in the pre-concentration step. Although the stripping peaks for lead and cadmium are well-defined, considerable peak overlap was still observed.

4.2.2 Instrument Standardisation

The need for instrument standardisation is exemplified by the fact that multivariate calibration models developed on one *near-infrared* (NIR) instrument do not necessarily concur with results obtained from other similar systems (Wang *et al.*, 1991). The main reasons for these discrepancies has been suggested to be due to different instrumental responses between instruments, a shift in the wavelength over time, and sample variation between batches. The latter implies that a calibration model developed from one batch cannot be applied to another batch. To overcome this, the calibration would need to be performed on the second instrument employing the entire calibration set. This was deemed impractical and costly, especially should the samples were also deemed to be physically or chemically unstable.

Four standardisation methods are described below: classical, inverse, direct and piecewise. The first two directly apply corrections to the calibration model whilst the latter two apply a correction to the response from the second instrument in order to equalise the response to that of the primary instrument.

4.2.2.1 Classical Calibration Model

This assumes a linear relationship between the response and the concentrations of the analytes such that:

$$R_1 = CK_1 \qquad\qquad 4.1$$

where R_1 is the response on the first instrument, C is the concentration matrix for the full dataset and K_1 is a "sensitivity" matrix. The same equation can be written for the response on the second instrument so that:

$$R_2 = C(K_1 + \Delta K) \qquad\qquad 4.2$$

The equation thus holds for the standardisation subset (Equation 4.3).

$$R_2{}^S = C_S(K_1 + \Delta K) \qquad\qquad 4.3$$

Substituting ΔK into Equation 4.2 and using Equation 4.1 leads to the estimation of R_2:

$$R_2 = R_1 + CC_S{}^+(R_2{}^S - R_1{}^S) \qquad\qquad 4.4$$

where $C_S{}^+$ is the pseudo-inverse of C_S. It is also assumed that the linear relationship is valid on both instruments and the concentrations of the analytes contributing to the responses must be known.

4.2.2.2 Inverse Calibration Model

This involves the calculation of regression vectors (b_1 and b_2) for a given analyte on both instruments. The estimation of b_2 is shown in Equation 4.5.

$$b_2 = R^+{}_1 c + (R^+{}_2{}^S - R^+{}_1{}^S)c_S \qquad\qquad 4.5$$

This can thus be employed to predict analyte concentrations on the second instrument. If there is more than one analyte concentration then regression vectors (*c* and *b*) are replaced by concentration and regression matrices (*C* and *B*) respectively. This differs from the classical method in that only the concentration of the target analyte need be known.

4.2.2.3 Direct Standardisation

This approach involves the correction of spectra acquired on the second instrument to "match" the spectra acquired on the primary instrument. Furthermore, the calibration model is untouched. The two data subsets acquired on both instruments are related in Equation 4.6:

$$R_1{}^S = R_2{}^S F \qquad\qquad 4.6$$

F is the square transformation matrix whose dimensions are dependent on the number of wavelength variables. It is calculated from "a relatively small subset" in which singular value decomposition (SVD) is employed to project the raw spectra on to the space of R_1. The generated scores are thus employed to generate the transformation matrix; the added bonus being that noise is filtered out.

With direct standardisation, the concentration of the analytes is not required. This implies that certified or standard reference materials could be employed in place of the actual sample sets.

4.2.2.4 Piecewise Direct Standardisation (PDS)

A limitation with the standardisation methods previously discussed is that the number of subset samples must not exceed the *rank* of R_1. Furthermore, the direct standardisation approach involves employing the whole spectrum acquired on the secondary instrument to be fitted to every spectral point on the spectra acquired on the primary instrument. However, if spectral variations are apparent, each spectral point will more likely be related to neighbouring spectral measurements than to the secondary instrument.

This led to the development of a "piecewise" approach in which each spectral point on the primary instrument is reconstructed from a number of spectral points in a small window on the secondary instrument. This implies that a spectral point at 800nm on the secondary instrument will not influence the spectral point at 550mn on the primary instrument (Wang *et al.,* 1995). The transformation matrix, F is constructed by placing all of the transformation coefficient vectors for each response into the main diagonal and zeros everywhere else (Wang *et al.,* 1992). Thus have:

$$F = \text{diag}(b_1^T, b_2^T, \ldots, b_i^T, \ldots, b_p^T) \qquad\qquad 4.7$$

The term b_i^T is a "vector of transformation coefficient for the *i*th response channel". **P** refers to the total number of channels. The window size is determined by the number of elements in b_i^T. Each calculation of b is via the computation of a number of multivariate calibration regression models, usually PLS (Bouveresse *et al.,* 1996). The transformation matrix can then be employed to standardise any response acquired on the secondary instrument (Wang *et al.,* 1992).

A comparison of the four standardisation methods were performed and piecewise direct standardisation (PDS) was found to be the most efficient (Wang *et al.,* 1991). The classical and inverse methods worked efficiently but required very large sample subsets. Initial findings indicated that the application of PDS did not lead to better RMSEP values than if the calibration was repeated in full on the secondary instrument ($1.2 - 1.6 \times$ error increase). The authors expressed that this was acceptable for real world samples due to the small amount of subset samples employed. In a later study, it was observed that the PDS improved considerably as the number of subset samples increased (Wang *et al.,* 1992). This further led to PDS giving better RMSEP values than if the calibration was repeated in full on the secondary instrument. This is attributed to the method taking advantage of the "larger rank" obtained via a high quality instrument. This further leads to the elimination of on-site calibration.

4.2.2.5 PDS, Additive Background Correction and Sample Subset Selection

The application of PDS in the work reported in this thesis was in conjunction with additive background correction. Although better RMSEP values were reported in Section 4.2.2.4, PDS is merely a "multiplicative correction" method (Wang *et al.,* 1995). In the majority of cases, there are also "additive differences". Examples are the drift in the source wavelength "between measurements of background and sample with a single-beam spectrometer" as inherent in FT-NIR, and in dual-beam spectrophotometers, in which the two beam paths "are never identical". This can be due to an element in the sample path differing from one in the background path. If this element is thus different in the second instrument, the additive background will be significant.

Two methods had thus been proposed. The first involved taking the second derivatives of the acquired spectra on each instrument, whilst the second involved mean-centring the spectra according to the maximum mean value obtained from the first instrument. However, both were deemed to be overly inefficient. The first method only leads to a minimal improvement due to the structure of the background. If the background was a line of constant slope the standardisation would be greatly improved. The second method is deemed only plausible if both instruments possess near-identical backgrounds. Instead, mean-centring the individual set of transfer samples led to the removal of the differences observed in the baseline. This resulted in the generation of a vector containing a correction term corresponding to each wavelength variable.

A disadvantage reported with PDS was that there was no official guidance in determining the actual number of samples in the subset to be standardised (Bouveresse *et al.,* 1996). This is crucial in performing PDS efficiently. The selection method previously employed involved selection of samples with the highest leverage (Wang *et al.,* 1991). Instead a more representative subset was selected via the Kennard-Stone algorithm (Kennard & Stone, 1969) (Bouveresse *et al.,* 1996). In brief, it begins by selecting two spectra that are the farthest from one another as determined by a scores plot (PC2 vs PC1). The next stage involves three steps in which the next sample is selected:

- Calculation of the Euclidean distances between the spectrum under consideration and the spectra already present in the subset

- The smallest Euclidean distance between the spectrum under consideration and the already selected spectra is selected

- The spectrum that has the largest Euclidean distance is determined and added to the subset

The iteration continues until the required number of samples has been attained. It must be stressed that the algorithm is applied to the raw data and that all outliers have been omitted. Outliers are samples within a dataset that do not possess any similarities with the other present samples. PCA score plots can identify potential outlying samples. The selection of the subset samples described above must cover the entire calibration range (Bouveresse *et al.,* 1996; Bouveresse & Massart, 1996).

4.2.2.6 PDS and Electrochemistry

The application of PDS is not restricted to NIR spectra (Wang *et al.,* 1991). Any analytical method that generates multivariate responses can be processed with PDS such as chromatography, flow injection analysis, fluorescence spectroscopy and voltammetry. One example is the application of PDS for the determination of zinc, cadmium, lead and copper via differential pulse polarography on the same instrument but over a number of given days (Herrero & Ortiz, 1997). Although there are many inter-metallic compounds such as gold-zinc, copper-cadmium, copper-zinc, copper-tin, nickel-zinc, etc, the focus in the literature has been on copper-zinc inter-metallic compounds due to their presence in a number of analytical samples ranging from environmental to pharmaceutical. Their presence can be reflected by the observation of a depressed and/or shifted polarographic peak.

Differential pulse polarography is generally employed if the concentration of the target analyte(s) exceeds 1ppm (for each analyte) leading to the minimisation of the effect of inter-metallic compound formation. Partial least squares (PLS) regression was employed leading to the successful calibration of the target metal ion peaks. The

application of PDS implies that there was no need to repeat the entire calibration on the same instrument at a later stage and hence no requirement to construct a new PLS model, since the original PLS model sufficed. Furthermore, the number of measurements was reduced from 28 to 8. The combination of PLS modelling with PDS calibration transfer has led to the development of powerful chemometric techniques.

4.2.3 Chemometrics on a PDA

Personal Digital Assistants (PDAs) are finding great use in the scientific community and this can be useful when performing on-line or at-line analysis. PDA technology itself is also taking a step forward as *wireless* communications are incorporated which could make at-line/on-line analysis even simpler. For example, evaluation of a "trial of networked PDAs in the NHS" which would permit doctors and other medical staff instant access to sensitive information and also omitting the need to carry bulky items about their person is being carried out (Turner *et al.,* 2004).

With the advent of PDAs, field-deployed instruments can be easily controlled in the absence of a PC or even a laptop. The literature reports an amperometric analyser constructed in-house using "off-the-shelf electronic components" (Avdikos *et al.,* 2005). The PDA employed was very basic with a three-button keypad. Extensive information is given by the authors with regards to the construction and operation of the device. It can operate as "stand-alone" or be connected to a PC. The mode of operation was to perform a self-calibration prior to acquiring a univariate calibration curve and then to perform a prediction of the last sample measured. The target compound was ascorbic acid which is present in some pharmaceutical tablets. The electrolyte used was 50mM phosphate buffer at pH5 in 0.5M KCl. Good agreement had been achieved when compared to a standard potentiostat (Autolab PSTAT10). A "relative discrepancy of only 0.4%" was attained.

Although PDAs are being employed to acquire data, there has so far been no mention in the literature of the application of PDAs to perform on-line or at-line predictions via multivariate calibration regression models. Perhaps the main reason can be attributed to the hardware performance of the PDAs, for example with regards to

storage capacity and internal memory. It must also be stressed that certain program applications on PDAs do not contain the same functionality as their PC-based counterparts mostly due to the lack of processing power of the PDA.

This chapter will therefore discuss preliminary findings of the employment of PDAs for the quantitative determination of specific target metal ions using PDS and multivariate calibration regression models developed in both Chapters 2 and 3 for voltammograms acquired on carbon-ink and gold-ink screen-printed electrodes respectively. Furthermore, the subject of automation is discussed with reference to a proposed screen-printed electrode delivery system.

4.3 Materials and Methods

4.3.1 Development of Prototype PDA Application

LabVIEW 7.1 Student Edition with the PDA Module (National Instruments, Austin, TX, US) was purchased and installed on a stand-alone PC with a 750MHz Athlon AMD processor and 354MB RAM running Windows XP Service Pack 2 (Microsoft, US).

A Pocket PC 2003 iPAQ PDA (HP) was employed to run the application. In order to allow communications between the PDA and the PC, Microsoft ActiveSync was also installed on the PC.

A number of multivariate calibration regression models which had been previously created in LabVIEW were transferred on to the PDA and loaded into the PDA application.

4.3.2 Incorporation of Piecewise Direct Standardisation

The data analysis packages developed in Section 2.3.3 in both the Matlab and LabVIEW environments were upgraded with the incorporation and implementation of the piecewise direct standardisation (PDS) algorithm. Focusing on LabVIEW, a new

virtual instrument (VI) was constructed and incorporated into the main program. The front panel is displayed in Figure 4.1.
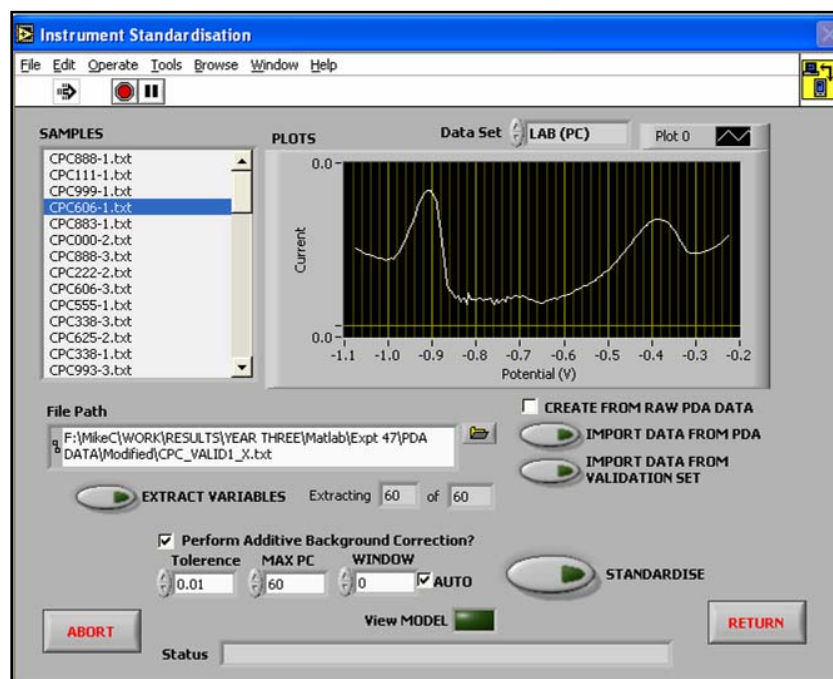
Figure 4.1: The front panel for the instrument standardisation virtual instrument (VI)

As can be seen, the listbox contains the identities of the samples whose voltammograms are correspondingly displayed in the XY graph on the right. Once opened, the user navigates to the text file containing the dataset to be standardised followed by the validation set. The variables are then extracted, and in the example above, there are a total of 60 samples. The user is then able to view the voltammograms from either dataset.

The dataset is then standardised. The user selects whether additive background correction (ABC) is carried out, and also has the option of selecting whether to manually input a window size, or to let the PC select it by checking the "Auto" checkbox. If the window size is set to zero, only direct standardisation (DS) is performed.

This virtual instrument (VI) will also crop the original datasets (training and validation) should the dataset to be standardised contain fewer variables than the original data. Referring to Figure 4.1, there is a checkbox entitled "Create from Raw

PDA Data". If checked, and the user clicks on the "Import Data from PDA" button, a new window appears which allows the user to create the PDA dataset from the individual text files. The user is thus presented with a new interface. The user must input the sample identity for every sample until all samples have been identified. The VI also constructs the appropriate matrix along with the concentration matrix and vector of sample labels, which are all then randomised accordingly. The "raw PDA data" originates from an in-house custom-built PDA application described in Section 4.3.3.

4.3.3 Acquisition on Carbon-Ink Screen Printed Electrodes via the PDA

Instrumentation

The PDA program to acquire the voltammograms was created in-house in C# (Microsoft Visual Studio .Net, Microsoft, US) by Mr Paul Knight at Cranfield University at Silsoe, UK. This was downloaded on to the PDA (a Pocket PC 2003 iPAQ (HP, USA)). The PDA was connected to the multi-channel potentiostat (constructed in-house) which contains four channels with which four electrodes can be connected (Figure 4.2). A more detailed description of the device is given in Appendix A5. The C-SPE-O batch of carbon-ink screen-printed electrodes was employed throughout.

Figure 4.2: In-house custom-built multi-channel potentiostat and Pocket PC PDA

Acquisition

A number of samples were randomly selected from the batch of standards that had been prepared and employed in Section 2.4. The range of standards encompassed the training and validation sets. Due to electrode availability, the samples were only

measured once. A 100µl drop was placed on the working area of the respective electrodes.

Acquisition of the voltammograms occurs in a two-step process. The first involves running the deposition sub-program. After the allotted time, the user must then return to the main program and commence the scan (stripping). After the scan, the user must then press STOP. On doing so, a text file is automatically generated; the corresponding filename is the time in hours, minutes and seconds. It is these text files that are imported into the "Import Data from PDA" VI (Section 4.3.2). Table 4.1 displays the experimental parameters.

Table 4.1: The instrumental parameters employed by the PDA acquisition

Parameter	Value
Electrochemical method	DPASV
Deposition potential (mV)	-1100
Deposition time (s)	165
Initial potential (mV)	-1100
Final potential (mV)	-200
Step potential (mV)	5.0
Modulation amplitude (mV)	50
Modulation time (ms)	50
Interval time (ms)	500
Scan rate (mV/s)	10

Measurements were performed at ambient temperature with the standards employed in Section 2.4.2.

Twenty standards were recorded (Section 4.4.2.3). An extra three samples labelled as "unknowns" were also recorded in order to test the applicability of the data treatment model. These samples had also been employed in Section 2.4.1 and their concentrations are displayed in Table 4.2 for convenience.

Table 4.2: Concentration values for the three "unknown" samples

Sample	Cd	Pb	Cu
UNK1	200	80	160
UNK2	100	20	0
UNK3	0	140	180

The raw data was then either transferred on to a desktop PC for further processing, or manipulated on the PDA itself.

Finally, a number of voltammograms from the extract of the certified reference material (CRM) soil sample described in Section 2.4.2 were acquired on the PDA under the instrumental conditions detailed in Table 4.1.

4.3.4 Acquisition on Gold-Ink Screen Printed Electrodes via the PDA

The instrumentation has already been described in Section 4.3.3. Data was also acquired via a PDA under the instrumental conditions detailed in Table 4.3 but employing the AO batch of gold-ink screen-printed electrodes (Section 3.4.2).

Table 4.3: Instrumental parameters employed for the acquisition of the voltammograms via the PDA

Parameter	Value
Starting Potential (mV)	0.0
End Potential (mV)	500
Deposition Potential (mV)	0.0
Deposition Time (s)	30
Potential Pulse (mV)	50
Pulse Width (ms)	50
Period (ms)	500
Scan Rate (mV/s)	10

Measurements were performed at ambient temperature with the standards employed in Section 3.4.2. A total of 29 standards were recorded (Section 4.4.3.2) in addition to an extra five samples labelled as "unknowns". These samples had also been employed in Section 3.4.2 and their concentrations are displayed in Table 4.4 for convenience.

Table 4.4: Concentrations (in ppb) of the metal ions in the "unknown" samples

Sample	As	Hg	Cu
UNK1	600	400	200
UNK2	400	0	400
UNK3	0	600	0
UNK4	0	600	200
UNK5	500	200	200

The raw data was then either transferred on to a desktop PC for further processing, or manipulated on the PDA itself.

Finally, a number of voltammograms from the extract of the certified reference material (CRM) soil sample described in Section 3.4.3 were acquired on the PDA under the instrumental conditions detailed in Table 4.3.

4.3.5 Automation of Data Acquisition

Two main objectives were sought in the initial attempts at developing an automated system:

- The construction in the LabVIEW environment (version 6.1, National Instruments, Austin, TX, US) of a program to control an in-house custom-built sample preparation device in addition to acquisition of DPASV voltammograms via communication with an Autolab PSTAT10
- The conception and implementation of a screen-printed electrode delivery system

Section 4.4.4 discusses the procedures and findings in greater detail. All potentials are stated relative to Ag/AgCl reference electrodes (+0.197V vs SHE).

4.4 Results and Discussion

The PDA application developed in the LabVIEW environment will be discussed (Section 4.4.1) followed by the implementation of piecewise direct standardisation (PDS) with regards to DPASV voltammograms acquired on carbon-ink (Section 4.4.2) and gold-ink (Section 4.4.3) screen-printed electrodes. Finally, preliminary work on the automation and proposed electrode delivery system will be briefly discussed (Section 4.4.4).

4.4.1 Functionality of the Prototype PDA Application

Figure 4.3A displays the main screen of the PDA application with the options of loading a model, viewing its properties, acquiring a set of voltammograms and carrying out predictions. When the user imports a model, the properties can be viewed via the "Properties" option (Figure 4.3B).

| (A) | (B) |
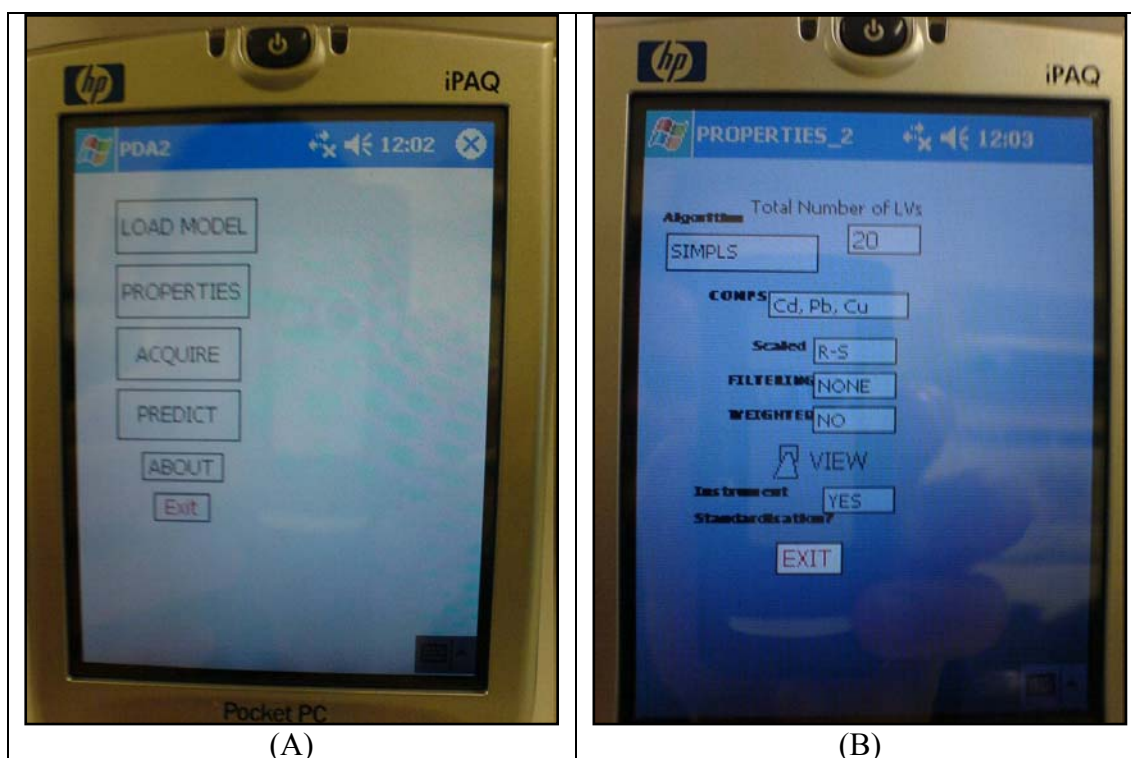
Figure 4.3: The prototype PDA application constructed in the LabVIEW environment:

(A) The welcome screen and root menu; (B) The properties screen of an imported model

The "About" option (Figure 4.3A) informs the user that the current PDA application is only capable of loading models that were created with either no data pre-treatment, range-scaling, or mean-centring. No other treatments such as smoothing with the Savitzky-Golay algorithm, and the application of weights are currently available. As can be seen, the model was generated with the SIMPLS algorithm employing 20 latent variables (Figure 4.3B). The three target metal ion components were cadmium (Cd), lead (Pb) and copper (Cu). It is also evident that range-scaling and "Instrument Standardisation" was performed.

Figure 4.4A displays the "Acquisition" screen in which the datasets containing the voltammograms of the "unknown" samples are imported. There is also the option of importing the raw PDA files. If the checkbox is checked, a new screen appears in which the user imports the PDA text files generated by the multi-channel potentiostat (Section 4.3.3). PDA application, and labels them accordingly (just as in the desktop PC version). Figure 4.4B displays the predicted values.
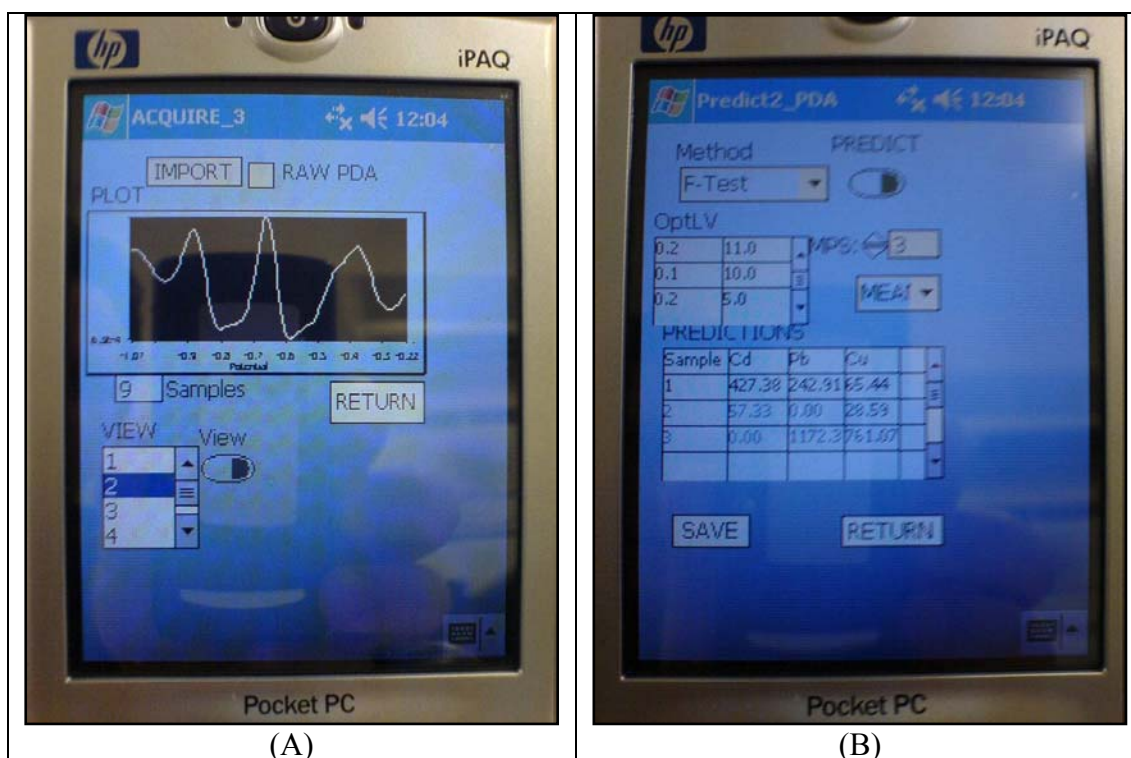
Figure 4.4: The prototype PDA application constructed in the LabVIEW environment:

(A) The "Acquisition" screen; (B) The "Prediction" screen

The voltammograms are thus viewed by selecting the appropriate number in the list box and then selecting "View" (Figure 4.4A). The user selects the method employed to generate the optimum latent variables (OptLV) via the "Method" text ring (Figure 4.4B). The two choices available are *F-test* and *Min RMSEP* (Section 1.5.5.2). The user must then enter the number of measurements per sample (MPS). Selecting "Predict" displays the predicted concentration values for each target metal analyte. As can be seen from the table, there were only three "unknown" samples, and the three target metal analytes were cadmium, lead and copper. The default display in the table is the "Mean" values. The user is therefore able to select from the drop-down menu whether the mean, standard deviations, or relative standard deviations are to be viewed. The final option for the user is to be able to save the predicted results. The filename is automatically generated with the time in seconds appended to the end of the default filename so that the data file is appropriately date-stamped.

It had been observed that many of the functions that are employed in the Windows (desktop) version of LabVIEW (whether version 6.1 or 7.1) cannot be employed in the PDA application. For example, the PDA cannot use *Property Nodes* which add extra functionality to an indicator or a button, such as greying out (i.e. disabling) a button until a specific event has occurred.

4.4.2 Quantitative Determination of Cadmium, Lead and Copper via the PDA

All measurements were performed on the C-SPE-O batch of carbon-ink screen-printed electrodes (Section 2.4.2.2).

4.4.2.1 Conversion of PDA Data

When the voltammograms are acquired via the PDA, they are not stored as DPASV voltammograms. Instead, the *forward* and *backward* currents are stored. The forward current is the current prior to the application of the 50mV potential pulse (S_1); the backward current is the current prior to the removal of the pulse (S_2) (Figure 1.7). Modules were constructed in both the Matlab and LabVIEW environments to convert the PDA data into DPASV voltammograms. Figure 4.5 displays a portion of the LabVIEW interface to illustrate the conversion.
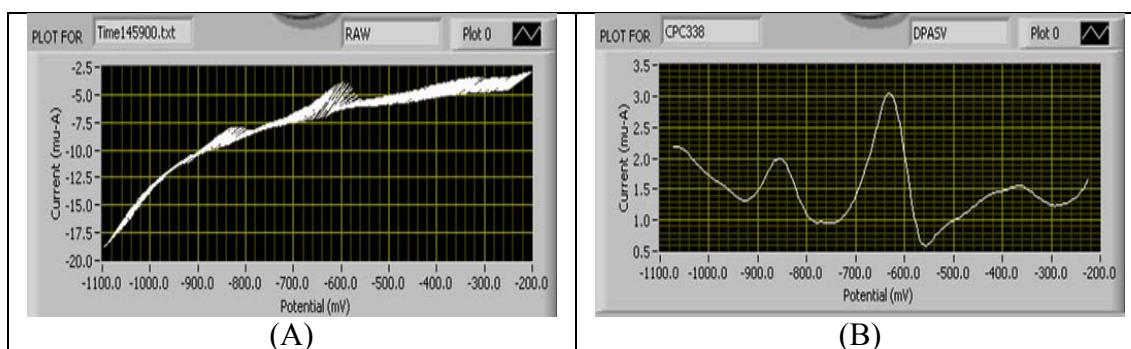
Figure 4.5: Post-processing of an acquired PDA data file corresponding to sample CPC338: (A) RAW; (B) Derived voltammogram

The filename which changes to the assigned sample name when the conversion ensues, may also be observed.

4.4.2.2 Univariate Approach

Figure 4.6 shows the voltammetric profiles obtained for a series of measurements performed on the carbon-ink screen-printed electrodes (C-SPEs) and acquired with the PDA.
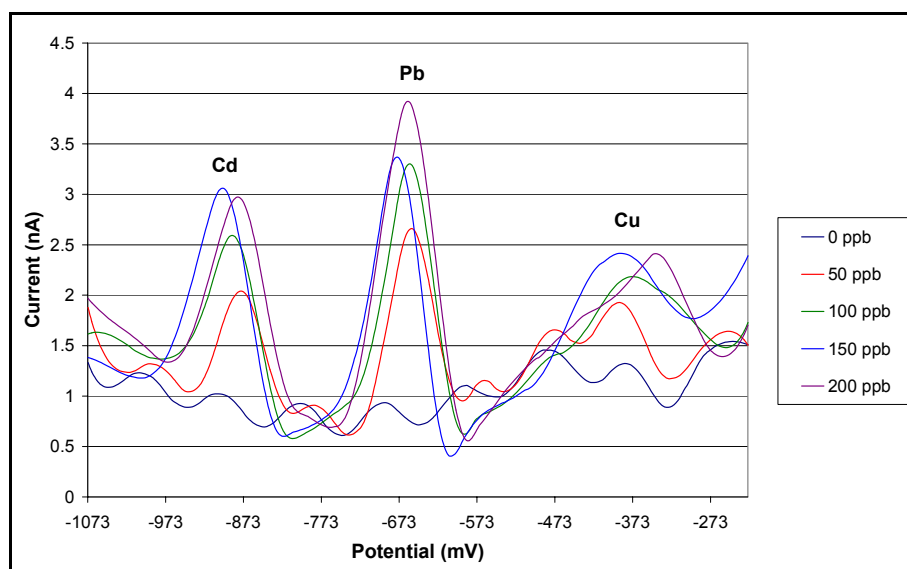
Figure 4.6: Voltammetric profiles for a set of standard samples measured on C-SPEs connected to a multi-channel potentiostat controlled by a PDA. Each standard contained the same metal ion concentration (ranging from 0 ppb to 200 ppb in increments of 50 ppb) in 200ppm mercuric (II) nitrate, 1% nitric acid, 0.1M KCl. Deposition time: 165s; Deposition potential: -1.1V; Range: -1.1 to -0.2V.; Scan Rate: 10mV/s.

A considerable difference in peak height for cadmium and lead between the 0 ppb and the 50 ppb standards may be observed; however there is only a marginal difference with regards to the copper. It is yet unclear as to why there are significant and cyclic waves within the 0 ppb standard which could give rise to low signal-to-noise ratios, hence higher detection limits. One postulation is that this phenomenon is due to electronic interference on or around the connectors, and/or improper shielding. Figure 4.7 displays an auto-correlation plot which supports the claim that there is indeed a cyclic tendency.
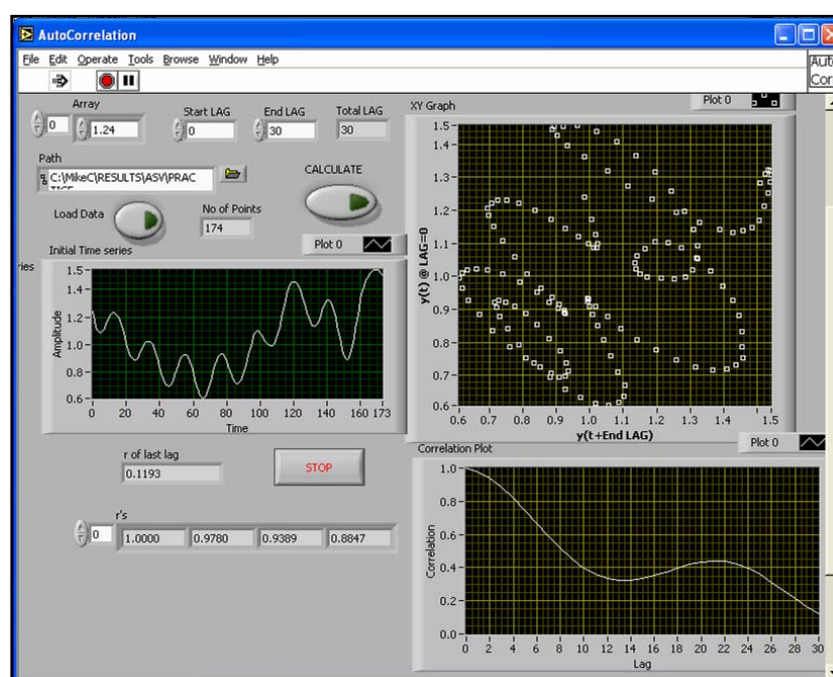
Figure 4.7: Auto-correlated plot for the 0 ppb standard acquired on the PDA under the experimental conditions described in Figure 4.6

The relatively high correlation coefficient values observed, even after a lag of 5 potential indices, confirm that there are cyclic tendencies. If there was no auto-correlation, the "Correlation plot" would be very close to zero irrespective of the lag (except for when the Lag = 0; the correlation coefficient = 1).

It has also been observed that there are no major increases in peak intensity from 50 ppb to 200 ppb except for cadmium. This implies that the correlation between the concentration and the peak areas will have a tendency to be of a non-linear nature. Figure 4.8 displays the resultant univariate calibration curve.
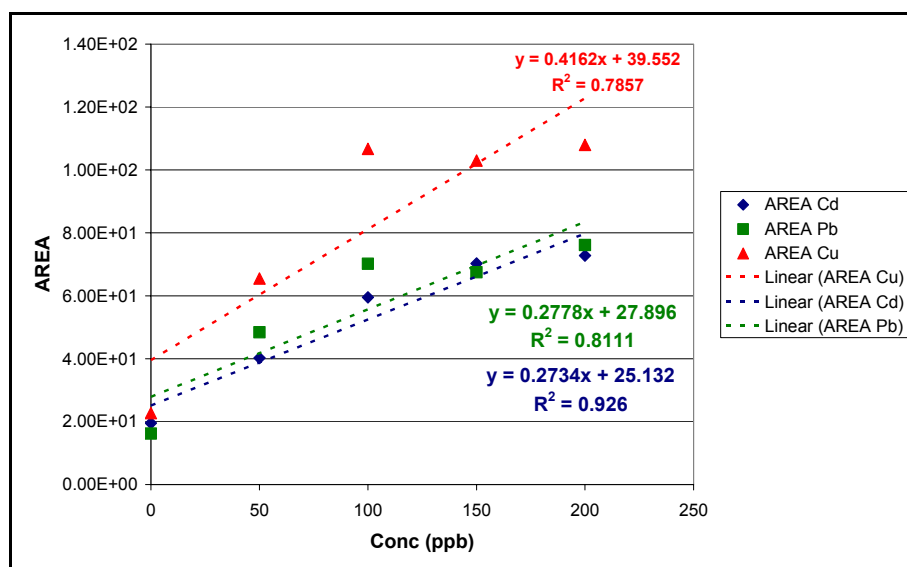
Figure 4.8: Univariate calibration plot of the calculated areas against standard concentrations from the voltammograms acquired on the PDA under the experimental conditions described in Figure 4.6. Concentration values of 0, 50, 100, 150, and 200 ppb for each metal. Error bars not shown due to each standard being measured only once.

As expected, the coefficients of determination (R^2) are relatively poor for lead and copper at 0.8111 and 0.7857 respectively compared with 0.9828 and 0.9782 from the Autolab PSTAT10 (Section 2.4.1.1) . Inclusion of the validation concentrations at 90, 130 and 180 ppb did not improve the coefficients (0.7656 and 0.7694 respectively). The coefficient for cadmium also decreased from 0.926 to 0.8861. This is also reflected in the predictions of the "unknown" samples as shown in Table 4.5.

Table 4.5: The predicted concentration values calculated from the respective slopes and intercepts from Figure 4.8 and the inclusion of the validation concentration set (IVCS) and compared to the actual values

Components	TRUE			Figure 4.8			IVCS		
	Cd	Pb	Cu	Cd	Pb	Cu	Cd	Pb	Cu
UNK1	200	80	160	131	117	110	128	116	109
UNK2	100	20	0	101	1	75	98	(-10)	71
UNK3	0	140	180	(-18)	118	172	(-21)	116	175

This leads to an overall root mean square error (RMSE) of 41.45 ppb for Figure 4.8, and 42.19 ppb for inclusion of the validation concentration (IVCS). This further

implies that the univariate approach cannot be employed for prediction of the unknown samples.

4.4.2.3 Multivariate Approach

Due to the small number of samples employed for multivariate calibration on the PDA (Table 4.6) high RMSEP values were obtained for the prediction of the unknown samples compared to Section 2.4.1.2.

Table 4.6: Concentrations of the standards employed in the acquisition of the respective voltammograms via a PDA. Experimental conditions described in Figure 4.6

Standard	Cd	Pb	Cu
1	130	130	180
2	180	180	130
3	0	150	200
4	90	90	130
5	150	200	50
6	150	150	150
7	130	130	130
8	90	90	90
9	180	180	180
10	100	100	100
11	200	200	200
12	50	50	50
13	0	0	0
14	0	50	150
15	0	100	0
16	0	200	0
17	200	0	150
18	150	0	150
19	0	100	100
20	150	100	0

The optimum latent variables suggested via the minimum RMSECV and the F-test were identical throughout. A *U vs T* plot (not shown) indicated that there were no outliers. The average overall RMSEP values calculated for the unknown samples were 69.65, 70.44 and 72.60 ppb for data pre-treatment with range-scaling, no scaling and mean-centring respectively.

4.4.2.4 Instrument Standardisation

The voltammograms of the standards acquired via the PDA (Table 4.6) were transferred to a desktop PC and combined into a 20×3 matrix. The corresponding voltammograms were thus extracted from the relevant training and validation sets (employed in Section 2.4.1) and combined into a second matrix. The two matrices were then employed to effect the instrument standardisation. The voltammograms pertaining to the "unknown" samples (Table 4.2) were transferred from the PDA to the PC, standardised by the transformation matrix (Section 4.2.2.4) and predictions carried out. Figure 4.9 compares the calculated overall RMSEP values for standardised datasets.
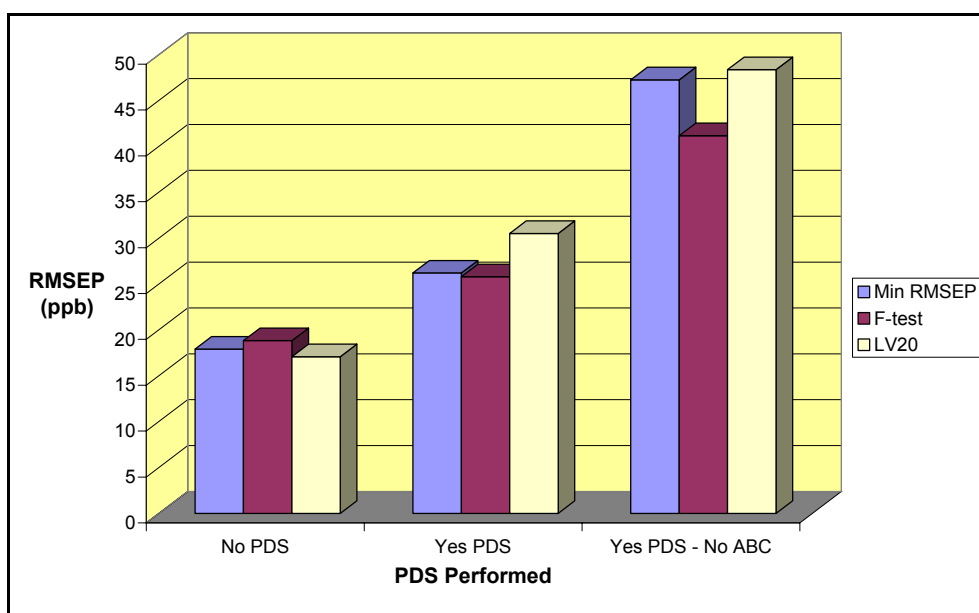
Figure 4.9: RMSEP plots for range-scaled data after SIMPLS modelling: (A) "No PDS" implies that predictions were carried out on the unknown samples measured on the Autolab; (B) "Yes PDS" implies that predictions were carried out on the unknown samples measured on the PDA and thus having undergone instrument standardisation. Window size: 87; (C) "Yes PDS – No ABC" implies (B) but with no additive background correction (ABC).

A similar trend was also observed for mean-centred data (not shown). As can be seen, the predictions performed for the unknown samples acquired on the Autolab are better than those acquired on the PDA by 10 to 15 ppb. However a number of reasons were considered for this. One is the minimal number of voltammograms acquired on the PDA pertaining to the standards in Table 4.6 leading to minimal variability in the

modelling implying the generation of a weak transformation matrix. Another is due to the fewer number of variables in the voltammograms acquired on the PDA (175) compared to the Autolab (184), the voltammograms acquired on the latter must thus be cropped leading to the possible loss of information. However as only the first few variables at either end of the voltammogram are omitted, this is considered not to have a significant impact. It was also observed that the omission of additive background correction (ABC) worsened the predictive abilities of the SIMPLS model in conjunction with a weaker transformation matrix with an increase of $20 - 25$ ppb. It must be stressed that the data analysis was performed both in the Matlab and LabVIEW environments, and that identical results were attained.

A final point to mention is that setting the window size to zero, and thus invoking the *direct standardisation* method as opposed to *piecewise direct standardisation*, leads to very poor standardisation models generated; so poor that the results are not shown (very high RMS values exceeding 10^8 were obtained!). This is due to singular matrices being generated during the modelling stages of the standardisation via the SVD function (warning messages are displayed in Matlab). No results could be obtained in the LabVIEW version due to the LabVIEW SVD function outputting an empty matrix rather than a singular one.

4.4.2.5 Application of PDS to CRM Soil Extract

Figure 4.10 shows the overlaid voltammograms acquired on the PDA of the prepared CRM samples with increasing concentrations of cadmium, lead and copper compared to voltammograms acquired on the Autolab PSTAT10.
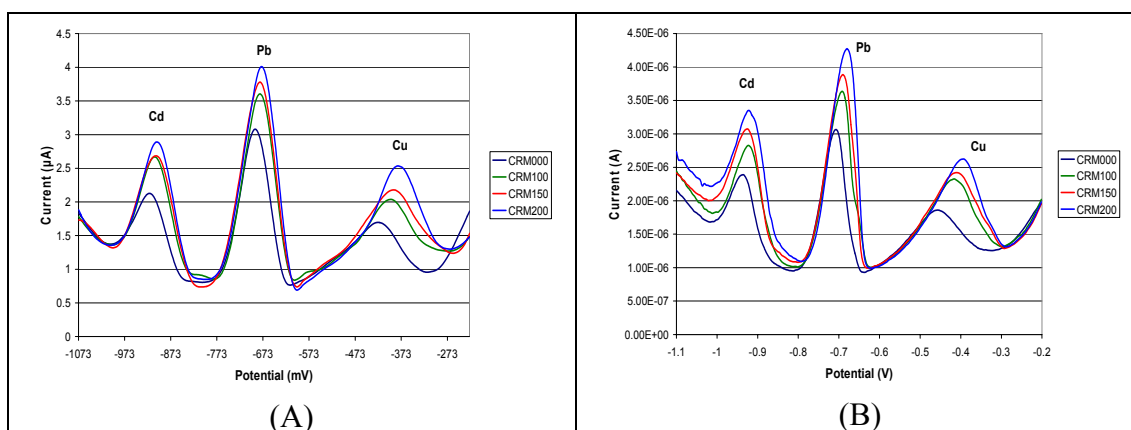
Figure 4.10: Overlaid voltammograms of Cd, Pb and Cu in the CRM samples acquired on the PDA (A) and Autolab PSTAT10 (B). Instrumental conditions: deposition time: 165s; deposition potential: -1.1V; scan rate: 10mV/s)

It is evident to see that the extraction procedure employed has been very successful in extracting the target metal ions. However, the peak heights of the voltammograms are less distinguished (Figure 4.10A), especially at the 100 and 150 ppb level. Standard addition experiments were also performed (Figure 4.11).
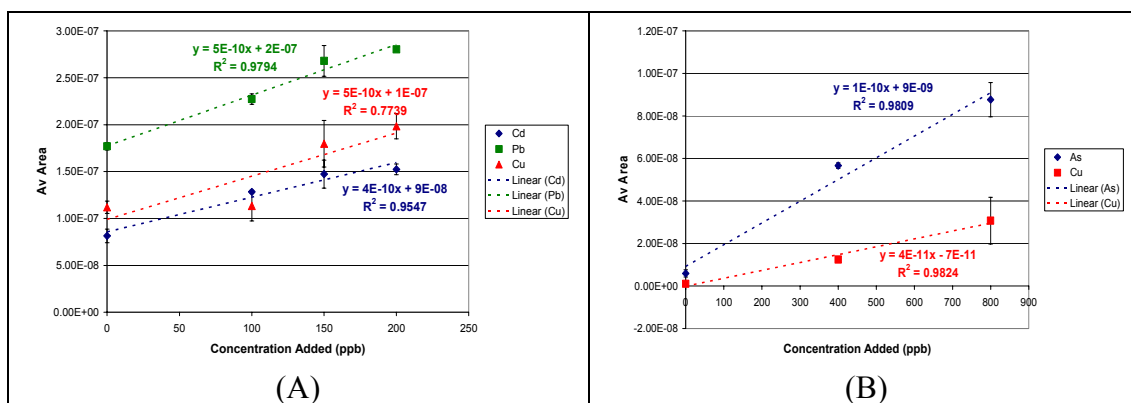
Figure 4.11: Standard addition performed after calculation of the peak areas obtained in Figure 4.10: (A) PDA; (B) Autolab PSTAT10

As is clear to see in Figure 4.11B, the coefficients of determination (R^2) are encouraging. The magnitude of the error bars increase with concentration, particularly for copper (concurs with previous findings). However, the calibration is not as good for the PDA-acquired data (Figure 4.11A). This is evident in the coefficients of determination (R^2) and the increase in the heights of the error bars. Table 4.7 compares the calculated % recoveries of metals from the soil on the PDA and Autolab via standard addition whilst Table 4.8 compares the different multivariate calibration

regression models. The method employed to calculate % recoveries of the respective metals is to calculate the amount of metal in mg/kg from the predicted concentrations taking into account the dilution factors and then to divide by the true amount as stated on the information sheet supplied with the CRM sample.

Table 4.7: Determination of the % recovery for Cd, Pb and Cu present in the CRM via standard addition (Figure 4.12)

Target Ion	% Recovery (Flame AAS)	% Recovery (PDA)	% Recovery (Autolab)
Cd	68.49	114.5	118.5
Pb	26.34	93.01	93.01
Cu	38.32	63.32	63.32

Table 4.8: Determination of the % recovery for Cd, Pb and Cu present in the CRM via prediction with the PDS and SIMPLS models (Minimum RMSEP). Key: P = PDA; A = Autolab Note: PDS only applied to PDA voltammograms

Element	Flame AAS	Raw		Range-scaled		Mean-centred		Weighted	
		P	A	P	A	P	A	P	A
Cd	68.49	6.90	60.23	15.29	-	17.89	61.88	37.81	68.18
Pb	26.34	33.89	47.05	32.98	-	34.02	47.35	64.56	40.76
Cu	38.32	39.85	19.96	35.54	-	40.57	18.90	57.90	27.16

The % recoveries calculated from the voltammograms acquired on the Autolab agree more with the % recoveries determined by flame AAS than do the PDA voltammograms. This can be attributed to the poor transformation matrix generated by the PDS algorithm as explained in the previous section. This has led to the SIMPLS models employed not being sufficiently robust to predict the analyte concentrations. This is re-enforced by the % recovery calculated via standard addition (Table 4.7) in which the values calculated from the voltammograms acquired via the PDA and the Autolab are in very close agreement, for example 63.32% for copper.

4.4.3 Quantitative Determination of Arsenic, Mercury and Copper via the PDA

All measurements were recorded on the AO batch of gold-ink screen-printed electrodes (Section 3.4.1.2).

4.4.3.1 Univariate Approach

The voltammetric response obtained on the PDA was very poor compared to the response on the carbon-ink screen-printed electrodes (Figure 4.6). Figure 4.12A displays the overlaid voltammograms acquired on the PDA with the corresponding univariate calibration curve in Figure 4.12B.
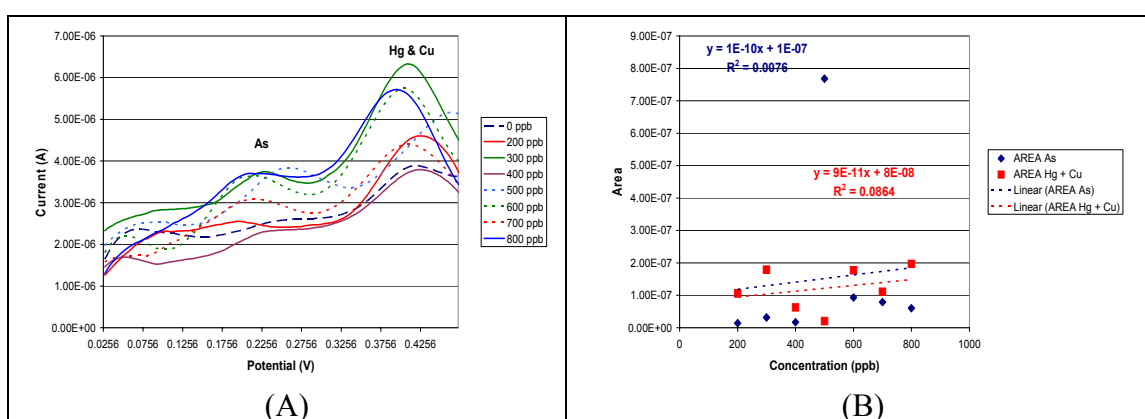
Figure 4.12: The overlaid voltammograms (A) and calibration curves (B) for arsenic (III) and the "mixture" of mercury (II) and copper (II) acquired on the AO batch of gold-ink screen-printed electrodes. Instrumental conditions as in Table 4.3

As can be seen by the coefficients of determination, there is no correlation (Figure 4.12B). In the case of arsenic, it is due to the huge area calculated at 500 ppb. However, as the concentration increases, the computed areas decrease. Calculating the coefficient of determination for the 200, 300 and 600 ppb arsenic concentrations, a linear relationship ($R^2 = 0.9992$) is observed. The mercury and copper "mixture" is no better. This is mostly due to no distinct peak being observed from the baseline and so determining the area was difficult.

It must be stressed here that only one measurement per sample was acquired due to a limited number of electrodes. As can be seen, the arsenic peak intensity does not increase as the concentration increases (Figure 4.12A). This can be attributed to a

number of factors such as the poorer performance of the AO batch (discussed in Section 3.4.2) and the sensitivity of the in-house built multi-potentiostat compared to the "more sophisticated" Autolab. Employing the coefficient of determination of 0.9992, the arsenic concentration was determined in the sample labelled "UNK2" (Table 4.4) which had also been measured on the PDA. The predicted concentration was 487 ppb (400 ppb true). No other predictions were performed.

4.4.3.2 Multivariate Approach

Table 4.9 displays the standards measured on the gold-ink screen-printed electrodes via the PDA.

Table 4.9: The standards measured on the PDA under the same experimental conditions as detailed in Table 4.3

Sample	As	Hg	Cu	Sample	As	Hg	Cu
AHC822	800	200	200	AHC000	0	0	0
AHC628	600	200	800	AHC733	700	300	300
AHC888	800	800	800	AHC555	500	500	500
AHC662	600	600	200	AHC333	300	300	300
AHC222	200	200	200	AHC753	700	500	300
AHC840	800	400	0	AHC533	500	300	300
AHC648	600	400	800	AHC337	300	300	700
AHC666	600	600	600	AHC777	700	700	700
AHC622	600	200	200	AHC000	0	0	0
AHC024	0	200	400	AHC-U1	600	400	200
AHC824	800	200	400	AHC-U2	400	0	400
AHC444	400	400	400	AHC-U3	0	600	0
AHC262	200	600	200	AHC-U4	0	600	200
AHC026	0	200	600	AHC-U5	500	200	200
AHC864	800	600	400	--	--	--	--

The leave-one-out cross-validation plots for the dataset that had been acquired on the PDA (not shown) were poor due to the insufficient number of samples and the fact that only one measurement per sample was performed. For example, the plot for copper contained a maximum at the 14th latent variable which can also be attributed to the presence of an outlier but is more likely due to the absence of a discernible peak in the +0.45V range. Application of range-scaling and mean-centring does not improve matters. Furthermore, the variance captured in the Y-block for the first latent variable

decreases from 75 → 67 → 12% respectively. Application of weights to the dataset does not improve matters (Figure 4.13).
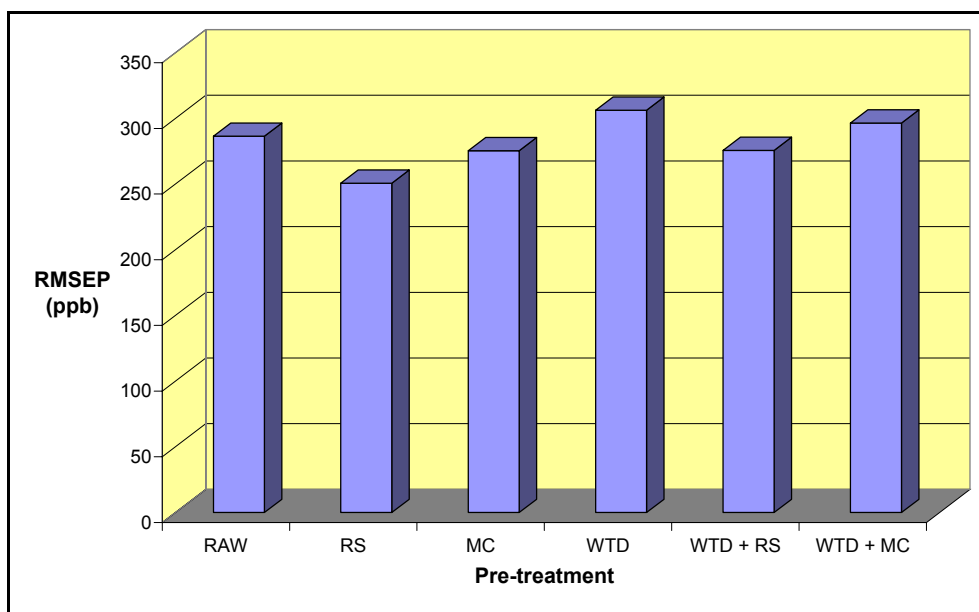
Figure 4.13: RMSEP plots for the prediction of the "unknown" concentration values (Table 4.4) with respect to the different pre-treatment techniques

As can be seen, the model constructed via the range-scaled data offers the better predictions in both the non-weighted and weighted datasets (with ~240 and ~260 ppb respectively). Overall, the models are very poor due to only a small number of samples employed.

4.4.3.3 Instrument Standardisation

Due to the overall poor quality of the voltammograms obtained with the AO batch employed on the PDA "field" device, it was envisaged that instrument standardisation would not be successful in contrast to what had been observed with the carbon-ink screen-printed electrodes. Figure 4.14 compares two voltammograms of the same standard but acquired on different instruments.
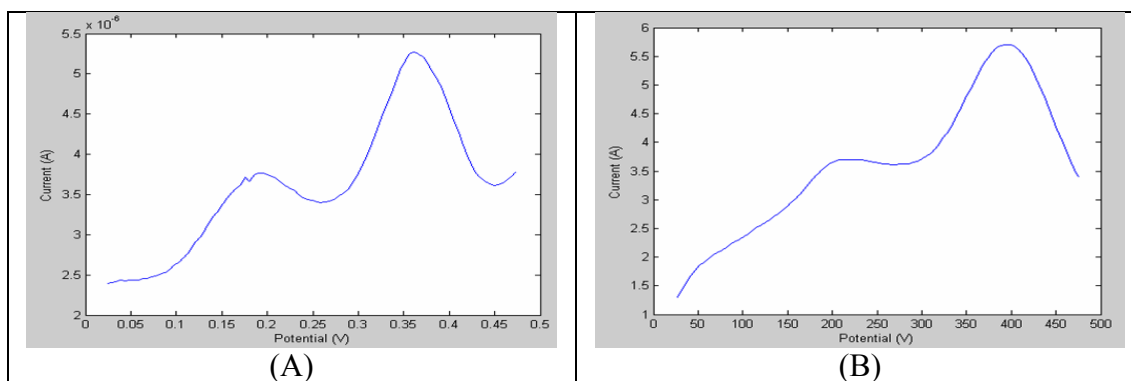
Figure 4.14: Voltammograms of standard AHC662 (600 ppb As and Hg; 200 ppb Cu) acquired on an Autolab multi-PSTAT10 (A) and the PDA-controlled in-house built multi-potentiostat (B)

Instrument standardisation was performed with the piecewise direct standardisation (PDS) algorithm incorporating additive background correction (ABC). Figure 4.15 displays the RMSEP plots for the predictions of the three target metal ions under different pre-treatment techniques.
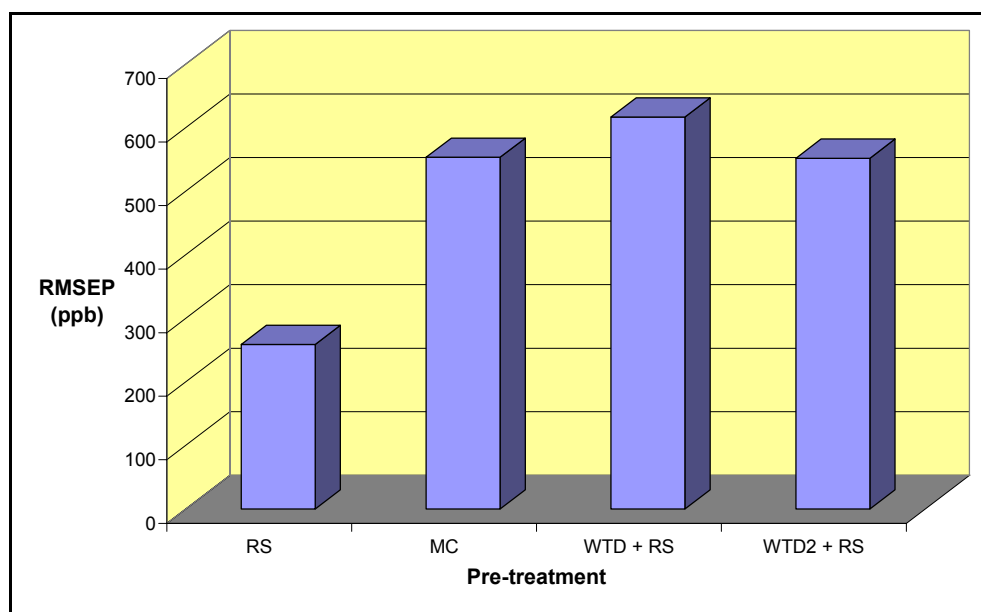
Figure 4.15: The RMSEP plots for the predicted concentrations of arsenic (III), mercury (II) and copper (II) in 4M HCl from standardised PDA data

Data pre-treatment with range-scaling (RS) has led to a better overall RMSEP value with ~250 ppb compared to an average of ~550 ppb for the other pre-treatment techniques. The "WTD2" term implies that the PDA data was also weighted in order to observe whether improvements could be made, in other words, whether the PDA data should also be weighted, as was suggested in Section 4.4.2. One would therefore

expect improved PDS models to be attained. Although there is a slight improvement of ~ 100 ppb in the RMSEP values compared to the normal default weighting (Figure 4.15), it was deemed necessary to make alterations to the main programs at a later stage.

In short, there are insufficient samples to properly generate a PDS model (transformation matrix) and hence to standardise the voltammograms that were acquired on the PDA-controlled "field" device. This is partly attributable to the limited electrodes available. Furthermore, the qualities of the voltammograms were very poor compared to those recorded on the AI batch on the Autolab multi-PSTAT10 instrument (Section 3.4.2).

4.4.3.4 The CRM Soil Extract

Figure 4.16A displays the overlaid voltammograms acquired on the AO batch of electrodes via the PDA. Only one measurement per sample was performed. The profiles are not distinctive although a small flattened peak in the CRM000 voltammogram for arsenic is observed at +0.2V. This peak is also more pronounced after the addition of 800 ppb arsenic. However, a "tailing off" for copper in the +0.45V region is not observed. This has therefore made it difficult to ascertain the peak area. This has thus contributed to the relatively poorer standard addition curve as shown by the coefficient of determination (R^2) value of 0.923 (Figure 4.16B).
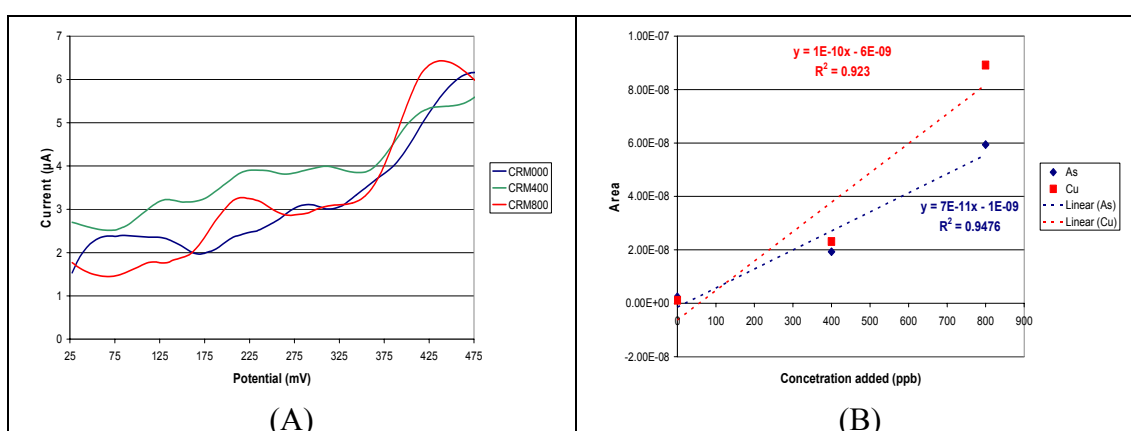
Figure 4.16: Overlaid voltammograms (A) and standard addition curve (B) for As and Cu recorded on the AO batch of gold-ink screen-printed electrodes via the PDA

As the extrapolation of both arsenic and copper at Y=0 does not lead to a "negative" value on the X-axis in conjunction with negative values on the Y-axis at X = 0 ppb (Figure 4.16B), this implies that no arsenic and copper was present in the CRM solution. This conforms to what was observed in Section 3.4.3.2 in that the SIMPLS models predicted no presence of both arsenic and copper, but is in contrast to what was observed in Section 3.4.3.1 for the standard addition performed on voltammograms acquired on the Autolab in which an extraction value of ~90% was attained. However, it must be stressed that there are only three data points in the standard addition curve (Figure 4.16B). Furthermore, the samples were only measured once on the gold-ink screen-printed electrodes which, as has been shown elsewhere, are not very reproducible.

Importing models in which data pre-treatment in addition to instrument standardisation was performed gave a highly over-predictive arsenic content in which the % recovery was in excess of 350 and 800 for un-scaled and range-scaled data respectively. The poor performance was expected due to instrument standardisation not being highly successful in the previous section.

4.4.4 Development of an Automated System

The advantages of developing an automated data acquisition system negates the effects of human error and increases the number of experiments performed 100-fold compared to manual acquisition (Richards *et al.,* 2003). Section 4.4.4.1 briefly describes the development of a sampling preparation device. Section 4.4.4.2 describes the preliminary efforts to develop a screen-printed electrode dispenser. Finally, Section 4.4.4.3 details the preliminary attempts at creating an interface in LabVIEW to acquire DPASV voltammograms directly from the hardware controller.

4.4.4.1 Development of the Sample Preparation Unit

The sample preparation device was developed in-house in order to automatically prepare samples prior to the acquisition of a dual pulse staircase voltammogram (Richards *et al.,* 2003). The device was labelled a "Robotic Technician" or "Robotech" for short. A program had been constructed in Borland Visual C++

(Windows 98) to control the array of pumps in order to effect the sample preparation within the mixing chamber. Solid-state electrodes were employed throughout and connected to an external analytical system with which the program communicated. Full functionality and operating conditions of the program and the device are detailed in the reference.

Due to the author's prior knowledge of the LabVIEW programming environment, a new program was constructed to communicate with the sample preparation device. The advantage of this action is that the program would be much easier to upgrade as advances were made. Furthermore, the data analysis package developed in the LabVIEW environment (Chapter 2) could either be implemented as a sub-program, or the data generated by the new program (labelled as "Robotech II") could be exported in a custom format so as to be easily manipulated by the data analysis package.

Acquisition of the DPASV voltammograms was by an Autolab PSTAT10 (Eco Chemie, Holland) which was connected to the sample preparation device by a parallel printer cable. With both the Borland Visual C++ and LabVIEW versions, the GPES 4.9 application (Eco Chemie, Holland) was also initiated. In order to allow communication between the two programs and for data acquisition to commence, a custom-built macro was executed (Figure 4.17).

```
Procedure!Method = VA
Procedure!Open("C:\MikeC\LabVIEW\ROBOTECH II\DPASV-AuSPE")
DIO!SetMode("P2","A","IN")
Repeat(250)
System!Beep
DIO!WaitByte("P2","A","1")
Procedure!Start
System!Beep
Dataset!SaveAs("C:\MikeC\LabVIEW\ROBOTECH II\DATA\DEFAULT")
EndRepeat
```

Figure 4.17: The MACRO script executed by the GPES application

In short, the method is loaded (VA), then the Procedure File is loaded. The GPES is then notified that an external source will trigger the Autolab via "P2" and channel "A". The "IN" function informs the GPES that it is an "incoming" signal. In the above case, there will be 250 measurements. The GPES is then told to WAIT until the signal is received before acquiring the voltammogram. When the signal is received, the voltammogram is acquired (deposition followed by stripping as detailed in the

Procedure file). When the voltammogram has been acquired, the system emits an audible signal and the file is saved as "Default". The MACRO loops round and waits for the next input signal, or terminates when all 250 measurements have been made.

4.4.4.2 A Proposed Development for a Screen-Printed Electrode Dispenser

The circuitry for the screen-printed electrode delivery system was designed, built and implemented into the "Robotech II" LabVIEW program. The circuit board was equipped with four LEDs. The purpose of the LEDs is to inform the user as to what the status of the dispenser is. Figure 4.18 illustrates the simplified electronic design of the device.
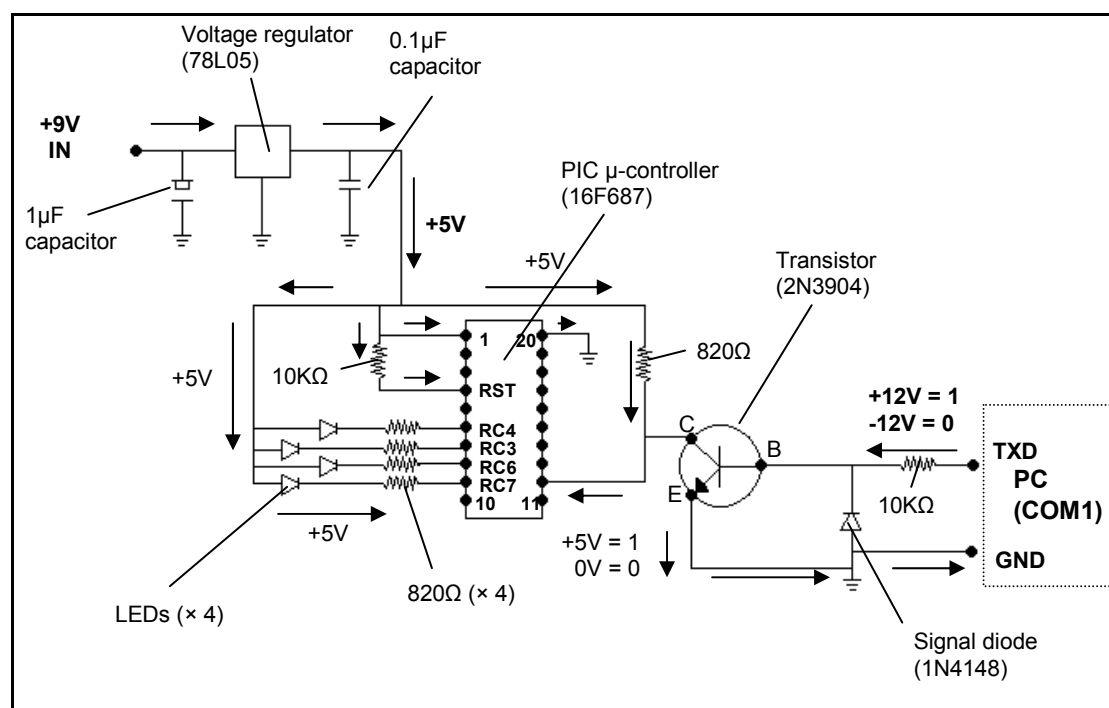
Figure 4.18: Simple electronic design of the screen-printed electrode delivery device

This circuit board would later be contained within the dispenser unit which itself would be powered by an external power source either by AC mains via a built-in transformer or a 9 or 12V DC input. This would be connected to the PC via the serial port (COM1: 9600 baud rate; 8-bits; 1-stop bit; 0-parity). The outlined approach is executed as follows:

- The voltage regulator ensures that the 5V passes through the entire circuit. The capacitors (1 and 0.1μF) act to "steady the flow of current".

- The PIC (programmable integrated circuit) micro-controller designates which LED is to be lit dependant on what signal it receives from the COM1 port of the PC. A small program written in BASIC is embedded within the controller

- The appropriate LED is lit when the corresponding pin on the PIC is set to 0V; this creates a potential difference which allows the current to flow and thus the LED to be lit. The LED is unlit when the corresponding pin on the PIC is set to 5V.

The design of the delivery system is currently envisaged to be in the form of a rotating disc containing four grooves which each hold a single screen-printed electrode. A stack of pre-cut electrodes sits above the disc. As the disc rotates, and one of the grooves passes beneath, an electrode is released. This electrode remains in the groove until it reaches the connector and is employed. After the acquisition of a voltammogram, the disc is rotated. As another electrode is selected, the previous one is ejected. Figure 4.19 displays the overall instrumental set up of the proposed system.
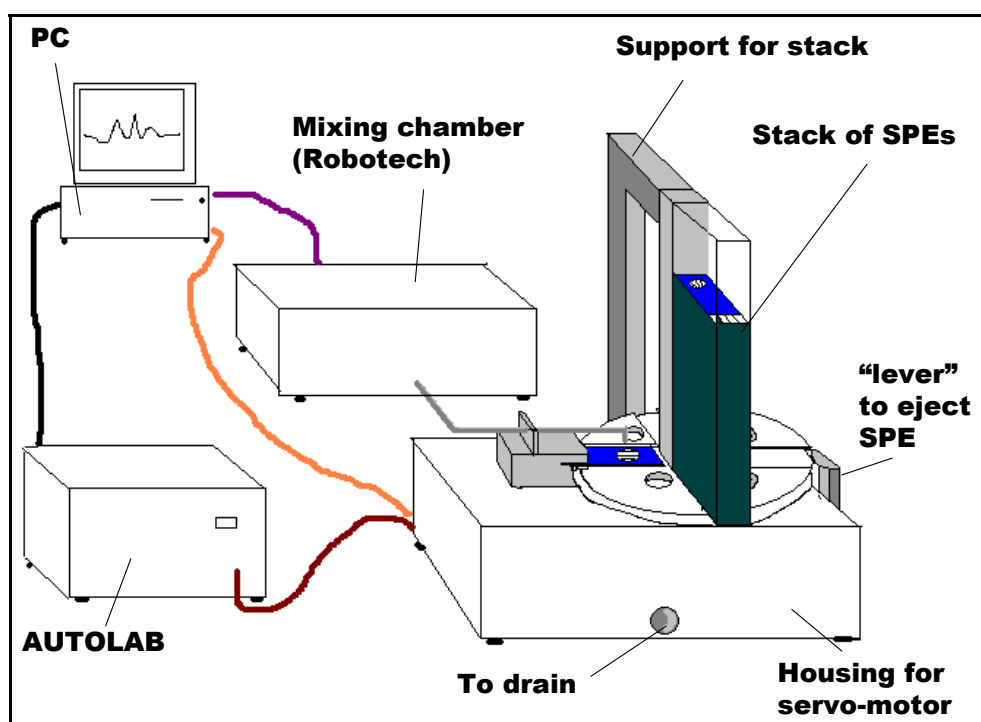
Figure 4.19: Proposed instrumental set up for automated system

There are at present four LEDs. Each of these informs the user that a specific function is being carried out. Table 4.10 lists the functions of the LEDs along with their bit values.

Table 4.10: Function of the four LEDs and corresponding bit patterns

LED (and pin)	Function	Numeric	Hexadecimal	Bit pattern
Δ (RC4)	Rotate disc	1	1	0001
γ (RC3)	Contacting	2	2	0010
β (RC6)	Eject SPE	4	4	0100
α (RC7)	New SPE	8	8	1000

If the Δ-LED is lit an appropriate signal is sent to the rotor to rotate the disc for a specified number of seconds (depending on the speed of rotation of the rotor coupled with the ability to be able to vary the rotational speed). Alternatively, if the γ-LED is lit, then a solenoid is activated so as to ensure contact is made between the terminals of the screen-printed electrode and the device.

An alternative design is illustrated in Figure 4.20. It consists of a substrate disc of carbon-ink screen-printed electrodes (blue). The areas in white are blank but contain drainage holes for the cleaning cycle. Once every electrode on the disc has been exhausted, the terminal connectors retract whilst the drop dispenser swivels to the right. The next disc is dropped for the next round of measurements. Should there be insufficient discs, the program controlling the automated process would pause the acquisition until new discs are placed in the holder.
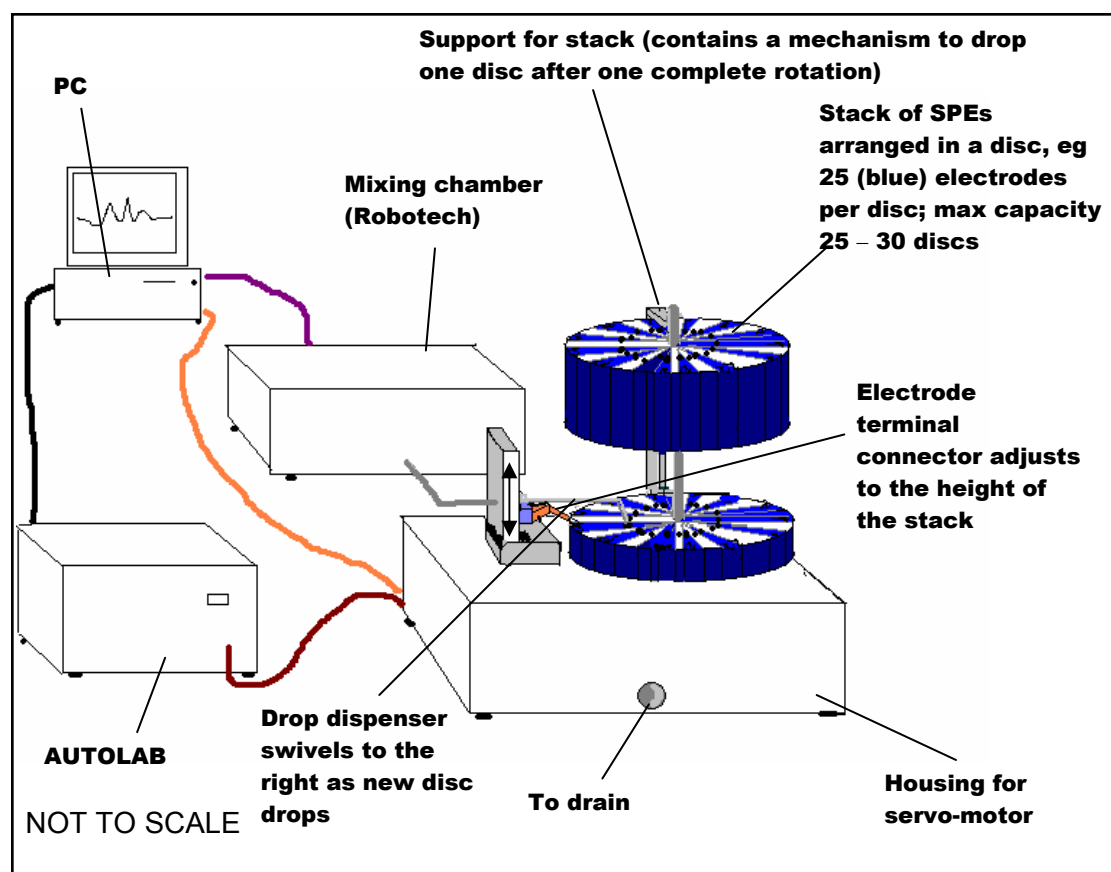
Figure 4.20: An alternative design for the screen-printed electrode dispenser

In order to avoid cross-contamination, a second tube could be employed to remove the drop after acquisition of the voltammogram. The important part of the design is that the height of the terminal connector is adjustable and rises as a new disc drops. Sophisticated mechanisms would be required. However the advantage of this design over the one in Figure 4.19 is that there is no risk of an electrode failing to fall into a groove. Furthermore, there is no need to cut out individual electrodes from the sheets.

For each measurement to be made, the following steps would be performed (Figure 4.19):

1. Disc rotated one-eighth of a turn clockwise which contains a new SPE. The corresponding LED (Δ) is lit.

2. Sample preparation performed by pumps as determined by the concentration values of each component for that particular sample

3. Disc rotated one-eighth of a turn clockwise which places SPE in position. The corresponding LED (Δ) is lit.

4. Contact is made with the terminals of the SPE. The corresponding LED (γ) is lit.

5. Sample is injected on to the working area of the SPE by the sample preparation device.

6. The Autolab is triggered. DPASV voltammogram is recorded by the Autolab and stored in a text file with a default filename.

7. The main program copies the default file with a new name and deletes the default file. This is to prevent loss of data should problems arise.

8. The program extracts the new voltammogram from the new file and displays it.

9. The terminals of the SPE are disconnected. The corresponding LED is unlit.

10. A new SPE is placed on to the disc whilst an older SPE is ejected. The appropriate LEDs are lit when this occurs.

11. The disc is rotated one-eighth of a turn. The corresponding LED (Δ) is lit.

12. The wash cycle is performed. Go to Step 2.

Ten measurements were performed for the analysis of cadmium, lead and copper in 200 ppm mercuric (II) nitrate, 1% v/v nitric acid, and 0.1M KCl. As the dispenser does not exist, the samples were added manually on the SPE. The purpose for the run was to ensure that the LEDs functioned when required, and that the dispenser could actually be physically implemented into the design. Figure 4.21 shows the user interface for the "Robotech II" application after acquisition and displaying of a voltammogram.
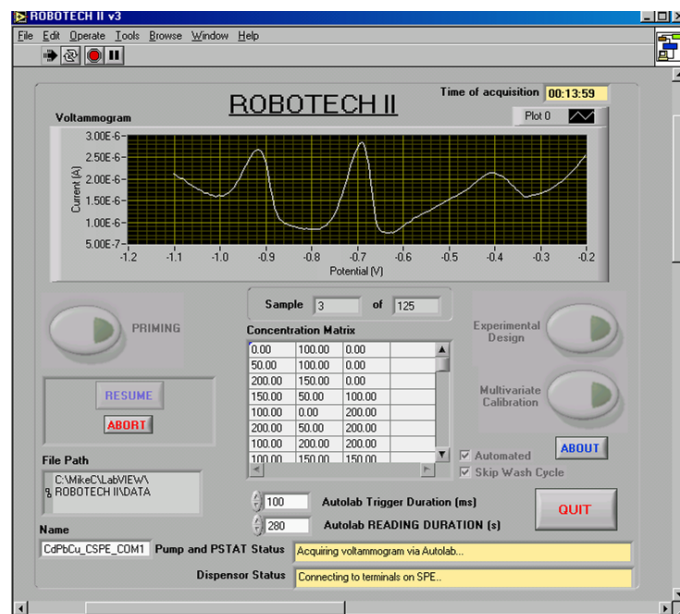
Figure 4.21: User interface of application after acquisition of voltammogram (100 ppb)

Note that the concentration values displayed in the concentration matrix table in Figure 4.21 do not correspond to the concentrations of the samples measured in this instance. The measurements were performed to ensure that the program could acquire and store the data whilst controlling the other peripherals such as the pumping unit and the dispenser. The voltammograms obtained for 50, 100, 150 and 200 ppb are displayed in Figure 4.22.
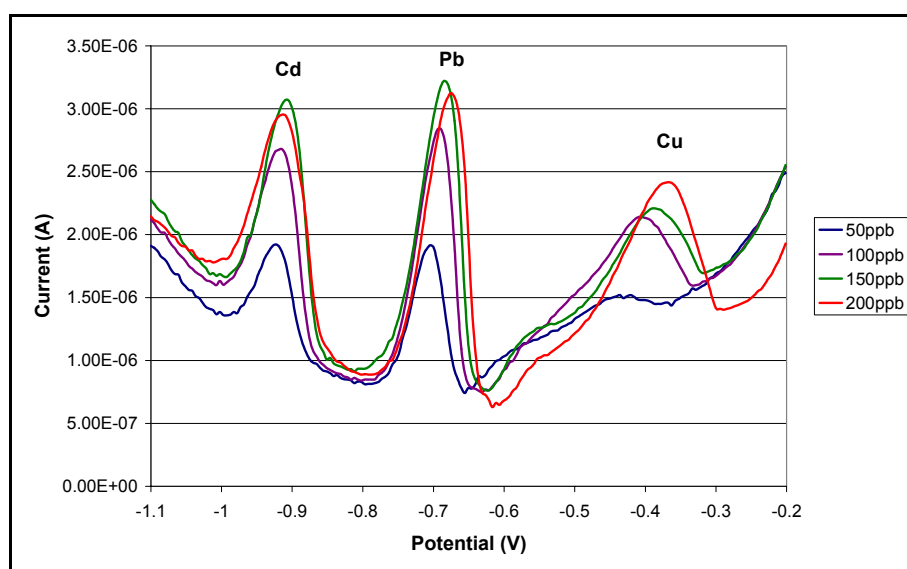
Figure 4.22: Voltammograms acquired via the Robotech II unit for 50, 100, 150 and 200 ppb Cd, Pb and Cu in 200 ppm mercuric (II) nitrate, 1% nitric acid and 0.1M KCl. Instrument parameters: deposition potential: -1.1V; deposition time: 165s; scan rate: 10mV/s

There is a substantial increase in the peak heights from 50 to 100 ppb and then from 100 to 200 ppb. However, the 150 ppb peak appears to be more intense than the 200 ppb peak. This can be attributed to the SPE itself seeing as only one measurement per sample was performed.

Finally, for automation to be successful, the system must be able to operate for long durations as specified by the operating time and sampling frequency. The system was thus set up to do so by connecting a single carbon-ink screen-printed electrode to the Autolab PSTAT10 and placing the electrode in a beaker containing 200 ppb cadmium, lead and copper in 200 ppm mercuric (II) nitrate and 0.1M KCl. DPASV voltammograms were successfully acquired over a period of 24 hours.

4.4.4.3 The Direct Acquisition of DPASV Voltammograms with LabVIEW

It would be highly advantageous to be able to omit the acquisition of the DPASV voltammograms via the GPES software which controls the Autolab PSTAT10 potentiostat by constructing the equivalent interface in the LabVIEW environment. This is thus feasible seeing as other workers have constructed specific modules in order to acquire and process data (Avdikos *et al.,* 2005).

As the Autolab potentiostat was connected to the PC via the ISA slot, preliminary investigations were performed in order to determine whether such a module could be constructed to control the potentiostat. This involved constructing a monitoring module to capture the signals transferred to and from the PC via the ISA slot on all available channels during the acquisition of a voltammogram. PCA score plots indicated that of the eight channels, the first, second, and fifth channels were involved with the acquisition of the signal; the sixth channel was involved in the deposition stage. A Matlab script was then written to convert the captured data from the first, second and fifth channel and construct a voltammogram (Figure 4.23).
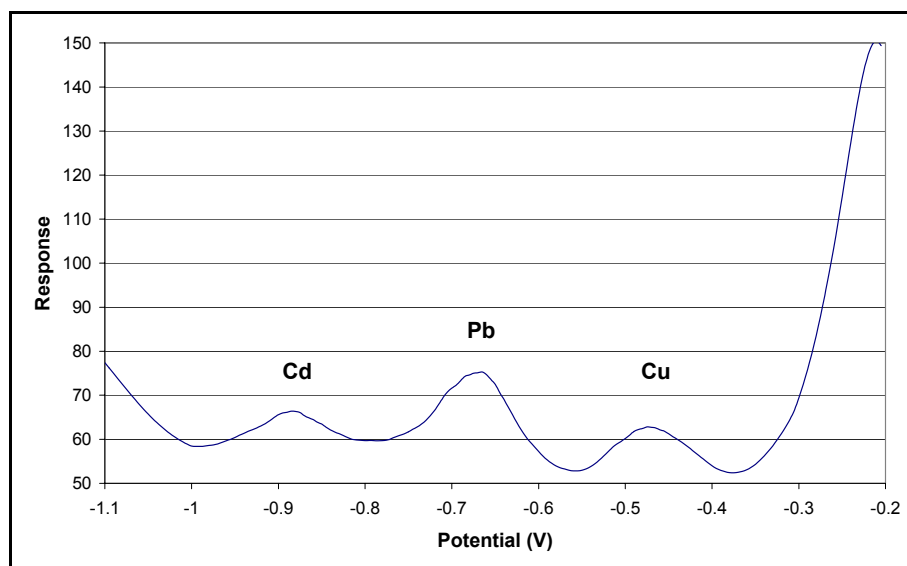
Figure 4.23: Calculated voltammogram (200 ppb Cd, Pb, Cu) from the measured raw data points contained in Channels 1, 2 and 5 (dep time: 80s; dep pot: -1.1V; scan rate: 10mV/s; modulation amplitude: 50mV; interval time: 0.5s; modulation time: 50ms; range: -1.1 to -0.2V)

The "voltammogram" in Figure 4.23 was acquired by calculating the difference between channels 1 and 5, and then subtracting channel 2 from this difference. Three distinct peaks are visible which could correspond to cadmium, lead and copper respectively.

There is thus strong evidence that a module can be constructed. However, it must be borne in mind that it is unclear how the electronic signals are conveyed within the ISA card (supplied by Eco Chemie) as well as within the hardware controller. Furthermore, it is not known how the GPES program processes the data/signals exchanged between the PC and the hardware controller. Another point to overcome is how to inform the hardware controller to apply a deposition potential of -1.1V or -0.7V. Further analyses of the generated data files are thus warranted.

There is thus great potential for the development of an automated data acquisition system coupled with data pre-treatment and multivariate calibration techniques; models are developed in the laboratory, exported to the PDA which then acquires voltammograms in the field and predicts the amounts of the target analytes after application of a standardisation routine. This therefore satisfies the requirements for at-line analysis.

4.5 Conclusions

Although disposable screen-printed electrodes offer the possibility of at-line analysis, flow injection systems, as described in Section 4.2.1, applied to the ARTDEMO project would provide "real-time" readings to feed into an overall process management protocol. An alarm can be immediately raised should problems, such as contamination in the form of an excess of metal ions, arise. However, following the raising of an alarm, at-line analysis employing a PDA and a field-based instrument to confirm any problems will still be preferable to the laborious process of sampling, transportation to a central laboratory and finally measurement on a laboratory-based instrument.

Initial attempts to incorporate chemometric techniques via a custom-built PDA application have proved successful. This follows from the successful implementation of the instrument standardisation techniques of direct standardisation and piecewise direct standardisation with additive background correction. It has been reported that both range-scaling and mean-centring offer improved predictions of both voltammograms acquired in the laboratory, and on the PDA via instrument standardisation.

Regarding carbon-ink screen-printed electrodes, weighting does not improve the predictions when employed in conjunction with PDS. It has been postulated that applying no weights to the voltammograms acquired on the PDA decreases the efficiency of the PDS modelling. This therefore implies that the weighting must be applied to all datasets. However, with regards to the gold-ink screen-printed electrodes, instrument standardisation has not been successful. This is mostly due to the poor qualities of the acquired voltammograms attained via the PDA. This implies that further investigations are warranted should the quantitative determination of the presence of arsenic, mercury or copper in the "field" via a PDA be a requisite.

Initial attempts at developing an automated sample preparation and screen-printed electrode dispensing unit have proved successful. The "Robotech II" program constructed in LabVIEW is currently able to control the sample preparation unit,

communicate with an external instrument (the Autolab PSTAT10), acquire the voltammograms and store the relevant data over a designated period of time. The implementation of the dispenser circuit into the main "Robotech II" program has also proved successful. The next stage of the development would involve four steps:

1. The construction of the dispenser itself
2. Replacing the Autolab PSTAT10 controller with either the in-house custom-built multi-potentiostat device or another specially constructed device which can also be controlled by the LabVIEW program.
3. Finalisation of the "Robotech II" program
4. Patenting and marketing the package

The thesis has so far reported on the successful application of multivariate calibration regression to electrochemical data, namely differential pulse anodic stripping voltammograms. However, for a complete and robust system the same chemometric tools must be applicable to the screening and quantitative determination of organic analytes in the environment. Chapter 5 presents such an application to polynuclear aromatic hydrocarbons (PAHs), in particular anthracene, phenanthrene and naphthalene.

CHAPTER 5:

APPLICATION OF MULTIVARIATE CALIBRATION ON 2D AND 3D FLUORESCENCE SPECTRA FOR THE QUANTITATIVE DETERMINATION OF ANTHRACENE, PHENANTHRENE AND NAPTHALENE

5.1 Overview

This chapter focuses on applications of the data analysis package developed in Chapter 2 for the quantitative determination of anthracene, phenanthrene and naphthalene via emission spectra (2D) and excitation-emission matrix (EEM) spectra (3D). Examples of applications of chemometrics to fluorescence spectra for the quantitative determinations of specific target components are given. The main focus of the discussion will be on partial least squares (PLS) regression. Other multivariate calibration techniques are also discussed. The materials and methods employed will be detailed followed by an in-depth discussion of the results obtained. Finally, an overall conclusion is drawn.

5.2 Introduction

Fluorescence spectroscopy has become a widely used method of analysis over the last two decades. The general theory and background of the application has already been given in Section 1.4.3. The main advantages have already been stated such as relative cheapness in instrumentation, high sensitivity, and non-destructive nature of the analytical method itself. Emission, excitation and synchronous fluorescence spectroscopy are some of the variants which will be discussed below. These can lead to the formation of 2D fluorescence data, whilst excitation-emission matrices (EEMs) lead to the formation of 3D fluorescence data, and thus an overall generation of huge amounts of data which can be dealt with adequately with chemometrics.

The application of chemometrics to fluorescence spectroscopy has also seen a substantial growth in recent years. Some of these applications are discussed below, in

particular to organic data, mainly originating from pharmaceutical and environmental sectors. These include the application of fluorescence spectroscopy in conjunction with partial least squares (PLS) regression, in addition to other chemometric techniques such as artificial neural networks (ANNs), and principal component analysis (PCA). Finally, a brief discussion will follow on the application to polynuclear aromatic hydrocarbons (PAHs).

5.2.1 Fluorescence and PLS

As far back as the early eighties, the application of PLS to fluorescence and other spectroscopic and electro-analytical techniques has been investigated. PLS had been employed by Wold *et al.,* as far back as 1983 to resolve severely overlapping fluorescence spectra originating from a mixture of humic acid, lignin sulphonate and an "optical whitener" (Sjostrom *et al.,* 1983). They also compared PLS against PCR and ridge regression (RR) which involves the addition of "a small constant to the diagonal elements in the moment matrix before its inversion". PLS performed much better than the other two techniques. This was mainly due to the ability of PLS to take the concentration (*Y*) matrix into account. These authors were in fact one of the first scientists to apply PLS to fluorescence emission spectra.

Mixtures containing phenol, *o*-cresol, *m*-cresol, and *p*-cresol were resolved with excitation fluorescence spectroscopy and PLS (del Olmo *et al.,* 1996). These compounds are used as starting blocks for the synthesis of more complex compounds such as herbicides and insecticides. Their spectra overlap considerably which is a common "downside" to fluorescence spectroscopy. Fortunately, PLS regression was able to assist in resolving the spectra. The PLS1 algorithm was employed, which implies that separate models were built for each of the four components.

The authors had acquired three different datasets: emission spectra (constant excitation wavelength of 219 nm); excitation spectra (constant emission wavelength of 298 nm); and synchronous spectra (constant difference between emission and excitation of 80 nm). Although synchronous fluorescence spectra tends to give better spectra with refined peaks, the authors found that better resolution with the PLS1 algorithm occurred with the data acquired via the excitation fluorescence spectra.

Savitzky-Golay filtering was unable to assist in the resolution of the overlapping spectra with regards to emission and synchronous fluorescence. Furthermore, the optimum latent variables calculated for each component were derived via a statistical F-test on the PRESS values (Section 1.5.5.2) which calculates the F-ratio between the impending PRESS values and the minimum PRESS value (Haaland & Thomas, 1988a; Haaland & Thomas, 1988b).

The PLS1 algorithm was also employed in the simultaneous determination of the overlapping fluorescence spectra of naproxen (a substitute for aspirin), salicylic acid and acetylsalicylic acid (Navalon *et al.,* 1999). Without the appropriate chemometric tools, the components within the given sample would require separating out via classical means such as gas or liquid chromatography. It was also reported that finding "a common emission wavelength" for the acquisition of an excitation spectrum was very difficult with regard to the three components due to the inherent loss of sensitivity of one component with respect to the others. Contrary to (del Olmo *et al.,* 1996), the emission spectra were recorded instead. The F-test ((Haaland & Thomas, 1988a; Haaland & Thomas, 1988b)) was employed to determine the optimum latent variables for each component. The authors also found that mean-centring the data significantly improved the predictive abilities of the PLS1 models (Navalon *et al.,* 1999).

Fluorescence spectroscopy coupled with PLS has been applied to a wide variety of pharmaceutical formulations. The gain in popularity is understandable when one considers the traditional time-consuming and expensive methods employed, such as GC-MS and capillary zone electrophoresis, when determining the presence and subsequent amounts of components in given samples (Sorouraddin *et al.,* 2005). Recent years have seen the growth in the application of PLS and other chemometric tools in such fields as environmental, clinical, and biomedical including the analysis of drugs (Martos *et al.,* 2000).

The preparation of pharmaceutical products involves many compounds. Many steps can be employed, and each step may require analysis to ensure that fabrication and processes are conforming to the appropriate protocols. If each step contains a mixture of a given number of components, it is extremely important to be able to resolve and

quantify these components, in the least-expensive but most productive and conclusive manner. Fluorescence spectroscopy coupled with PLS meets these criteria. Pyridoxine, acetylsalicylic acid and codeine are found in many pharmaceutical products in a differing number of combinations. PLS could successfully predict the amount of each component in the given pharmaceutical products. This was due to its ability to resolve the overlapped fluorescence spectra obtained for each product (Martos *et al.,* 2000).

Human urine exhibits fluorescent properties. One way of assessing the efficiency of a given drug such as an antibiotic is to analyse the content in urine. One such group of antibiotics employed is the quinolones (Espinosa-Mansilla *et al.,* 2004). However, a marked improvement in the bactericidal properties of the quinolones is observed when 6-fluoro and 7-piperazinyl groups are chemically added to the molecule. It was reported that "conventional" 2D analysis, for example the emission spectra, was not possible due to the extreme overlap attained. However, when the EEM was acquired for each component, the authors were able to ascertain the optimum excitation wavelengths leading to the respective emission spectra.

Caffeine is also employed in many pharmaceutical preparations alongside acetylsalicylic acid (Moreira *et al.,* 2005). HPLC is normally employed to determine the amounts, but at a relatively high cost. Two PLS1 models were developed for the simultaneous determination of the caffeine and acetylsalicylic acid via emission fluorescence spectroscopy. In this instance, the analytes were "pulverised" and mixed together in different amounts ranging from 50-170mg/g for acetylsalicylic acid and 5-20mg/g for caffeine. These were "mixed with lactose, maize starch, talc and magnesium stearate in the 70:15:10:5 or 80:10:7:3 w/w proportions". Although there was considerable spectral overlap, it was concluded that the application of PLS was sufficient to satisfactorily resolve the two components in a given "pharmaceutical formulation" without the need to perform any preparatory steps in the laboratory.

The discussion has so far been focussed on organic compounds employed in the manufacture of pharmaceutical products. Little has so far been said of the detection, whether on-line, at-line or off-line, of organic compounds present in the environment. Organic compounds of interest are the polynuclear aromatic hydrocarbons (PAHs),

which will be discussed in greater detail later (Section 5.2.3). However, these PAHs will be a fraction of the natural organic matter (NOM) found for example in rivers and lakes which can be employed as sources for drinking water (Marhaba *et al.,* 2003). It was stated that NOMs are "complex mixtures of organic materials".

Dissolved organic carbon (DOC) can give a good indication to the NOM present in drinking water. Fluorescence spectroscopy is employed since some of the organic materials will contain fluorophores. The PLS models generated were able to predict the DOC in the given river water samples. It was also suggested that "other parameters linked to the organic content in water such as chlorophyll-a, and chlorine demand" could also be determined. However, it was equally stressed that the detection of outliers was of great importance. Glyphosphate, which is employed as a herbicide to destroy weeds, was successfully determined in the presence of its metabolite (aminomethylphosphonic acid) in river water via fluorescence emission spectroscopy and the PLS1 modelling algorithm (Meras *et al.,* 2005). However both components must first undergo a derivatisation reaction with 4-chloro-7-nitrobenzofuran (NBD-Cl) in order for them to attain fluorescent properties. However, this only takes five minutes compared to one hour in another method suggested by other workers (Colin *et al.,* 2000). Good recoveries were obtained for glyphosphate (83 to 94%) and aminomethylphosphonic acid (104 to 120%) (Meras *et al.,* 2005).

It must be stressed that in the references detailed above, the PLS optimum latent variables were selected by the F-test on the PRESS values (Haaland & Thomas, 1988a; Haaland & Thomas, 1988b).

5.2.2 Fluorescence Incorporated with Other Chemometric Tools

Although PLS regression has been shown in many instances to be the most efficient chemometric tool in calibrating analyte concentrations, there are a number of other tools that can also be employed, either instead of, or in addition to PLS, such as PCA, and ANNs. In Section 5.2.1, it was mentioned that dissolved organic carbon (DOC) in river water was determined via fluorescence spectroscopy and PLS. The same can be done for the DOC in sea-water (Persson & Wedborg, 2001). A good description of the three groups of humic substances was given. The first is *humin* which is "insoluble in water at all pHs", *humic acid* which is not soluble at < pH 2, and *fulvic acid* which is "soluble at all pHs".

A detailed explanation was given by the authors on emission, excitation and synchronous fluorescence spectroscopy; PCA (in particular the interpretation of scores and loadings); and the fundamentals of PLS regression. It had been emphasised that samples originating from different sources would not contain the same concentration of chromophores which would thus affect the single-channel emission spectrum. This was why EEMs were generated instead. Furthermore, the spectra required correcting with Rhodamine B dissolved in ethylene glycol at a concentration of 8g/l and "the reflection spectra of barium sulphate", which are due to such phenomena as the aging of the lamp.

The authors also stressed that the presence of certain metal ions can affect the fluorescence spectra. For example, iron and copper can lead to quenching of the spectra whilst magnesium and calcium also "have a slight affect on the spectra". The PCA models were able to class which samples originated from specific locations, for example, below 10m in the Baltic Sea. Furthermore, the humic substances were classed in terms of terrestrial or marine origin. The PLS models were able to illustrate the effects of salinity. For example, there was a greater scattering of samples in the score plot due to the mixing of humic substances from marine and terrestrial origins.

Humic substances, which originate from the "degradation of plant and animal organic matter", were also studied with fluorescence spectroscopy in conjunction with multivariate curve resolution alternating least squares (MCR-ALS) instead of PLS

(Antunes & Esteves da Silva, 2005). In fact, it was described as being the best tool for analysing the generated EEMs since "they do not follow a rigorous multilinear model". MCR-ALS relates the concentrations of the humic substances to the trends of the fluorescence intensities in addition to the pure emission and excitation spectra being recovered from the EEM which lead to the extraction of information regarding the "major fluorophore structures of the humic substances". In other words, the number of fluorophores in the EEM of the respective humic substance can be identified.

As mentioned in Section 1.5.4.6, the major advantage of ANNs is the ability to model both linear and non-linear data. The latter can also be observed in fluorescence data, such as in the analysis of Rhodamine B, butyl Rhodamine B and Eosin B (Zhang *et al.,* 1997). PLS was shown to be ineffective as poor models were created. Successful modelling occurred with ANNs. In addition to this, the authors constructed an algorithm which "pruned" the hidden nodes when combined with back propagation. The *pruning* itself involved determining the *rank* and the respective *singular value decomposition* (SVD) matrices. In essence, the optimum number of hidden nodes is determined, and this would lead to the number of *epochs* being reduced, which itself would ensure that the data was not being over-fitted.

ANNs have also been employed to select wavelength variables from a dataset containing a large number of variables in each spectrum. This in essence is a form of data pre-treatment. The variant of ANNs employed is *Kohonen* (K-ANN). A square "Kohonen map" is generated. The ANN is trained and after a certain number of epochs, the wavelengths are grouped into certain cells due to their similarities. The selected wavelengths are those that are closest to the centre of the neuron. And thus, 73 wavelengths were selected out of 151 for a 10×10 K-map for the PLS modelling of phenol, *o-*, *m-*, and *p*-cresols (Todeschini *et al.,* 1999), and 70 wavelengths were selected out of 361 for a 10×10 K-map for the PLS modelling of three "non-steroidal anti-inflammatory drugs" (Capitan-Vallvey *et al.,* 2000). In both cases, better PLS models were generated than when the entire spectrum was chosen.

Other chemometric tools employed for the selection of "important" wavelengths include PCA, orthogonal signal correction (OSC) and wavelet analysis (Eriksson *et*

al., 2000). Each will remove variables (wavelengths) that will not correlate with the corresponding Y-variables, in other words, that will not contribute positively to the model. In essence, these pre-treatment techniques are performed in order to "enhance the predictive power" of the models, for example, PLS. It was foreseen by the authors that OSC in combination with wavelet analysis (compression of the number of variables (wavelengths) in the raw data) would lead to more powerful PLS models, even though the authors had to conclude on this occasion that the best PLS model was attained with wavelet analysis only.

Other Y-properties that are modelled as well as concentration values are ash content (as in the production of white sugar), "thickness of a tablet coating", polymer strength, "ethanol content in wine, or the viscosity of a solution". A thorough description was given with regards to PLS modelling, and signal correction such as OSC and wavelet analysis (WA). It was stated that the PLS weights (*W*) play an important role in the modelling for they interpret which "spectral variables are *influential* for the modelled responses, and which are not". The PLS scores can be used to identify strong outliers, whilst plots of the residuals can spot "weak" outliers.

Multiplicative signal correction (MSC) was also looked at but was found to remove certain X-variables that contributed significantly to the model. In addition, the OSC algorithm failed to completely remove the Rayleigh scattering, as had been hoped. The authors had not previously removed it, for example by background subtraction, due to the need to observe whether OSC could actually assign the peaks as irrelevant to the model (uncorrelated with Y) and then remove it. The authors were forced to conclude that the Rayleigh scatter did hold some relevant information with regards to the data. Finally, the WA was able to reduce the number of variables from 4000 down to 250, and thus leading to an overall faster computation.

5.2.3 Application of Fluorescence to Polynuclear Aromatic Hydrocarbons

Polynuclear aromatic hydrocarbons (PAHs) have been the subject of much research over the last two decades. Due to their carcinogenic and toxic properties, great concern has been afforded them. The chemicals are released into the environment via such activities as wood and coal burning, exhaust from combustion engines (whether

petrol or Diesel), tobacco smoke, as well as other processes such as the reactions of "saturated hydrocarbons in oxygen-deficient conditions" (Ribeiro & Ferreira, 2005). In the environment, they can be found in air, soils, sediments and water. Phototoxicity has become a major concern. When exposed to sunlight, the structures of the PAHs change and in most cases become more toxic than their parent compounds. The reason for the increase in toxicity is attributed to the absorption of UV and visible radiation by the π-orbital (Kemp, 1991).

A study was carried out in 2001 in order to characterise the PAH profiles within a road tunnel in Gothenburg, Sweden (Wingfors et $al.,$ 2001). Sampling points were placed at the entrance and exits of the tunnels. The PAHs that had been adsorbed on to the collectors were treated with a mixture of methanol/water (3:1) and analysed with HPLC employing an octyldecylsilica column. A fluorescence detector coupled to the exit valve of the column detected the PAHs as they were eluted. The traffic composition had also been recorded, for example, the number of cars, vans, and lorries. These were calibrated against the PAH concentration with PLS. This implies that the generated models can later be employed to ascertain the composition of traffic in other tunnels and urban areas based on the profiles/concentrations of the PAHs detected.

As far back as the early eighties, synchronous fluorescence spectroscopy has been employed for the detection of PAHs such as anthracene (Thompson & Pardue, 1983). During this time, more attention was paid to developing the appropriate instrumentation instead of actually assessing the environmental implications of the PAHs. Today, it is a different matter. Determining the environmental impact of these PAHs is of paramount importance, and this itself has led to the development of even more sophisticated instrumentation coupled with applications of multivariate analysis techniques.

It was long understood that single-wavelength emission fluorescence spectroscopy was limited in "its ability to analyse complicated multi-component samples when they have severely overlapping emission and/or excitation spectra" (Patra & Mishra, 2002). Synchronous fluorescence spectroscopy was developed in order to increase the selectivity and avoid the need to perform separation steps. Furthermore, peaks that are

observed in normal emission spectra can be intensified with synchronous fluorescence spectroscopy.

Further improvements in the synchronous fluorescence spectroscopic method have been proposed such as derivatisation of the spectra which leads to enhancement of the identification of species. The addition of surfactants such as hexadecyl-trimethylammonium bromide also increases the fluorescence intensities. However, there are always instances where interfering peaks may also be intensified. In this regard, fluorescence quenching can be applied, for example to certain PAHs.

There is the application of multivariate calibration in the form of PLS regression to the synchronous fluorescence spectra (Ferrer *et al.,* 1998). In this instance, PLS1 and PLS2 models were constructed for the resolution of ten PAHs including anthracene, phenanthrene, naphthalene and pyrene. Three synchronous spectra were recorded for each sample at $\Delta\lambda = 10$, 50 and 100nm. However, the 10nm spectra were rejected due to no distinction being attainable between the PAHs. A number of pre-processing techniques were performed and it was found that mean-centring sufficed. Optimum latent variables were obtained via the F-test on the PRESS values (Haaland & Thomas, 1988a; Haaland & Thomas, 1988b). Overall, PLS1 models provided the better predictions but longer times were required since ten PAHs were being analysed (Ferrer *et al.,* 1998). PLS2 on the other hand was better at predicting synchronous spectra that were similar to one another.

ANNs have also been applied to the determination and resolution of overlapping synchronous fluorescence spectra of PAHs (Ferrer *et al.,* 1999). Synchronous spectra at $\Delta\lambda = 50$ and 100nm were recorded in a micellar medium and concatenated. PCA was then performed to reduce the number of variables, and hence increase the computational speed. The score matrix (14 columns) was then subjected to the ANN in back propagation mode. There were thus 14 input nodes, 10 hidden nodes, and 11 output nodes (the last node attributed to the surfactant, Brij-35). Comparisons had also been made with PLS and PCR.

With regard to the ANN, one hidden layer sufficed for the model for when more layers were added, results had not improved in addition to computational time being

increased. The F-test was performed in order to determine the optimum number of latent variables (Haaland & Thomas, 1988a; Haaland & Thomas, 1988b). Overall, the PLS model was found to be the best at prediction. However, the other calibration models had properties that were better than the PLS, for example PCR seemed to predict anthracene better whilst ANN was better at predicting phenanthrene and naphthalene (Ferrer *et al.,* 1999).

The importance of being able to simultaneously determine the presence of a number of PAHs in the environment has been discussed. Synchronous fluorescence spectroscopy has played a large role in assisting in this determination. Coupling with multivariate analysis techniques such as PCR, PLS and ANNs have assisted these determinations even further. However, there are other techniques available in determining PAHs. One recent technique involves the development of a DNA biosensor in the form of a sol-gel derived array (Doong *et al.,* 2005). This was used to detect for PAHs in water as well as serum samples.

The biosensor successfully detected naphthalene and phenanthrene in water but failed to quantify fluoranthene and benzo[*a*]pyrene. However, benzo[*a*]pyrene did show toxic effects at low concentrations in the serum. A fluorescence microscope was employed to detect the fluorescent dyes. This was coupled to a hardware controller which was connected to a PC. The LabVIEW program acquired the data and processed it. Overall, the developed biosensor was capable of detecting PAHs in both water and serum.

PAHs are toxic and carcinogenic. Their derivatives are no exception. For example, naphthalene derivatives are deemed "organic pollutants" due to originating from the degradation of pesticides (Ortega-Algar *et al.,* 2003). Two important derivatives are α-naphthol and β-naphthol (OH group attached to the 1 and 2 position on naphthalene respectively). The former is found in pesticides for it is very efficient at killing pests but at the same time has a measurable toxicity towards mammals. However, it may contain traces of β-naphthol which is far more toxic than the α-form. Ironically, the β-form is used in dyes in both the pharmaceutical and cosmetic industries.

The detection and determination of PAHs and their derivatives will continue to be an important and much researched subject, especially as certain environmental bodies continue to increase legislation, and reduce acceptable limits.

5.3 Materials and Methods

Additional background information will be given of the three main target analytes and the reason for their selection in this study (Section 5.3.1). A description of the experimental design follows for acquisition of (2D) emission spectra (Section 5.3.2) and (3D) excitation-emission matrix (EEM) spectra (Section 5.3.3). Application to a real soil sample is also described (Section 5.3.4).

5.3.1 Specific Target PAHs

The Environmental Protection Agency (EPA) has listed 16 PAHs ("Priority PAHs") to be of great concern due to their hazardous effects on the environment in addition to the health and safety of living organisms. These are acenaphthene, acenapthylene, anthracene, benz(*a*)anthracene, benzo(*a*)pyrene, benzo(*b*)fluoranthene, benzo(*ghi*)perylene, benzo(*k*)fluoranthene, chrysene, dibenz(*a,h*)anthracene, fluoranthene, fluorene, indeno(*1,2,3-cd*)pyrene, naphthalene, phenanthrene and pyrene.

Three of the PAHs employed in this study were anthracene, phenanthrene and naphthalene. Anthracene and phenanthrene are three-ringed planar molecules (Figures 5.1 and 5.2 respectively) whilst naphthalene is a two-ringed planar molecule (Figure 5.3). These make interesting compounds for study by fluorescence spectroscopy due to the distinctive peaks attained in conjunction with the determination of the ability of the multivariate calibration regression modelling algorithm being capable of successfully distinguishing between them.

5.3.1.1 Anthracene

Anthracene is extracted via fractional distillation from coal tar in which its abundance is 0.5% (Sharp, 1990). The colourless crystals of anthracene melt at 217°C; the boiling point is at 355°C. Anthracene is normally employed as an intermediate for the manufacture of anthraquinone, which involves electrophillic substitution at the "9" and "10" positions, and its derivatives. It can undergo chlorination, nitration and sulphonation. It is also employed in wood preservatives, coating materials, insecticides and in the synthesis of alizarin (a red dye) (Morrison & Boyd, 1992). The structure of anthracene is shown in Figure 5.1.
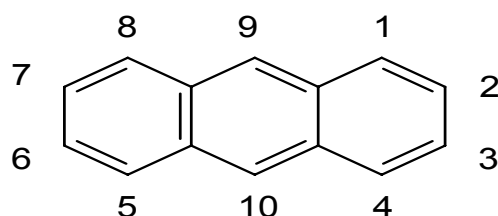
Figure 5.1: The structure of anthracene (C$_{14}$H$_{10}$)

The numbers indicate the positions where substitution occurs, for instance 1,5-dinitroanthracene indicates that a nitro group (NO$_2$) has been substituted in the 1 and 5 positions. Anthracene has also been employed in the production of smoke screens, crystals used in scintillation counters, and as an organic semiconductor (Hawley, 1987). Incomplete combustion of fossil fuels is the main cause of anthracene in the environment. It has been identified in ground and surface waters employed for drinking in addition to ambient air, vehicle exhaust emissions, cigar and cigarette smoke, smoked foods and shellfish (ATSDR (Agency for Toxic Substances and Disease Registry), 1990).

5.3.1.2 Phenanthrene

Phenanthrene also contains three fused benzene rings (Morrison & Boyd, 1992). The colourless crystals of phenanthrene melt at 100°C; the boiling point is 340°C (Budavari *et al.,* 1989). Phenanthrene is nearly insoluble in water, but is soluble in glacial acetic acid and many organic solvents. It is also obtained from coal tar via

fractional distillation. It has been identified in foods, surface and ground waters, and air. It occurs in the environment via the incomplete combustion of wood and fossil fuels, vehicle emissions, iron and steel works and incinerators. Figure 5.2 shows the structure of phenanthrene, which also possesses five resonance structures.
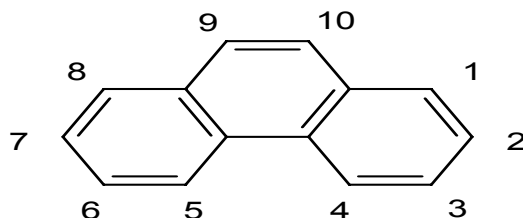
Figure 5.2: Structure of phenanthrene ($C_{14}H_{10}$)

Phenanthrene is employed in the manufacture of dyes, explosives, drugs, and phenanthrenequinone (Sax & Lewis, 1987). Phenanthrene can be absorbed through the skin (Storer *et al.*, 1984). *In vivo* studies have indicated that dihydrodiols were the primary metabolites generated via epoxidation at the 1-2, 3-4 and 9-10 carbons (Nordqvist *et al.*, 1981).

5.3.1.3 Naphthalene

The structure of naphthalene is of two benzene rings fused together (Figure 5.3). At room temperature it is in white crystalline form with a melting point of 80°C and a boiling point of 218°C. The crystals are described as having a "tarry smell" (Sharp, 1990). It too is obtained from coal tar and is most abundant at 5% (Morrison & Boyd, 1992). It is also obtained from petroleum by the removal of methyl groups from methylnaphthalene fractions with hydrogen at 750°C and 10-70atm (Sharp, 1990). Figure 5.3 displays the structure of naphthalene.
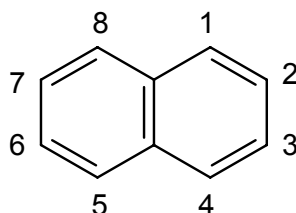
Figure 5.3: Structure of naphthalene ($C_{10}H_8$)

Naphthalene is soluble in most organic solvents and only very slightly miscible with water (Budavari *et al.,* 1989). It tends to undergo electrophilic reactions more so than nucleophilic due to its aromaticity (Morrison & Boyd, 1992). It is employed as an intermediate in addition to a starting material in industrial processes such as the manufacture of chemicals, plastics and dyes (Sandmeyer, 1981). Its main application is the synthesis of phthalic anhydride. Other applications include intermediates in the manufacture of celluloid, synthetic resins, lubricants and solvents. As a product, naphthalene is directly employed as an insecticide and an antiseptic for the intestines.

Naphthalene enters the environment much in the same way as anthracene and phenanthrene but also from mothballs. Naphthalene is said to react with hydroxyl radicals in the atmosphere leading to degradation. In waters and soil, naphthalene does not bioaccumulate over time (ATSDR (Agency for Toxic Substances and Disease Registry), 1990). Exposure to naphthalene can lead to neurological, gastrointestinal, kidney and liver problems. It can seriously irritate the skin and eyes, and be lethal to adults at amounts of 5 – 15g, and to children at amounts of 2 – 3g (Sandmeyer, 1981).

5.3.2 Acquisition of 2D Fluorescence Spectra

5.3.2.1 Experimental Design and Sample Preparation

Training samples were prepared in different concentrations of anthracene, phenanthrene and naphthalene (0, 100, 200, 400 and 600 ppb) in HPLC grade acetonitrile (Fisher, UK) via the aid of a full factorial experimental design generated by an in-house program developed in LabVIEW (National Instruments, Austin, TX, USA). Validation samples were also prepared in different concentrations of the above PAHs (150, 300 and 500 ppb) in HPLC grade acetonitrile (Fisher, UK). Figure 5.4 shows the overall design incorporating both sets.
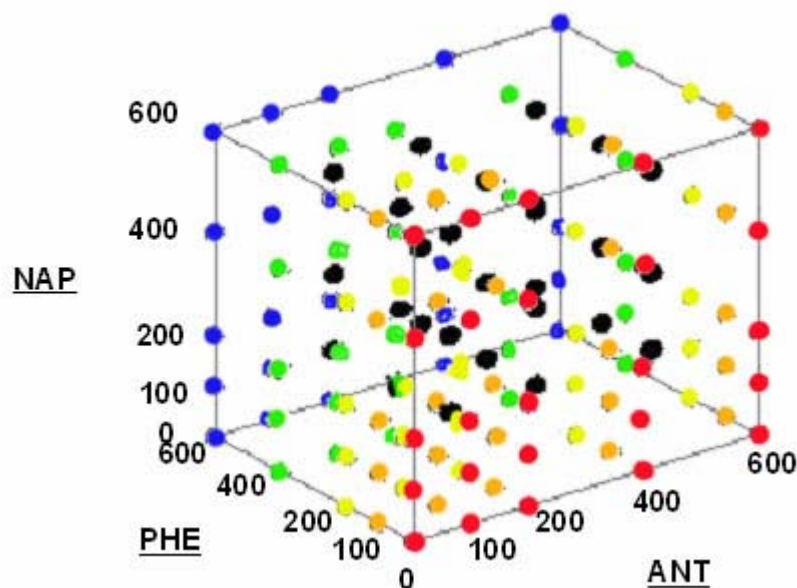
Figure 5.4: Overall experimental design: Training set (coloured with respect to PHE concentration) consists of five levels and three factors resulting in the preparation of 125 samples containing anthracene (ANT), phenanthrene (PHE) and naphthalene (NAP) at concentrations ranging from 0, 100, 200, 400 and 600 ppb; Validation set (●) consists of three levels and three factors resulting in the preparation of 27 samples containing ANT, PHE and NAP at concentrations ranging from 150, 300 and 500 ppb

In addition to the above, a set of "unknown" samples were prepared. These were excluded from any training and validation procedures and were employed to determine the strengths and efficiencies of models generated. Table 5.1 displays the concentrations of the three target PAHs in each of the five samples.

Table 5.1: Concentration values of anthracene (ANT), phenanthrene (PHE) and naphthalene (NAP) in the five "unknown" samples

Sample	ANT	PHE	NAP
UNK1	400	150	600
UNK2	500	300	400
UNK3	200	400	600
UNK4	150	500	600
UNK5	0	0	500

The PAHs were purchased from Dr Ehrenstorfer GmbH (via QMx Laboratories Ltd., Thaxted, UK).

5.3.2.2 Data Acquisition

Emission spectra of the randomised training and validation samples were acquired in triplicate on a FluoroMax-2 fluorimeter (ISA Instruments S.A (UK) Ltd., Jobin Yvon-Spex, Middlesex, England) via the Datamax software (v2.20; GRAMS/32 v4.11 level II) also by Jobin Yvon-Spex on a personal computer operating under Windows 2000 (Microsoft, US) over a period of four days. Table 5.2 displays the operating parameters of the fluorimeter.

Table 5.2: Instrument parameters employed for the acquisition of emission spectra

PARAMETER	VALUE
Start Scan (nm)	200
End Scan (nm)	500
Excitation Signal (nm)	254
Increment (nm)	1
Integration Time (s)	0.1
Slit width (mm)	1.175
Lamp	Xe (150W)
Cell	1 ml quartz cuvette

The "unknown" samples detailed in Table 5.1 were also acquired in triplicate. In between measurements, the cell was rinsed with HPLC grade water followed by HPLC grade acetonitrile.

Prior to acquiring the data on each respective day, the FluoroMax-2 fluorimeter required calibrating after initialisation of the hardware and the software. A set procedure was followed in order to achieve this (ISA, 2004). In short, the excitation grating was first calibrated (maximum peak at 467 ± 0.5nm), followed by the emission grating (maximum peak at 397nm).

5.3.2.3 Data Processing

The SPC files generated by the Datamax program were converted into standard text files by the GRMSCNVT program (also by Jobin Yvon-Spex). The SPC data files contain both a text and a binary portion. The text portion contains information regarding the instrument parameters such as the start and end wavelengths employed.

The binary portion contains the intensity values and corresponding wavelengths. The GRMSCNVT program converts the binary portion into a separate text file which can be read by any text editor. The text file will contain two tab-separated columns: the first the wavelength values; the second the intensity (in counts per second) at the respective wavelength. The appropriate text files were combined into respective data matrices via another in-house custom-built program developed in the LabVIEW environment. This program was also able to simultaneously construct the concentration matrices (Y) from the respective filenames which contain the following identification labels as illustrated in Table 5.3.

Table 5.3: Examples of sample IDs and corresponding concentration (in ppb) values for anthracene (ANT), phenanthrene (PHE) and naphthalene (NAP) in the training set

Sample ID	ANT (A)	PHE (P)	NAP (N)
APN602	600	0	200
APN114	100	100	400
APN246	200	400	600

Data pre-treatment, modelling, validation and predictions were performed via the custom-built data analysis package described in Section 2.3.3. Section 3.3.5 summarised the capabilities of the package. In addition to importing DPASV voltammograms, the package is capable of importing fluorescence spectra. Flow-diagrams of the functionalities of the package are detailed in Appendix A3.

5.3.3 Acquisition of 3D Fluorescence Spectra (EEMs)

5.3.3.1 Sample Preparation

The samples analysed here were identical to those analysed with the FluoroMax-2 fluorimeter (Section 5.3.2).

5.3.3.2 Data Acquisition

Excitation-Emission matrix (EEM) spectra were randomly acquired in triplicate in the same manner as in Section 5.3.2.2 on a prototype SPEX 3D spectrofluorometer (ISA Instruments S.A (UK) Ltd., Jobin Yvon-Spex, Middlesex, England) incorporating a

charge-coupled device (CCD) imaging camera (ISA SPECTRUMONE, Jobin Yvon-Spec, Horriba Group, UK). It is unique as it contains no moving parts and is thus portable. The instrument was connected to a personal computer with a 750MHz Athlon processor (AMD, USA) and 128MB RAM (now upgraded to 384MB) running Windows 98 (Microsoft, US) via a National Instruments (NI) General Purpose Interface Bus (GPIB) peripheral component interface (PCI) card (AT/GPIB-TNT (PnP), National Instruments, US). The instrument was controlled by Datamax software (v2.20; GRAMS/32 v4.11 level II) also by Jobin Yvon-Spec. A sub-program within the main program (CCD) was employed to acquire and store the EEM in a "Galactic" SPC file.

Table 5.4 shows the operating parameters for the SPEX 3D spectrofluorimeter.

Table 5.4: Instrument parameters employed for the acquisition of EEMs

PARAMETER	VALUE (Pixel)
Emission (X) Start	227.7 nm (1)
Emission (X) End	724.5 nm (511)
Emission (X) Resolution	1.0
Excitation (Y) Start	73.9 nm (1)
Excitation (Y) End	691.4 nm (511)
Excitation (Y) Resolution	1.2
Exposure Time	1.0 s
Slit width (mm)	0.05
Lamp	Xe (75W)
Cell	1 ml quartz cuvette

The "unknown" samples detailed in Table 5.1 were also acquired in triplicate. In between measurements, the cell was rinsed with HPLC grade water followed by HPLC grade acetonitrile.

Prior to acquiring the data on each respective day, the SPEX 3D spectrofluorimeter was employed to acquire a set number of EEMs of HPLC grade water (Fisher, UK) by filling a quartz cell (1cm path length). These EEMs would later be employed to calibrate the emission wavelengths after the entire sample EEMs had been acquired.

5.3.3.3 Data Processing

As with the FluoroMax-2 described in Section 5.3.2, when an EEM is acquired on the SPEX 3D spectrofluorimeter, a Galactic SPC file is also generated but is far more complex than the 2D version. Each SPC file has dimensions of 512×511. The first row contains the emission wavelength values corresponding to each column in the EEM. Text files cannot be generated directly from this file. Furthermore, an additional complication is that the wavelength parameters must be re-calibrated to coincide with the wavelength values of specific peaks in the FluoroMax-2 data.

The normal procedure for converting the EEM SPC file is to first generate SPC sub-files, which in this case would number 511. These 511 sub-SPC files would be converted into the corresponding text files. The calibration is applied (Section 5.4.3.1) and the 511 text files are combined into one large text file. This procedure was deemed inefficient due to the large number of SPC files (>400).

An alternative approach consisted of importing the data files into a commercial statistical analysis package (SIMCA 8.0, Umetrics, Sweden). Each Galactic EEM SPC file was imported into the program, and then exported as a text file. The file conversions were performed over a period of $3 - 4$ days. The text files contained additional lines of text which therefore required omitting. A program was constructed in the LabVIEW environment to accomplish this (Figure 5.5).
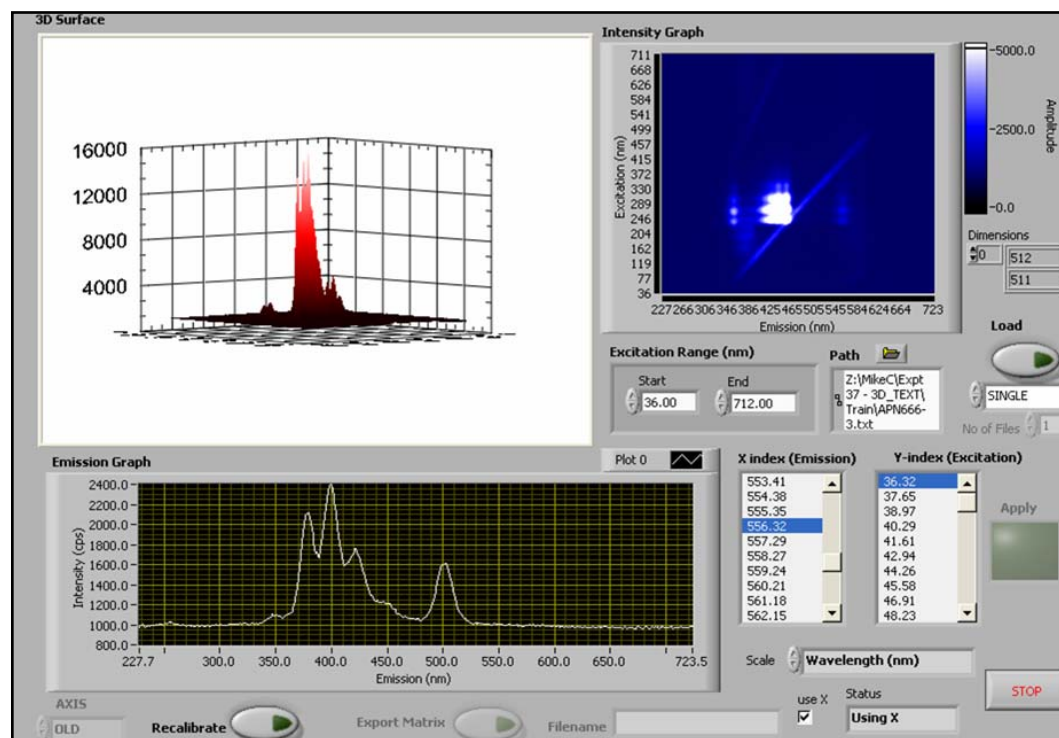
Figure 5.5: The user interface constructed in the LabVIEW environment to crop the EEM text files

The plots displayed in Figure 5.5 are from one EEM text file which had been created by the SIMCA application. The 3D surface plot displays the intensities (z-axis) against the emission and excitation wavelengths (x- and y-axes respectively). The intensity graph clearly demonstrates the active region on the CCD imaging camera. The diagonal line corresponding to the back-scattering (Rayleigh) effect is also visible. The emission graph displays the emission spectra at the excitation wavelength (or index) as dictated by the X-index (Emission) list-box. Other features of the VI are to be able to plot against CCD pixel number (index) or wavelength, as well as to view the spectra either along the emission or the excitation axis.

The custom-built data analysis package was currently incapable of directly processing the generated EEM text files. Figure 5.6 illustrates the steps involved in converting the SPC files acquired on the SPEX 3D instrument into the appropriate datasets.
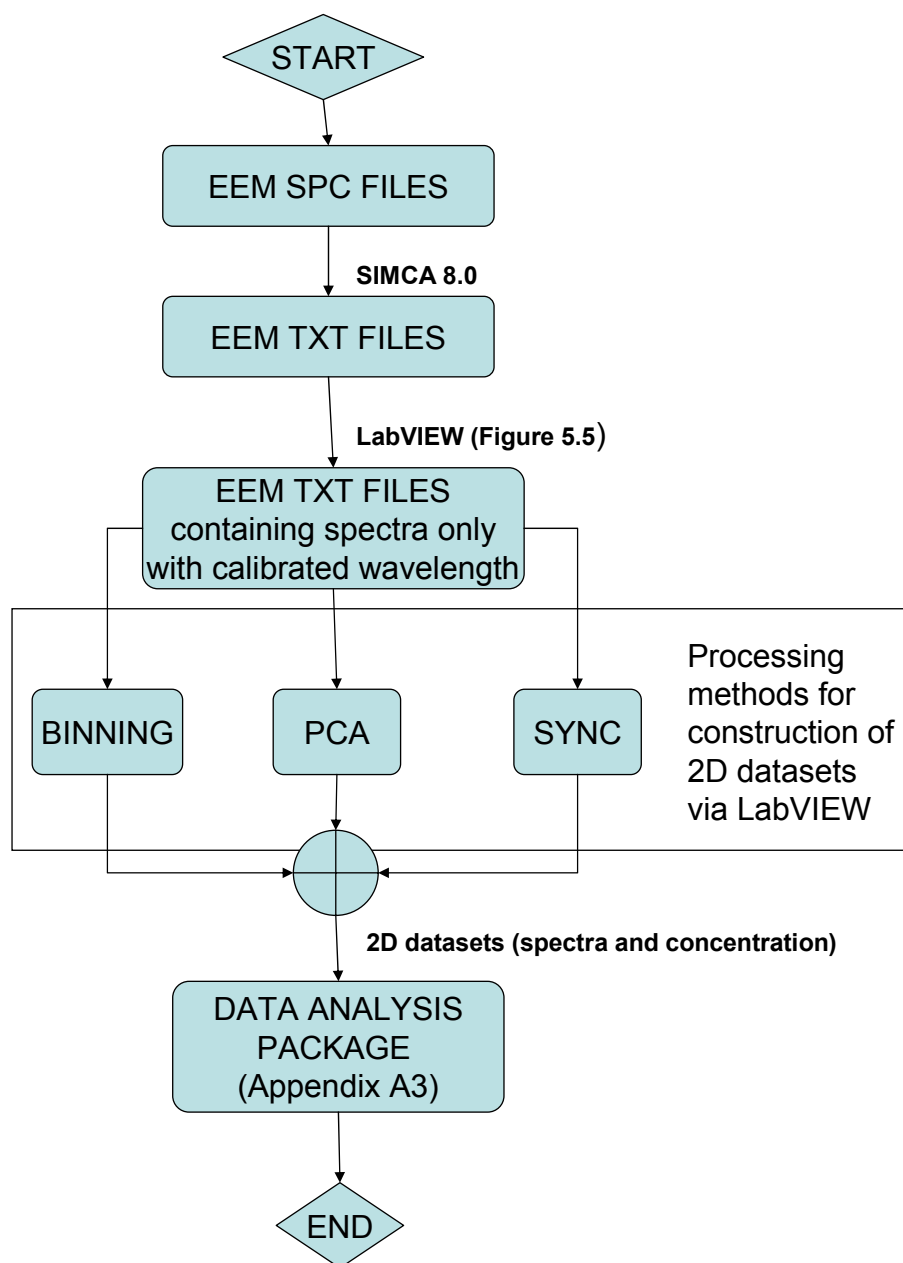
Figure 5.6: Flow-chart illustrating the conversion of the EEM files in SPC format to 2D datasets in text format via three processing methods (binning, PCA and extraction of synchronous spectra (SYNC) from the EEM files)

The processing methods illustrated in Figure 5.6 include subtraction of an averaged background EEM which is calculated from the EEMs containing no target PAH components (blanks). The processing methods are briefly explained below. Furthermore, the calibrated wavelength refers to calibration of the emission wavelengths and is explained in Section 5.4.3.1.

Binning

Hardware binning has been explained in Section 1.4.3.4. However, *software* binning can also be applied to a specified region of all the acquired EEMs. In this instance, it is defined as the summation of a given number of spectra. An experimental function was introduced which bins a given number of excitation and emission spectra from the EEM, calculates the outer product and extracts the main diagonal from it. These diagonals are iteratively combined into a 2D matrix for processing on the data analysis package.

PCA

The intensity graph in Figure 5.5 indicates that the fluorescence has been focused in one particular region of the CCD. The selected region of interest can have for example dimensions of 161 × 250. For each EEM, the region of interest was processed by binning the rows (emission spectra) and columns (excitation spectra) into respective vectors and concatenating them. Once all the EEMs had been treated in the same manner, PCA was performed with a minimum of 30 PCs. This value was chosen to explain the spectral data given that the total number of variables after concatenation exceeded 400. Ferrer had selected 14 PCs after concatenating synchronous spectra when analysing for PAHs in water samples (Ferrer *et al.,* 1999). However the concatenated spectra were similar whereas the emission and excitation spectra will be different with respect to the region of interest. Billa concatenated four emission spectra prior to PCA, however only 3 PCs were retained due to the very close resemblance of the spectra at the four different excitation wavelengths employed (Billa *et al.,* 2000).

Synchronous Spectra

The synchronous fluorescence spectra are generated from the respective EEMs from the difference between the excitation and emission wavelength (Luis *et al.,* 2004; Wiberg *et al.,* 2004).

Data Analysis

Further data pre-treatment, modelling, validation and predictions were performed via the custom-built data analysis package detailed in Section 5.3.2.3.

5.3.4 Extraction and Determination of Anthracene, Phenanthrene and Naphthalene in a Soil Sample

The soil sample employed originated from a given batch labelled as B4 (Morelands, Severn Trent Laboratories, UK). Table 5.5 lists the amounts of the target PAHs.

Table 5.5: Amounts of target PAHs in the soil sample labelled B4 (WSPC/C4507; Ref: 13010478; Order: 06015/478/AH; Morlands, Severn Trent Laboratories, UK)

Target PAH	Amount (mg/kg)
Anthracene (ANT)	0.76
Phenanthrene (PHE)	6.00
Naphthalene (NAP)	6.50

The extraction procedure detailed below was employed as a rapid method of determining qualitatively and quantitatively the presence of the three PAHs listed in Table 5.5 (Tomaniova *et al.,* 1998) The Soxhlet Extraction was not employed due to a more extensive procedure (Morrison & Boyd, 1992; Wingfors *et al.,* 2001).

- 1.47g of the soil was weighed into a 50ml plastic centrifuge tube
- 10ml of laboratory grade acetone was added
- Sonication for 40 minutes.
- Left to settle
- 600µl of clear upper phase transferred to a 15ml plastic centrifuge tube
- Evaporation in air
- 20ml of HPLC grade acetonitrile (ACN) added and shaken until residue dissolved

Measurements made in triplicate on a 2D and 3D fluorimeter (instrument details given in Sections 5.3.2.2 and 5.3.3.2 respectively).

5.4 Results and Discussion

Section 5.4.1 discusses the univariate approach to the quantitative determination of the three target analytes: anthracene, phenanthrene and naphthalene. Sections 5.4.2 and 5.4.3 discuss the multivariate approach to the fluorescence spectra acquired on the FluoroMax-2 spectrometer (2D) and the SPEX 3D spectrometer (EEM) respectively. Section 5.4.4 discusses the application of the multivariate calibration models to a soil sample.

5.4.1 Univariate Analysis

Figure 5.7A displays the individual fluorescence spectra of anthracene (ANT), phenanthrene (PHE) and naphthalene (NAP) at 400 ppb in acetonitrile, and Figure 5.7B displays the overlaid fluorescence emission spectra of a 600 ppb mixture of ANT, PHE and NAP (APN666) in acetonitrile, where it is clear to see the high precision of the FluoroMax-2 instrument.
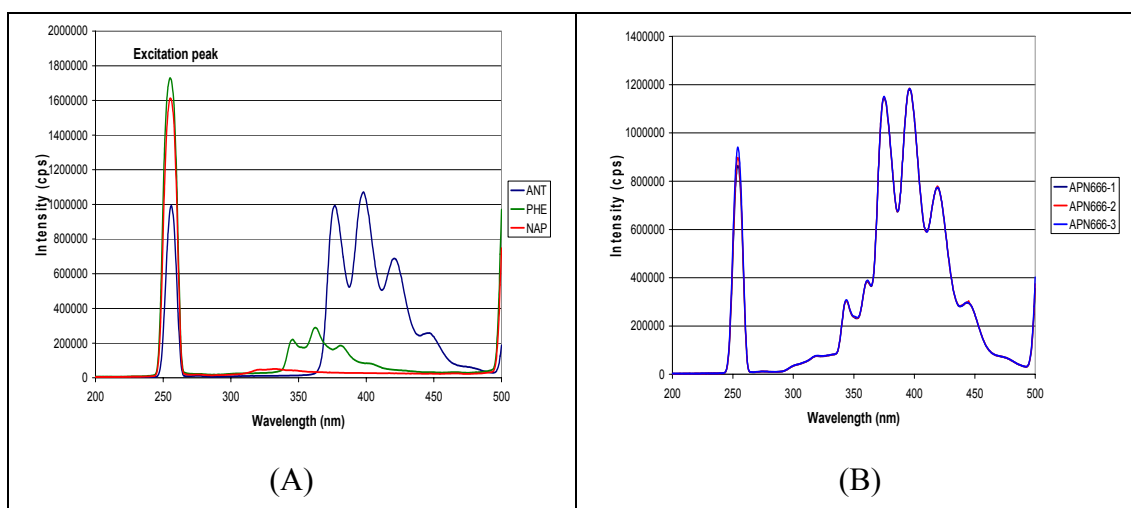
Figure 5.7: Individual emission spectra for 400 ppb ANT (blue), PHE (green), and NAP (red) in HPLC grade acetonitrile (A) and overlaid emission spectra of a mixture of 600 ppb ANT, PHE and NAP in HPLC grade acetonitrile (B). Experimental parameters as detailed in Table 5.2

All three spectra overlap to a certain extent (Figure 5.7A). The anthracene gives the most intense peaks with an almost five-fold increase in intensity compared to phenanthrene of the same concentration. Both the anthracene and phenanthrene

spectra contain three main peaks which correspond to the three fused benzene rings. As for naphthalene, the peak is very broad and weak. This implies that no major electron transitions between energy states occur and/or the excitation wavelength (254nm) is not optimum for naphthalene. Peak areas were calculated via an in-house program developed in the LabVIEW environment (National Instruments, USA). Figure 5.8 displays the calibration curve for all three analytes acquired individually.
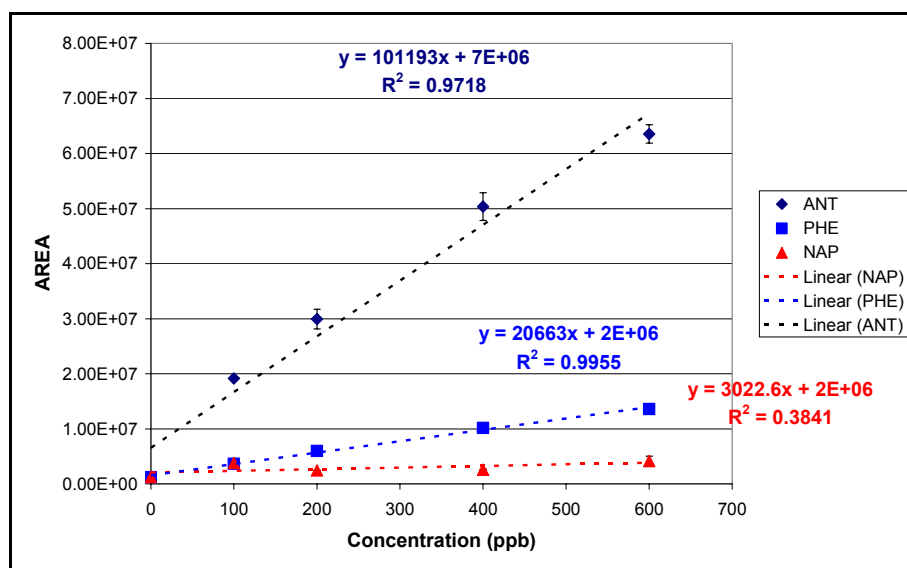
Figure 5.8: Calibration curves for anthracene (ANT), phenanthrene (PHE) and naphthalene (NAP). Experimental parameters as in Table 5.2

Phenanthrene has given the better coefficient of determination (R^2) of 0.9955 which corresponds to a correlation coefficient (r) of 0.9977. However, that of naphthalene is very poor ($R^2 = 0.3841$). With exception to the last sample containing only 500 ppb naphthalene (Table 5.1), prediction of the target PAHs in the samples could not be determined due to the overlap of the spectra of the three molecules (Figure 5.7). An average concentration of 186 ppb (n = 3) was predicted, implying an error of 63%. This can be attributable to a low signal-to-noise ratio of naphthalene. This clearly indicates the need for multivariate calibration.

5.4.2 Multivariate Calibration of 2D Emission Spectra

The SIMPLS modelling algorithm was employed throughout. This is in conjunction with the successful application in Sections 2.4 and 3.4 in addition to the review in

Sections 5.2.1 and 5.2.3. Tables 5.6 to 5.10 display the means, standard deviations, and percentage coefficient of variations for the prediction of the concentrations of the "unknown" samples displayed in Table 5.1 after applications of specific data pre-treatment techniques and multivariate calibration regression models, but in the absence of the validation set.

Table 5.6: The predicted concentrations of the unknown "real" samples with no data pre-treatment followed by modelling with SIMPLS (Optimum LV: 20 (ANT); 18 (PHE); 19 (NAP))

Sample	MEANS			STD DEV			% CV		
	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	356.46	183.28	573.18	13.91	1.71	2.50	3.90	0.94	0.44
UNK2	477.13	356.31	395.13	8.54	4.46	6.61	1.79	1.25	1.67
UNK3	180.12	462.41	575.61	4.36	1.44	1.16	2.42	0.31	0.20
UNK4	123.70	541.19	536.18	4.66	3.38	0.60	3.77	0.62	0.11
UNK5	0.00	0.00	485.76	0.00	0.00	3.73	NaN	NaN	0.77

Table 5.7: The predicted concentrations of the unknown "real" samples with range-scaling, Savitzky-Golay filtering (polynomial: 3; window: 41), followed by modelling with SIMPLS (Optimum LV: 20 (ANT); 20 (PHE); 20 (NAP))

Sample	MEANS			STD DEV			% CV		
	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	393.03	154.40	496.58	2.10	4.11	4.33	0.53	2.66	0.87
UNK2	512.07	300.38	348.31	5.21	2.91	2.40	1.01	0.97	0.69
UNK3	166.22	413.38	544.52	4.00	5.47	5.63	2.40	1.32	1.03
UNk4	94.21	488.41	536.88	14.19	4.51	17.20	15.06	0.92	3.20
UNK5	0.00	0.00	452.85	0.00	0.00	12.69	NaN	NaN	2.80

Table 5.8: The predicted concentrations of the unknown "real" samples with the excitation peak removed, and no data pre-treatment followed by modelling with SIMPLS (Optimum LV: 17 (ANT); 17 (PHE); 15 (NAP))

Sample	MEANS			STD DEV			% CV		
	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	313.63	164.08	532.80	6.59	2.26	4.20	2.10	1.38	0.78
UNK2	419.87	331.01	345.28	13.00	3.72	5.38	3.09	1.12	1.56
UNK3	141.90	455.99	538.46	9.31	5.02	5.37	6.56	1.10	0.99
UNK4	99.30	554.01	511.90	12.45	9.75	2.29	12.54	1.76	0.44
UNK5	21.01	25.53	506.42	1.69	0.95	7.48	8.05	3.72	1.47

Table 5.9: The predicted concentrations of the unknown "real" samples with the excitation peak removed, range-scaling, SG filtering (polynomial: 3; window: 41), followed by modelling with SIMPLS (Optimum LV: 20 (ANT); 20 (PHE); 18 (NAP))

Sample	MEANS			STD DEV			% CV		
	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	377.01	173.15	497.89	8.57	5.04	0.73	2.27	2.91	0.14
UNK2	490.45	316.99	338.73	25.10	2.15	6.11	5.11	0.68	1.80
UNK3	143.37	436.01	539.32	5.57	1.38	7.38	3.89	0.31	1.36
UNK4	89.11	526.24	532.54	3.28	3.67	4.06	3.69	0.69	0.76
UNK5	0.00	7.73	470.33	0.00	2.76	7.72	NaN	35.80	1.64

Table 5.10: The predicted concentrations of the unknown "real" samples with the excitation peak removed, range-scaling, SG filtering (polynomial: 3; window: 41), followed by compression with PCA (5 PCs) and subsequent modelling with NNPLS (Overall Optimum LV: 5)

Sample	MEANS			STD DEV			% CV		
	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	352.08	263.66	393.68	3.48	1.78	16.56	0.98	0.67	4.20
UNK2	363.35	462.44	410.99	4.21	2.20	24.15	1.16	0.47	5.87
UNK3	400.12	201.29	170.14	7.71	4.33	38.92	1.92	2.15	22.87
UNK4	418.54	189.22	42.23	11.01	7.17	40.49	2.63	3.78	95.86
UNK5	145.25	0	0.71	10.60	0	1.22	7.29	NaN	173.20

Table 5.11 displays the RMSEP values calculated for Tables 5.6 to 5.10.

Table 5.11: Calculated RMSEP values for Tables 5.6 to 5.10

Table	5.6	5.7	5.8	5.9	5.10
RMSEP (ppb)	35.58	42.76	56.11	47.40	272.94

Tables 5.6 and 5.7 show that the predicted concentration values are very close to the actual values with calculated RMSEP values of 35.58 ppb and 42.76 ppb respectively. Removal of the excitation peak leads to an increase in the RMSEP values (Tables 5.8 and 5.9) but the average RMSEP value for Table 5.9 is lower than that for Table 5.8 which enforces the need for range-scaling and filtering in that respect. Application of the neural network algorithm (with a 5-5-3 architecture) instead of SIMPLS has led to very poor predictions and a RMSEP value of 272.94 ppb (Table 5.10).

In contrast, inclusion of the validation set into the model leads to overall improvements in the calculated RMSEP values. Table 5.12 displays the RMSEP values for Tables 5.6 to 5.9. All predictions were performed after the incorporation of

the validation dataset into the model. In all cases, predictions were performed at the 20th latent variable for all components.

Table 5.12: Calculated RMSEP values for Tables 5.6 to 5.9 after inclusion of the validation set into the respective models

Table	5.6	5.7	5.8	5.9
RMSEP (ppb)	16.71	15.07	19.99	14.88

A vast improvement has been observed in the calculated RMSEP values by an average of 20 ppb. The removal of the variables up to and including the excitation peak in conjunction with range-scaling and Savitzky-Golay smoothing has led to the better reduction in the overall RMSEP values calculated. The opposite was observed for mean-centred data which yielded 215.19 ppb for inclusion of the validation set and 36.05 ppb for exclusion. The extreme difference observed in the calculated RMSEP values is proposed in Section 5.4.3.4.

In Section 2.4, the application of weights led to improved RMSEP values. However, the opposite was observed for the fluorescence spectra. This is attributed to the severe overlap of the three target compounds in the fluorescence spectra. Figure 5.9 displays the fluorescence spectra of a sample prior to and after weighting.
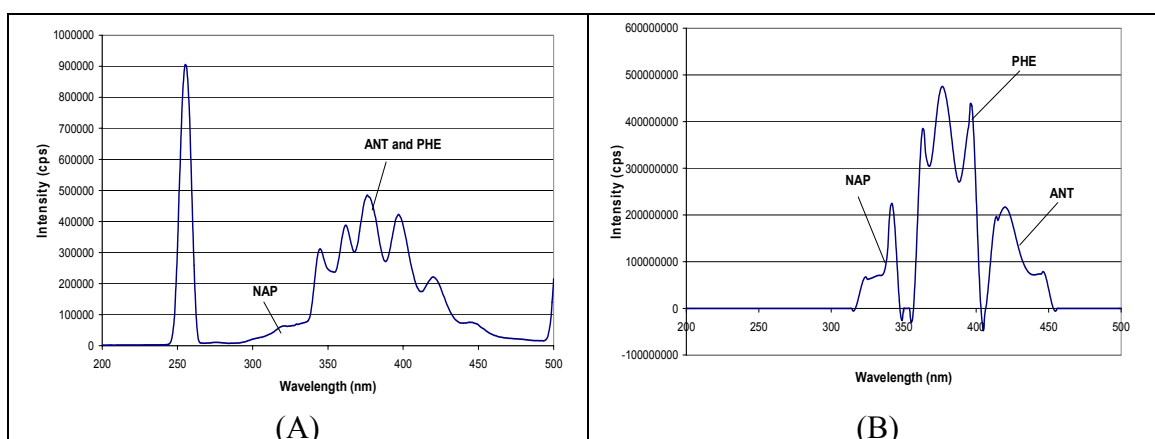
Figure 5.9: The fluorescence spectra of Sample APN166 (corresponding to 100 ppb ANT; 600 ppb PHE and NAP): Raw spectrum (A); Weighted spectrum (B) with the following conditions: high weight: 1000; low weight: 0.0001; SG wt poly: 3; SG wt win: 21.

Three distinct peaks have been generated along with an amplification of the naphthalene peak (Figure 5.9B). However, problems will arise as a result of the

overlapping nature of the peaks, especially if one of the PAH analytes is absent. For example, if phenanthrene is absent then the application of a low weight (0.0001) will severely reduce the portion of the spectra pertaining to anthracene and thus risking the loss of important information. This has thus resulted in the increase in the calculated RMSEP value up to 147.66 ppb. Further improvements must therefore be made to the weighting function so that peaks can be selectively weighted.

Should a separate validation dataset be absent, the data analysis package contains an additional option to split the training set into a separate calibration and test set. An advantage of this is a swift determination of the predictability of the generated model at the optimum latent variables as suggested by the leave-one-out cross-validation function. The last 15 spectra in the original training set were set aside as the test set. The SIMPLS model was thus constructed with the calibration set and predictions of the test set were performed. Table 5.13 shows the calculated correlation coefficients for the predicted concentrations and the actual concentrations at varying data pre-treatments.

Table 5.13: Correlation Coefficients (CC) and RMSEP values for the prediction of the test set data after modelling with SIMPLS: Set A to D are in the absence of the validation set; Set E to H are the same but have the validation set incorporated into the respective models. Sets A – H are defined in Table 5.14

PAH	ANT		PHE		NAP	
Set	CC (r)	RMSEP	CC (r)	RMSEP	CC (r)	RMSEP
A	0.99171	24.06	0.99824	13.02	0.99588	21.75
B	0.99468	19.26	0.99810	13.21	0.99594	22.00
C	0.98860	27.49	0.99804	15.10	0.99507	24.11
D	0.99283	21.70	0.99862	11.48	0.99704	19.53
E	0.99255	22.94	0.99826	13.53	0.99857	21.55
F	0.99160	24.80	0.99819	13.18	0.99704	18.23
G	0.98938	26.58	0.99803	14.64	0.99617	21.14
H	0.99360	20.75	0.99854	11.91	0.99750	17.11

Table 5.14: Description of Sets detailed in Table 5.13

SET	VALIDATION SET IN MODEL	DESCRIPTION
A, (E)	NO (YES)	No pre-treatment
B, (F)	NO (YES)	Range-scaled and SG filtered
C, (G)	NO (YES)	Removal of excitation peak; no other pre-treatment
D, (H)	NO (YES)	Removal of excitation peak; Range-scaled and SG filtered.

As can be seen, incorporation of the validation set into the respective SIMPLS models led to an improvement in the calculated RMSEP values, and in the correlation coefficients. Overall, the better sets are D and H with very high correlation coefficients and the lowest RMSEP values.

Figure 5.10 displays plots of the predicted concentrations against the actual concentrations for set D.
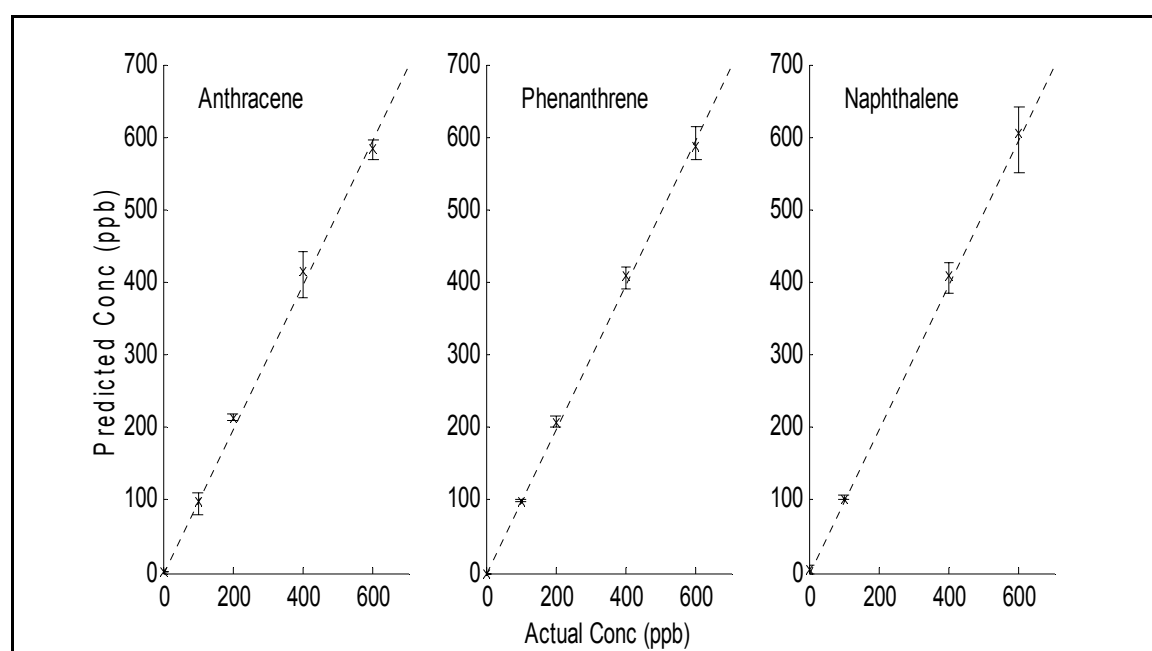
Figure 5.10: Plots of predicted against actual concentrations for the three target PAHs after removal of the excitation peak followed by range-scaling, Savitzky-Golay smoothing (polynomial: 3; window: 41) and modelling with SIMPLS

Overall, multivariate calibration has considerably improved the ability of the unknown samples to be predicted. In predicting the test set values, performing range-

scaling and filtering with the Savitsky-Golay function after removal of the excitation peak and prior to SIMPLS modelling leads to low RMSEP values.

5.4.3 Multivariate Calibration of 3D Excitation-Emission Matrix (EEM) Spectra

Section 5.4.3.1 briefly discusses the calibration of the emission wavelengths prior to processing by the different data pre-treatment techniques applied to the 3D fluorescence spectra: binning (Section 5.4.3.2), PCA (Section 5.4.3.3) and synchronous (Section 5.4.3.4).

5.4.3.1 Calibration of the Emission Wavelengths

The purpose of calibrating the emission wavelength is to ensure that the emission and synchronous spectra extracted from the EEM conform to the same respective spectra were they acquired on the FluoroMax-2 instrument. In order to achieve the calibration ten EEM spectra of HPLC grade water were acquired on two occasions over a period of four days prior to acquisition of the sample EEMs. It was therefore envisaged that all ten files could be employed to produce one calibration set instead of the need to develop two separate calibration wavelength sets and thus applying them to the samples accordingly. A basic statistical analysis was performed in order to deduce whether all ten files could be employed simultaneously (Table 5.15).

Table 5.15: Basic statistical calculations to determine whether all ten water files could be employed to form one calibration set

START	Water 1 → 5	Water 6 → 10	ALL 10 files
Average	183.4	182.4	182.9
Standard deviation	2.1909	1.5166	1.8530
%CV	1.19	0.83	1.01
END	Water 1 → 5	Water 6 → 10	ALL 10 files
Average	734.2	735.6	734.9
Standard deviation	3.6332	3.3615	3.3813
%CV	0.49	0.46	0.46

As can be seen in Table 5.15, the calculated % coefficient of variance (%CV) for the start and end wavelengths for all ten water files are not significantly different from the values calculated over the respective two occasions. This implied that all ten files

could be employed simultaneously to form average water EEM which would then be employed to adjust the emission wavelengths to all the measured standards and samples (EEMs).

5.4.3.2 Spectral Binning of Selected Wavelengths in the EEM

The intensity chart in Figure 5.5 indicated the region in which fluorescence was localised on the CCD. Two types of binning were performed. The first binning involved the summation of 91 emission spectra in that range within the EEM (index 145 ± 45). The second binning involved the summation of 91 emission (index 145 ± 45) and excitation spectra (index 120 ± 45). The outer product of these two vectors was obtained prior to extraction of the main diagonal.

Tables 5.16 to 5.19 show the means, standard deviations and percentage coefficient of variations for the prediction of the concentrations of the synthetic samples. All predictions were performed after the incorporation of the validation dataset into the model. In all cases, predictions were performed at the 20th latent variable for all components. SIMPLS was employed throughout.

Table 5.16: The predicted concentrations with no data pre-treatment for binned emission spectra

	MEANS			STD DEV			% CV		
Sample	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	384.8	150.5	556.9	1.30	1.99	24.56	0.3	1.3	4.4
UNK2	501.2	312.4	412.8	0.48	3.52	36.04	0.1	1.1	8.7
UNK3	195.6	409.6	651.9	7.09	9.91	52.15	3.6	2.4	8.0
UNK4	149.9	499.3	585.2	3.35	3.31	25.46	2.2	0.7	4.4
UNK5	0.0	0.0	504.3	0.00	0.00	13.71	NaN	NaN	2.7

Table 5.17: The predicted concentrations with data pre-treatment for binned emission spectra with range-scaling and Savitzky-Golay smoothing (polynomial: 3; window: 41)

	MEANS			STD DEV			% CV		
Sample	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	388.7	146.7	590.6	2.46	2.22	17.97	0.6	1.5	3.0
UNK2	500.0	304.6	428.2	3.58	13.16	29.92	0.7	4.3	7.0
UNK3	197.1	380.7	591.6	6.19	33.74	27.13	3.1	8.9	4.6
UNK4	149.9	495.9	570.3	0.76	11.22	27.44	0.5	2.3	4.8
UNK5	0.0	0.8	526.2	0.00	1.33	10.25	NaN	173.2	1.9

Table 5.18: The predicted concentrations with no data pre-treatment for binned excitation and emission spectra

Sample	MEANS			STD DEV			% CV		
	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	410.8	127.6	561.1	10.67	90.79	80.43	2.6	71.1	14.3
UNK2	492.5	273.8	325.6	28.07	95.25	122.75	5.7	34.8	37.7
UNK3	195.0	448.2	552.0	6.04	29.04	62.13	3.1	6.5	11.3
UNK4	163.4	498.6	570.8	7.90	27.50	25.95	4.8	5.5	4.5
UNK5	12.3	44.2	255.5	0.24	2.75	5.37	1.9	6.2	2.1

Table 5.19: The predicted concentrations with data pre-treatment for binned excitation and emission spectra with range-scaling and Savitzky-Golay smoothing (polynomial: 3; window: 41)

Sample	MEANS			STD DEV			% CV		
	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	381.1	138.5	574.4	5.51	9.39	18.70	1.4	6.8	3.3
UNK2	540.0	273.5	430.9	4.33	14.88	42.40	0.8	5.4	9.8
UNK3	198.8	364.0	516.0	5.21	31.94	45.90	2.6	8.8	8.9
UNK4	169.9	435.2	580.6	9.01	8.47	13.04	5.3	1.9	2.2
UNK5	87.9	115.0	491.1	4.71	1.92	7.89	5.4	1.7	1.6

Table 5.20 displays the calculated RMSEP values for Tables 5.16 to 5.19.

Table 5.20: Calculated RMSEP values for Tables 5.16 to 5.19

Table	5.16	5.17	5.18	5.19
RMSEP (ppb)	19.06	14.25	71.00	50.85

Range-scaling the datasets in which the emission spectra were binned in conjunction with Savitzky-Golay smoothing has given the better calculated RMSEP value (14.25 ppb). This is also an improvement from the better RMSEP value (14.88 ppb) obtained for 2D fluorescence spectra under the same data pre-treatment conditions in addition to removal of the excitation peak (Table 5.12).

Weights had been applied to both binned datasets but had failed to yield low RMSEP values. This can be attributed to the extreme overlap observed in the spectra. As a result, the weights have thus been applied at the respective index ranges when they should have not. The custom-built weighting function works better for the heavy metals due to the clear difference in the peak potential for each metal ion (Section 2.4.2). This is not the case for the three PAHs employed here, and thus it is more than likely that this "intelligent weighting" function needs revising.

It is evident that the approaches described for the binning of the emission spectra only (Tables 5.16 and 5.17) in conjunction with the SIMPLS modelling algorithm offer the best predictions, especially for sample UNK5. Furthermore, binning the excitation and emission spectra in conjunction with the SIMPLS algorithm led to difficulties in predicting the zero concentrations for anthracene (ANT) and phenanthrene (PHE) in sample UNK5 (Tables 5.18 and 5.19). Better precision was observed for predictions (low %CV values) in Table 5.16 compared against the other tables.

As had been performed in Section 5.4.2, the last 15 spectra in the binned training set were set aside as the test set. The SIMPLS model was thus constructed with the calibration set and predictions of the test set were performed. Table 5.21 shows the calculated correlation coefficients for the predicted concentrations and the actual concentrations at varying data pre-treatments and binning techniques.

Table 5.21: Correlation coefficient values for some of the predicted concentrations.
Key: RS: Range-scaling; SG: Savitzky-Golay smoothing (polynomial: 3; window: 41)

CODE	ANT	PHE	NAP
Binned emission spectra only	0.9995	0.9998	0.9982
Binned excitation and emission spectra only	0.9883	0.8922	0.7067
Binned emission spectra, RS and SG	0.9994	0.9995	0.9962

Binning the emission spectra only has given the best correlation coefficients compared with the binning of the excitation and emission spectra. It is feasible to conclude that the utilisation of the main diagonal of the outer product of the binned excitation and emission spectra is not beneficial.

5.4.3.3 PCA on Binned Spectra from the Respective EEMs

The dimensions of the selected region of interest were 161×250. For each EEM, it was processed by binning all the rows and columns prior to concatenating the vectors. Once all the EEMs had been treated in the same manner, PCA was performed with a minimum of 30 PCs. Additional pre-treatment techniques were also carried out such as range-scaling and mean-centring prior to modelling with the SIMPLS algorithm. Figure 5.11 displays the RMSEP plots attained.

	YV	NV	RMSEP
M-C		✔	186.78
	✔		178.39
R-S		✔	180.24
	✔		170.98
RAW		✔	188.04
	✔		187.72
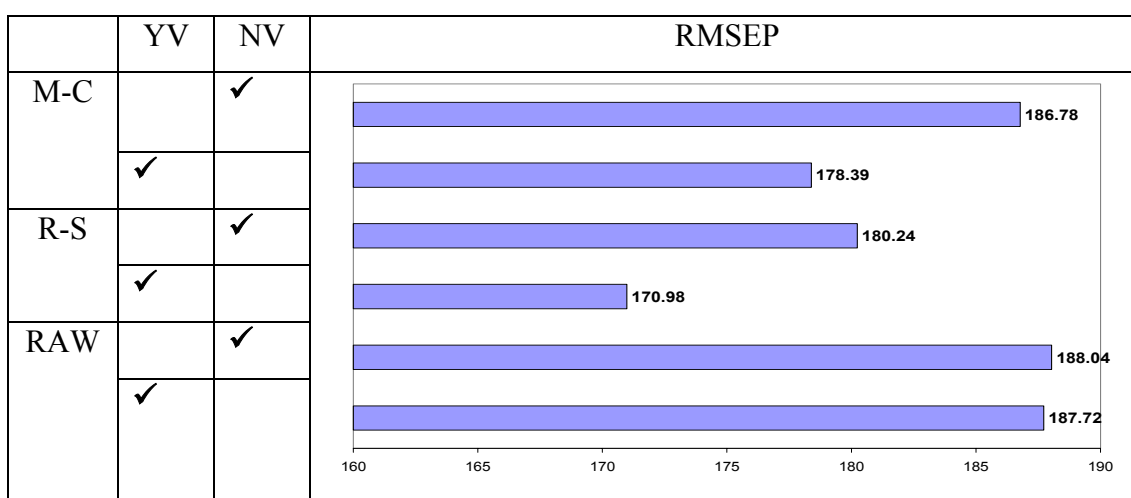
Figure 5.11: RMSEP plot for the prediction of the concentrations of anthracene, phenanthrene and naphthalene in a set of synthetic samples detailed in Table 5.1 from the processed datasets originating from the PCA scores of the binned rows and columns of the region of interest pertaining to the EEMs.
Key: YV: validation set included; NV: validation set not included;
M-C and R-S: mean-centred and range-scaled respectively with Savitzky-Golay smoothing
(polynomial: 3; window: 41)

The prediction of the PAHs following the range-scaling of the dataset is most accurate with a value of 170.98 ppb. Furthermore, the incorporation of the validation set into the model further improves the RMSEP values. The raw and mean-centred datasets have not performed as well. Overall, the respective models generated to predict the concentrations of all three PAH components have not been as effective as reported in Section 5.4.3.2.

5.4.3.4 Synchronous Fluorescence Spectra Extracted from the EEMs

Figure 5.12 displays the different spectra obtained at different synchronous wavelengths ($\Delta\lambda$ = 0, 20, 40, 60, 80 and 95 nm) as extracted from the EEMs (Meras *et al.,* 2005). Figure 5.13 displays the calculated RMSEP values obtained for the prediction of the three PAH analytes at varying synchronous wavelengths via latent variables suggested by the minimum RMSEP.
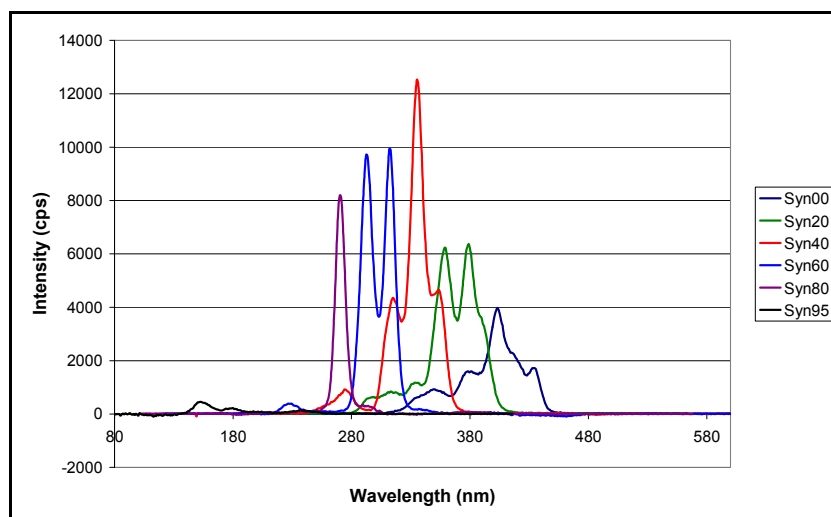
Figure 5.12: Synchronous spectra extracted from the EEM of a standard containing 500 ppb anthracene, phenanthrene and naphthalene in HPLC grade acetonitrile obtained at varying synchronous (Syn) wavelengths (Syn20 implies Δλ = 20nm)
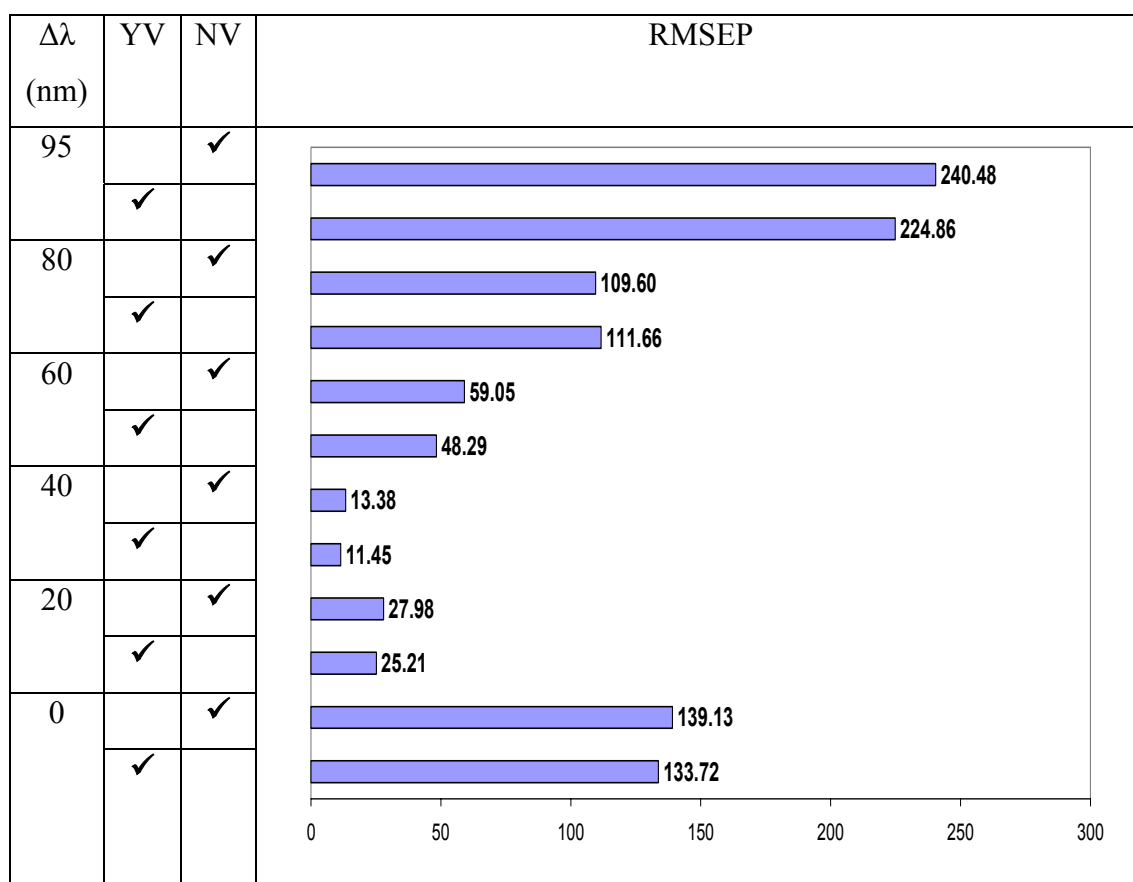
Figure 5.13: The calculated RMSEP values for the prediction of ANT, PHE and NAP at varying synchronous wavelengths (Δλ) via the minimum RMSEP.

Key: YV: validation set included; NV: validation set not included;

Less spectral information becomes available for $\Delta\lambda = 80$nm and 95nm, which implies that the respective models generated by the SIMPLS algorithm will be incapable of providing good predictions. A synchronous wavelength of 40nm has produced the datasets with the most spectral information and thus the better SIMPLS model with an RMSEP value of 11.45 ppb. This conforms to the findings reported in the literature in which three synchronous spectra of anthracene, phenanthrene, naphthalene and pyrene were recorded at $\Delta\lambda = 10$, 50 and 100 nm (Ferrer *et al.,* 1998). The authors rejected the spectra at $\Delta\lambda = 10$nm due to no distinction being attainable between the PAHs. Referring to Figure 5.13, in all instances except for $\Delta\lambda = 80$nm, the incorporation of the validation set into the respective models marginally improves the RMSEP values. This is reflected in the comparison of the LOO CV plots for $\Delta\lambda = 40$nm and 95nm (Figure 5.14).
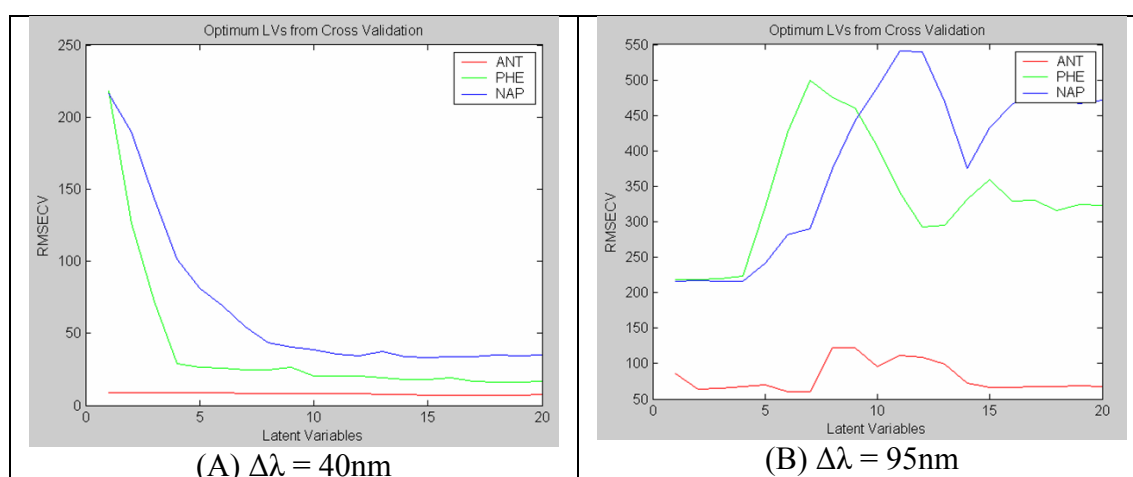
Figure 5.14: Comparison of the LOO-CV plots obtained for the SIMPLS models at synchronous wavelengths ($\Delta\lambda$) of 40nm (A) and 95nm (B)

The LOO-CV plot for $\Delta\lambda = 95$nm is thus very poor compared to that of $\Delta\lambda = 40$nm. The naphthalene curve is also very good at $\Delta\lambda = 40$nm. However, closer inspection of the scores plot (not shown) for the $\Delta\lambda = 95$nm indicated the presence of an outlier. Omission of this sample and thus re-modelling did not result in an improved LOO-CV plot. The overall RMSEP value of 175.13 ppb was calculated, which was only a slight improvement from 224.86 ppb.

Further pre-treatments involving range-scaling and mean-centring were performed on the dataset pertaining to $\Delta\lambda = 40$nm in addition to predicting via the minimum

RMSEP and the F-test (Section 1.5.5.2). However, no improvement was observed in the RMSEP values. The better RMSEP value was seen for the un-scaled synchronous fluorescence spectra at 11.45 ppb as determined by the minimum RMSEP. Except for the mean-centred dataset, there was a marginal improvement in the RMSEP values from 1.5 to 16 ppb when the validation set was incorporated in the model.

Focusing on the mean-centred data, there is some concern as to why the RMSEP value for incorporation of the validation set is so large compared to exclusion of the validation set. No outliers had been observed in the respective validation sets employed in both the 2D-F and SYN40 analyses. Figure 5.15 illustrates the possible cause of the high RMSEP values obtained for mean-centring of the binned emission spectra (Section 5.4.3.2).
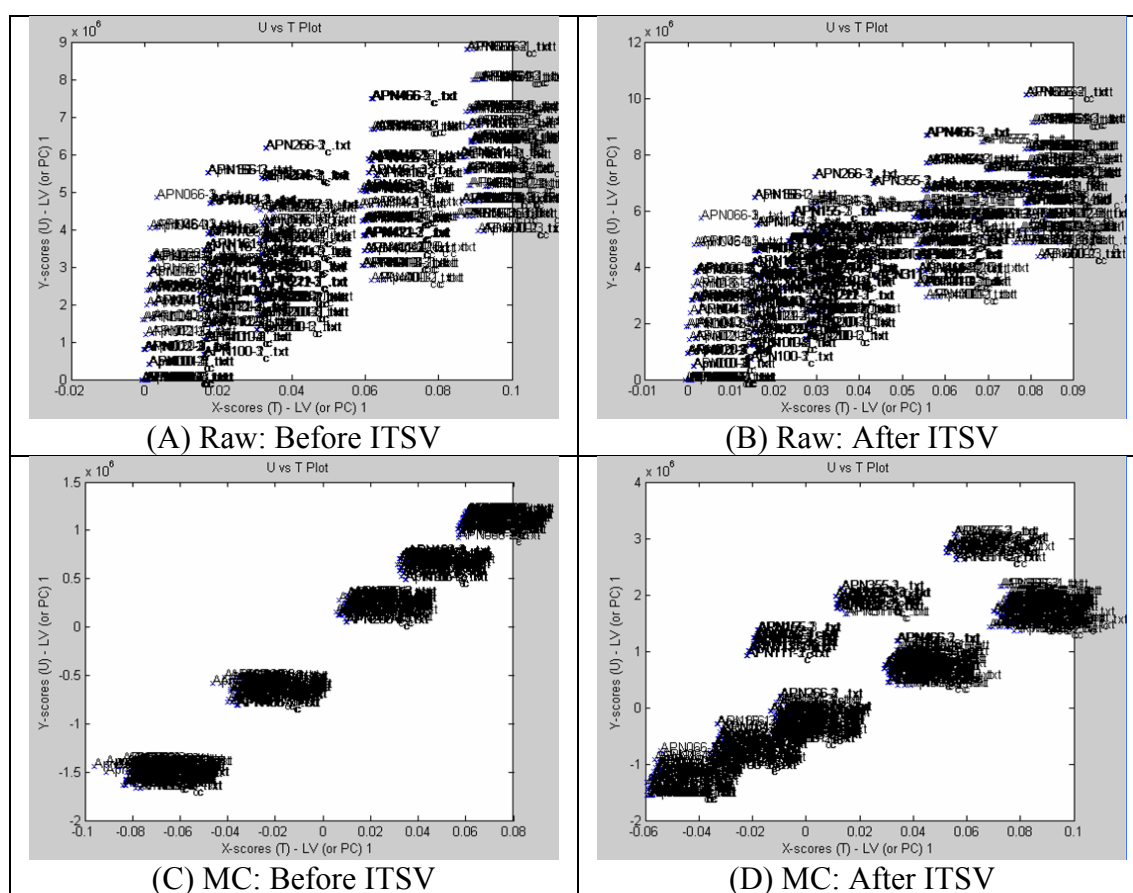
Figure 5.15: Comparison of U vs T plots before and after incorporation of the validation set

As can be seen in Figure 5.15C definite clusters have been formed. When the validation set is thus incorporated into the model, separate clusters result (Figure 5.15D). This most likely leads to significant underperformance of the model. This

further leads to the score of an unknown sample failing to be associated with any of the clusters, and thus results in a high RMSEP value. Further, the validation set is embedded within the training set for the raw data (Figure 5.15A and 5.15B).

Table 5.22 compares the calculated means for the prediction of the concentrations of the synthetic samples (Table 5.1) relating to the best RMSEP values calculated for the 2D fluorescence emission spectra and the synchronous spectra ($\Delta\lambda = 40$nm).

Table 5.22: Comparison of the calculated concentration means pertaining to the optimum calculated RMSEP values

Sample	2D Fluorescence (Emission)			ACTUAL			3D Fluorescence (SYN40)		
	ANT	PHE	NAP	ANT	PHE	NAP	ANT	PHE	NAP
UNK1	383.6	145.4	562.9	400	150	600	387.12	136.98	577.86
UNK2	493.2	308.7	395.7	500	300	400	505.84	303.56	399.26
UNK3	180.2	402.4	591.0	200	400	600	198.73	407.06	621.67
UNK4	126.7	492.0	577.2	150	500	600	150.44	506.44	598.35
UNK5	0.00	0.00	502.4	0	0	500	0.00	3.06	522.75

The calculated RMSEP values were 15.07 and 11.45 ppb for the 2D and 3D fluorescence spectra respectively. A very close agreement in the calculated concentrations is observed in the 3D compared to the 2D modes of acquisition. However, the concentration of naphthalene in the UNK5 sample is predicted much well in the 2D (502.4 ppb) than in the 3D (522.75 ppb). However, a very good agreement is seen in the UNK4 sample for the 3D mode of acquisition.

Overall, the synchronous wavelength of $\Delta\lambda = 40$nm has given the better RMSEP value for the prediction of the three target PAH components.

5.4.4 Application to Real Samples

Figure 5.16 displays the averaged fluorescence emission spectra of the B4 soil sample solution. For comparative purposes the emission spectra of the acetonitrile and a standard solution have been included.
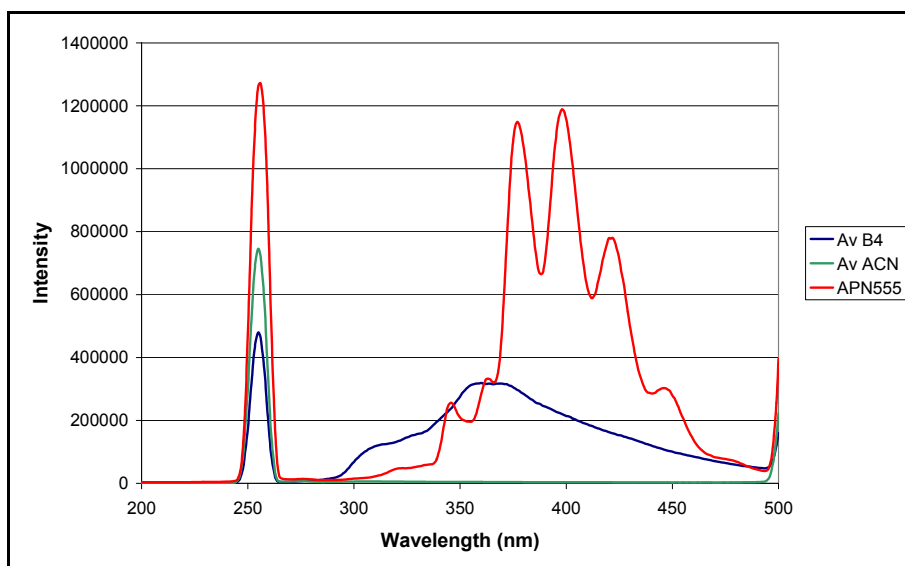
Figure 5.16: Fluorescence (2D) emission spectra of B4 soil sample after extraction with acetone (blue); blank consisting of acetonitrile included (green); standard solution: 500 ppb anthracene, phenanthrene and naphthalene in HPLC grade acetonitrile (red).
Instrumental conditions: Table 5.2

The anthracene peak gives a highly characteristic spectral response compared to the other components as can be seen in the spectrum for the standard solution. Looking at the spectrum for B4, one can infer that there is minimal presence of anthracene, and this complies with the true values listed in Table 5.5. The presence of the other PAHs can also be seen in the synchronous fluorescence spectra at $\Delta\lambda = 40$nm (Figure 5.17).
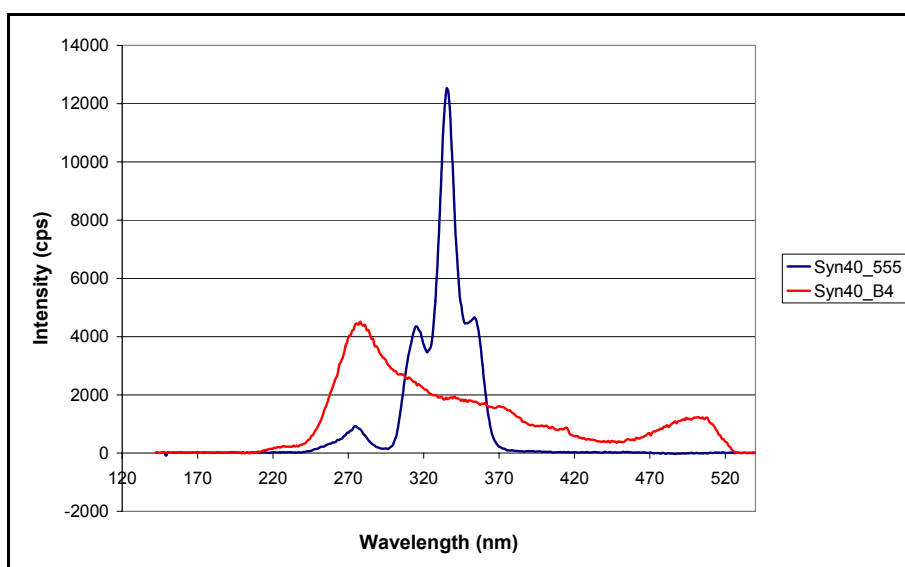
Figure 5.17: The synchronous fluorescence spectrum ($\Delta\lambda = 40$nm) for 500 ppb anthracene, phenanthrene and naphthalene in HPLC grade acetonitrile (Syn40_555), and soil sample B4 in acetonitrile (Syn40_B4)

Again one can infer the minimal presence of anthracene in the B4 sample. It is interesting to note that the maximum peak of the B4 sample (red) directly lies above the small peak at 280nm (blue). This could be attributed to naphthalene.

Based on the above observations, the SIMPLS models developed in the previous sections were deemed inadequate for the determination of the three main target PAHs in the B4 sample. This is mostly due to the different spectral profiles obtained. Furthermore, predictions had been performed and in many cases concentrations in excess of 1200 ppb were predicted for naphthalene and phenanthrene which were thus more than double the maximum concentration of 600 ppb in the training set. The test samples could have been diluted down with acetonitrile but at a risk of losing spectral information with regards to anthracene. It is possible that a reduction in quenching may have resulted in more characteristic spectra.

5.4.5 Prediction of Anthracene Only

Section 5.4.3.3 discussed the application of PCA to the binned emission and excitation spectra of the region of interest illustrated in the intensity chart in Figure 5.5 after concatenation of the vectors. The application of the respective SIMPLS models did not lead to good RMSEP values for the prediction of the three target PAH analytes. However, the literature has reported that the modelling of one target component leads to better RMSEP values than modelling all the components simultaneously; in terms of the NIPALS algorithm, PLS1 models are more robust than PLS2 models (Ferrer *et al.,* 1998). In this context, the effects of modelling only one component, in this case anthracene was investigated.

As anthracene produces the most intense spectral profile compared to the other two components, namely phenanthrene and naphthalene, the RMSEP values for the prediction of anthracene alone were calculated. Table 5.23 shows the calculated RMSEP values for the application of PCA to the binned emission and excitation spectra. Concurrently, Table 5.24 shows the calculated RMSEP values for the prediction of anthracene from synchronous spectra ($\Delta\lambda = 40$ nm) in conjunction with additional data pre-treatment.

Table 5.23: Calculated RMSEP values for the prediction of the concentrations of anthracene only in a set of "unknown" samples detailed in Table 5.1 from the processed datasets originating from the PC scores of the binned EEMs.

Scaling	Validation set included	Scaling	Validation set not included
None	13.30 ppb	None	12.96 ppb
Range-scaling	7.57 ppb	Range-scaling	8.16 ppb
Mean-centring	28.34 ppb	Mean-centring	12.96 ppb

In this instance, the range-scaled dataset has given the best prediction but the mean-centred data has performed worse. More importantly, the RMSEP values are much improved, for example the RMSEP value for the prediction of anthracene in the "unknown" samples for the range-scaled dataset with the validation dataset included in the model is 7.57 ppb compared to 170.98 ppb for all three components (Figure 5.11).

Table 5.24: Calculated RMSEP values (ppb) at varying pre-treatment techniques for $\Delta\lambda = 40nm$ as suggested by the optimum latent variables from the minimum RMSEP (M) and the F-test (F) but for anthracene (ANT) only. Key: YV: validation set included; NV: validation set not included

Pre-treatment	Min RMSEP	F-TEST
Raw (YV)	6.94	6.12
Raw (NV)	6.74	7.01
Range-scale (YV)	6.44	6.51
Range-scale (NV)	5.86	7.05
Mean-centre (YV)	134.97	6.55
Mean-centre (NV)	140.57	7.08

Finally, regarding the extraction of synchronous spectra from the respective EEMs (Section 5.4.3.4), the predictions via the optimum latent variables suggested by the F-test for the raw data (YV) is marginally better than those suggested by the minimum RMSEP (6.12 and 6.94ppb respectively). However, the opposite is true for range-scaling. However, range-scaling offers the best predictions in conjunction with the SIMPLS modelling algorithm (6.44 ppb and 5.86 ppb for YV and NV respectively). Overall, a viable method for the prediction of anthracene in samples containing a range of PAHs is possible.

5.5 Conclusions

Univariate analysis for the determination of the three PAHs employed in this study, namely anthracene, phenanthrene and naphthalene, cannot be performed successfully due to the overlapping spectra of the three respective molecules, under the experimental conditions applied. Not even the method of standard addition would be applicable, especially in real samples.

Multivariate calibration has considerably improved the ability of the unknown samples to be accurately predicted. In predicting the test set values, performing range-scaling and filtering with the Savitsky-Golay function prior to modelling leads to low RMSEP values. Application of weights in addition to the above-mentioned range-scaling and filtering can lead to reduction of the RMSEP values for specific components, for example anthracene but cannot currently be employed to deconvolute all three target PAHs due to the severely overlapping spectra. Comparisons of the SIMPLS models generated show that better RMSEP values are obtained when the validation sets are incorporated into the models.

Initial attempts at working with the data acquired on the 3D fluorimeter (SPEX 3D) have proved positive. The RMSEP values obtained were better than those from the 2D instrument (15.07 and 11.45 ppb for 2D and 3D respectively). Although it is usually not advisable to bin spectra due to the risk of losing important information, this seems to have not been the case. In fact, future work could involve varying the number of spectra binned.

A vast improvement has also been observed in the overall prediction of the target PAH analytes, especially with the application of synchronous fluorescence spectroscopy to the EEM with a $\Delta\lambda$ of 40nm. Naphthalene has been shown to be difficult to predict and this is reflected in the spectra obtained. Anthracene can still be successfully predicted in the presence of phenanthrene and naphthalene. This was observed for the datasets employed on two occasions, namely in Section 5.4.5 in which excellent RMSEP values were calculated (5.86 ppb).

Regarding the quantitative determination of the amounts of the three target PAHs in a real soil sample, initial attempts have proved unsuccessful. This is mostly due to the presence and hence interference of the other PAH molecules also present in the sample. Acetone was chosen mostly due to it being a well-utilised solvent for the extraction of PAH species from soil samples (Tomaniova *et al.,* 1998; Janska *et al.,* 2004). Furthermore, predicted concentrations exceeded the maximum concentration in the training set especially for phenanthrene and naphthalene indicating that dilutions should have been made.

Overall, for the simultaneous prediction of all three target PAHs, synchronous fluorescence spectra obtained from the EEM in conjunction with the SIMPLS algorithm gives the better RMSEP values.

CHAPTER 6: DISCUSSION AND CONCLUSIONS

6.1 Overview

A brief introduction will be given with a reminder of the aims and objectives. This is followed by a general discussion which will summarise the findings from Chapters 2 to 5. Further explanations of procedures undertaken will be given with discussions with reference to the literature. Finally, an overall conclusion and recommendations for future work is presented.

6.2 Introduction

The major aims and objectives of this thesis have been met with regards to the development of analytical techniques for water quality determination. In Chapter 2 quantitative determination of cadmium, lead and copper via multivariate calibration of DPASV voltammograms acquired on carbon-ink screen-printed electrodes was presented. The same was seen for arsenic, mercury and copper via multivariate calibration of DPASV voltammograms on gold-ink screen-printed electrodes (Chapter 3). However it must be stressed that the quantification in terms of the calculated root mean square error of prediction (RMSEP) was greater in magnitude than for the carbon-ink electrodes implying weaker predictive capability of the multivariate calibration modelling algorithm.

A key element of this thesis was the successful development and application of an in-house custom-built data analysis package constructed in both the Matlab and LabVIEW environments. It was initially designed to handle DPASV voltammograms (Section 2.3.3, 2.4.1 and 3.4.2) but later was adapted to handle fluorescence spectra (Section 5.4.2 and 5.4.3). Extra modules were also constructed to manipulate the excitation-emission matrix (EEM) spectra acquired on the 3D fluorimeter (SPEX 3D). Excellent quantification of three target PAHs was attained, with the most reliable data being obtained for anthracene.

The development of a personal digital assistant (PDA) application for the quantitative determination of cadmium, lead and copper in the field was reported (Chapter 4). This was in conjunction with the construction of a portable multi-channel potentiostat which was also controlled by the PDA. Screening and quantification of the target analytes was achieved on the PDA or on a desktop personal computer after transference of the data to it.

The next Section will discuss the findings from all four chapters.

6.3 General Discussion

6.3.1 Carbon-Ink Screen-Printed Electrodes

There have been many papers published over the last few years introducing new variants of PLS and/or other chemometric techniques and approaches to data pre-treatment. Chemometrics dates as far back as the early-to-mid nineteenth century (Brereton, 2003). Cauchy was one of the first to deal with matrices in 1829 and PCA was first used in 1879 for linear calibrations. The revolution came in the 1930s when PCA was applied to pyschometrics. It was not until the 1970s that PCA was applied to chemical data and the field of chemometrics was born.

The non-linear iterative partial least squares (NIPALS) algorithm has long been established, dating as far back as the 1970s. The statistically inspired modification of partial least squares (SIMPLS) algorithm introduced in 1993 superseded it producing better models at a faster rate leading to improved predictions, as has been reflected in the current research. However, coupled with the increase in technological advances, more complex and computationally time-consuming calculations are possible. This has led to the incorporation of large data sets such as those obtained from DNA microarrays (Datta, 2001; Datta, 2003b; Datta, 2003a).

The SIMPLS algorithm has been shown to be the better multivariate calibration regression tool for quantitatively predicting the concentration of cadmium, lead and copper in a given batch of samples. Although very good root mean square errors of

prediction (RMSEP) values were attained, there were still some issues leading to a cause for concern. These causes were not specifically related to the multivariate calibration regression model (SIMPLS) but also to the physical measurements themselves, namely the acquisition of the DPASV voltammograms on the carbon-ink screen-printed electrodes.

Two batches of carbon-ink screen-printed electrodes were employed in the study. The first (C-SPE-O) had been employed by Cooper who had optimised the experimental conditions for the DPASV acquisition of voltammograms with the electrodes in question (Cooper, 2004). The second batch (C-SPE-N), just like the first had been printed in-house (details given in Section 2.3.2). However it had been observed that the C-SPE-O batch of electrodes was more responsive than the C-SPE-N batch. At the same time, it was suggested that the reference electrodes of the C-SPE-O batch had suffered the effects of aging due to the shifted voltammograms observed.

Although the same insulator ink was employed for both batches, it was found that the ink employed for the C-SPE-N batch was more viscous than for the C-SPE-O batch. This implied that a thinning agent was required in the form of an organic solvent when applying to the screen. Due to propriety issues it was not possible to gain full details of the composition of the respective inks. However, the developed data analysis package was still capable of processing the acquired voltammograms on the C-SPE-O batch. The development of a custom-built program designed to re-align the voltammograms in order to increase the optimal performance of the model was described with better RMSEP values attained after SIMPLS modelling compared to no re-alignment. There are other parameters to consider with regards to the electrodes themselves. For example, the surface of the working electrode on each screen-printed electrode will be subtly different and thus possess slightly differing characteristics. This will in turn slightly alter the signatures of the voltammograms acquired.

The respective SIMPLS models were found to be more robust compared to the PCR, NIPALS and NNPLS (neural network) algorithms. To this effect, the focus of the study turned to the application of data pre-treatment prior to SIMPLS modelling. In particular, weighting of the individual variables in each voltammogram was attempted and led to a vast improvement in the RMSEP values, especially in conjunction with

outlier omission and range-scaling. Application of the weighting function involves a preliminary scan of the respective concentration matrix in order to "decide" whether the weights should be applied for a particular component. If so, then a weight is applied at the appropriate index range as defined by pre-set parameters. If a component is found to have 0 ppb concentration in the given sample then a low weight is applied.

In applying the weights to voltammograms that do not contain a corresponding concentration matrix (Y), the weighting function described in the previous paragraph is also able to account for this. Based on the pre-determined index ranges the function determines the standard deviation and if it exceeds a specified pre-defined limit, it applies the high weight, or else it applies the low weight. Overall the weighting function increases the signal-to-noise ratio substantially which has thus led to the improved RMSEP values attained. To the author's knowledge no similar function has been described in the literature.

Throughout the study, it was observed that the RMSEP values of lead were always much better (lower) than for both cadmium and copper. With reference to a DPASV voltammogram the lead peak was said to be on a horizontal baseline whereas the baselines for cadmium and copper were elevated, possibly due to hydrogen evolution and mercury stripping respectively at the extremes of the potential windows. However, another plausible cause for the superior performance of lead is possibly due to the inter-metallic bonding that can occur between cadmium and copper (Herrero & Ortiz, 1997), which contradicts the findings of Crosman and co-workers in 1975 who stated that cadmium did not form inter-metallic bonds with copper (Crosman *et al.,* 1975). The majority of inter-metallic studies have mostly been focused on copper and zinc. The effects of the inter-metallic bonding lead to a depression or shift of the voltammetric peak (Copeland *et al.,* 1974; Crosmun *et al.,* 1975). However, this is unlikely to be the cause of the shifted voltammograms encountered with regards to the C-SPE-O batch because all three peaks were shifted suggesting a possible reference electrode-related issue.

The quantitative determination of cadmium, lead and copper in a given certified reference material (CRM) soil sample by the SIMPLS models constructed with the

data analysis package yielded good agreement with the determinations performed by graphite furnace atomic absorption spectroscopy (GF-AAS). Better agreements were observed between the multivariate determination and GF-AAS than with standard addition. This can be attributed once again to cadmium and copper being respectively affected by the evolution of hydrogen and the stripping of the mercury. Overall, the % recovery was very relatively low (~15 – 30%).

However, it must be stressed that the aim of this exercise was not to develop a novel extraction process but merely to test the capabilities of the generated models and the data analysis package in processing voltammograms acquired from real samples. For percentage recoveries greater than 90%, an efficient but complex procedure is briefly described (Feeney & Kounaves, 2002). 5g of soil was placed in an extraction cell and placed in an oven at 200°C and pressurised to 20 – 30 bar. Acidified water was then pumped through the oven at 3ml/min. The extract would then be collected in a vial and cooled to 25°C. Cadmium, lead and copper were then determined by GF-AAS; the overall duration of the procedure lasting for 70 minutes.

Thus overall it can be seen that application of multivariate calibration regression modelling, in particular SIMPLS, combined with a number of pre-treatment techniques shows promise as a methodology for the determination of cadmium, lead and copper in environmental samples.

6.3.2 Gold-Ink Screen-Printed Electrodes

The majority of solid state working electrodes suffer from the cumbersome task of intense cleaning and polishing after a measurement has been performed, such as with the gold disk electrode. The need for disposable electrodes is paramount for analysis in the field hence allowing more rapid management decisions to be made as a result of the data generated. Preliminary work had been carried out in-house on gold-ink screen-printed electrodes for the determination of arsenic, mercury and copper by Cooper (Cooper, 2004). Optimal parameters obtained had been a deposition potential of 0.0V, a deposition time of 30 seconds and a 4M HCl electrolyte. However, copper peaks were found to overlap with mercury and suppress arsenic (Cooper, 2004). Multivariate Calibration techniques were thus applied to the voltammograms to

quantify the presence of all three metals, but were not as successful as reported in Chapter 2. The gold-ink screen-printed electrodes kindly supplied by Du Pont have been a contributory factor to this.

Section 3.4.1 presented an extensive analysis on a number of batches supplied by Du Pont; a company which has a long history of manufacturing electrode inks, such as the platinum-gold 7553 employed in the preparation of printed micro-circuits (Hoffman & Nakayama, 1968). It was found that the current ink formulations did not possess long shelf-lives. Furthermore the curing modes employed appeared to affect the overall analytical response of the working electrode. Table 6.1 displays the codes of the inks employed in addition to the optimum batches from the first and second batches.

Table 6.1: A comparison of the codes and methods of preparation of the batches of gold-ink screen-printed electrodes supplied by Du Pont

ID	Carbon	Ag/AgCl	Gold	Cure	Encapsulation	Quantity
D	E100735-155	5874	BQ331	Oven	E017257-1	15
2	BQ225	BQ164	BQ331	IR	BQ425	10
4	BQ225	BQ164	BQ331	Oven	BQ411	10
AO	BQ225	BQ164	BQ331	Oven	BQ411	30
AI	BQ225	BQ164	BQ331	IR	BQ411	75
BO	BQ242	BQ164	BQ331	Oven	BQ411	20
BI	BQ242	BQ164	BQ331	IR	BQ411	30

The gold ink is the same throughout (BQ331). The exact formulation of these inks is not known due to propriety issues. This in turn makes it difficult to deduce or propose the cause of the problems incurred. In Section 3.4.1, it was shown that the BI batch of electrodes had better repeatability and reproducibility. However, as only a limited number of electrodes were supplied the AI batch was employed.

Although the AI batch was highly repeatable for arsenic (III), the same cannot be said for mercury (II) and copper (II). It had been observed that the response of the 600 ppb mercury standard was very much higher than at 800 ppb. One postulation is that an inter-metallic formation occurs between the copper and mercury at that specific concentration. Although studies have been carried out on copper-mercury inter-metallic bonding (Ben-Bassat & Azrad, 1978) and dating far back as the 1930s, none

has so far at the time of writing, have been reported in the literature specifically on gold (-ink) working electrodes. This warrants further investigations.

Researchers are continuously looking into producing their own gold-ink screen-printed electrodes. For example it was reported in 2003 that by optimising the length of the alkanethiol encapsulant and the size of the gold nano-particles the gold-plated electrodes on a plastic substrate can be cured at a temperature of 150°C instead of the normal which is in excess of 200°C (Huang *et al.*, 2003). The fabrication of the gold film involved transferring a gold (III) solution ($AuCl_4^-$) into toluene. The aqueous phase was removed. An alkane thiol ranging in carbon length from 4 to 12 was added; the length dependent on the required size of the gold nano crystals. Table 6.2 displays the mole ratios of the thiol:gold.

Table 6.2: Mole ratios of the thiol:gold depending on the size of the required gold nano-crystals

Gold nano-crystal diameter (nm)	Mole ratio (Thiol:Gold)
1.5	4:1
5.0	1/12:1

Sodium borohydride dissolved in water (amount not given) was added to the organic phase. A reaction thus ensued for 3.5 hours at room temperature. A rotary evaporator was then employed to remove the toluene. The remaining black residue was immersed in ethanol and briefly sonicated. After washing with ethanol followed by acetone and left to dry, the residue was re-dissolved in toluene ensuring that it was saturated. A micro-pipette was employed to plate the tracts on to a given substrate (Du Pont Melinex sheets (polyethylene terephthalate)). After drying, the black film was heated in two stages; the first was to sublimate the alkanethiol; completion was achieved when the black residue turned to gold. The second heating stage involved melting the film, coagulating it and then solidification in order that a conducting gold film resulted.

The authors found that better conductivity was attained when the number of carbons in the alkanethiol ranged from 4 to 6. They had also performed stability studies. They had observed that due to a continuous evaporation of the alkanethiol, the shelf-life was reduced unless they were refrigerated. Further, they also observed that the larger

diameter gold nano-crystals were also prone to degradation due to the alkanethiol encapsulant evaporating as a result of the larger fraction of the gold. Overall, the authors concluded that hexanethiol was the better encapsulant coupled with gold nano-particle diameters of 1.5nm and an "annealing temperature" of 150°C.

Arsenic and copper are also reported to form inter-metallic complexes (Feeney & Kounaves, 2002). This may be the reason why no arsenic peak was observed on the gold-ink screen-printed electrodes in this study after extraction from the certified reference material (CRM) soil sample. Hydride generation atomic absorption spectroscopy (HG-AAS) showed that there was indeed arsenic in the extract with a % recovery of 11.76%. However it has been reported that copper can interfere with arsenic determination on HG-AAS if its concentration exceeds 4 ppm (Amankwah & Fasching, 1985). An alternative explanation is that the arsenic species was in the arsenate (As (V)) form. This seemed likely by the acquisition of voltammograms after reduction of arsenate to arsenite with L-cysteine. However, the presence of copper, and other metals such as mercury (2.42mg/kg), chromium (27.2mg/kg), zinc (140mg/kg) and antimony[4] (3.2mg/kg) may complicate the quantitative determination of the arsenate and was thus not attempted. Antimony (Sb) is also known to interfere with arsenic determination, especially in HG-AAS (de Moraes Flores *et al.,* 2002). However this is normally when antimony concentrations are in excess of arsenic. On examination of the CRM, the certified amount of arsenic (5.64mg/kg) is greater than that of antimony.

Overall, the major issue that has affected the outcome of this study is the quality of the gold-ink screen-printed electrodes which has made it difficult for the SIMPLS algorithm employed to construct models. Artificial neural networks (ANNs) could be employed but as had been observed with the carbon-ink screen-printed electrodes, the SIMPLS algorithm was more efficient. A future work could thus be to re-apply ANNs to the current data set and vary the number of epochs in addition to the input, hidden and output nodes; especially as the datasets are larger than when ANNs was previously employed.

---

[4] Not certified

6.3.3 Portable Field Devices and Automation

The construction of the personal digital assistant (PDA) application to predict the amounts of target analytes in given samples via pre-loaded multivariate calibration regression models has been successful (Section 4.4). As had been stated the application in its current state is restricted in functionality. For example, the application is able to perform range-scaling and mean-centring but not Savitzky-Golay smoothing or variable weighting. The latter cannot be performed because it too employs the Savitzky-Golay smoothing algorithm.

It was not possible to simply convert the LabVIEW functions that are employed in the desktop version of LabVIEW into the PDA version. Although there were some specialised PDA functions they were rather limited in use. Attempts had been made to build "PDA equivalents" but they proved unsuccessful. Alternative methods were thus sought. One employed the construction of a desktop module to convert the normal text files containing voltammetric data into specialised binary files which could be imported into the PDA application.

A current problem affecting the performance of the PDA application is the fact that when the model is imported, the PDA application is unable to determine whether instrument standardisation was performed. If a model is imported which was constructed in the absence of instrument standardisation, the "Properties" screen still indicates that it was performed. It is unclear as to why this phenomenon occurs. The predicted concentration values therefore disagree with the desktop version. This is mostly due to the input voltammograms being standardised by an empty piecewise direct standardisation (PDS) transformation matrix. When both a multivariate calibration model and a PDS model are imported on to the PDA, the predicted concentrations still differ. This is due to the standardisation PDA program attempting to perform Savitzky-Golay smoothing on the voltammograms pertaining to the target analytes.

Parallel to the work performed in this thesis, a miniaturised multi-channel potentiostat capable of acquiring signals from four channels simultaneously has been developed by co-workers. The PDA software application constructed to acquire the signals and

process the data was written in the C# language. It would be better if the acquired voltammograms could be processed immediately instead of the need to employ another application, namely the LabVIEW PDA application. It may therefore be advisable to "re-write" the PDA application in C# seeing as there is more functionality available in that language.

The LabVIEW version should not be dismissed. It has become employed in a wide variety of fields over the last decade. For example, dissolution testing of pharmaceutical tablets was performed in which the data was acquired, processed and stored by *one* program which had been built and validated in-house with LabVIEW (Johansson *et al.,* 2002). It was also able to perform NIPALS (PLS1 and PLS2) and allowed concentrations to be calculated in real time instead of the arduous task of sample extraction every twenty minutes, filtration and then measurements performed on HPLC and UV-VIS spectrophotometry.

This study has also reported the successful application of instrument standardisation in the form of piecewise direct standardisation (PDS) with additive background correction (ABC). However, the selection of the subset was performed manually which is likely to have affected the overall RMSEP values attained. Further improvements to these values could be achieved by increasing the number of samples in the subset and ensuring that the entire concentration range is covered. However, application of an algorithm is more desirable than manual selection. The Kennard-Stone algorithm has already been described in Section 4.2.5 (Kennard & Stone, 1969). Another algorithm involves the use of the "highest leverage" (Wang *et al.,* 1991). This ensures that there is sufficient information in the subset to "describe the difference between the two instruments" which in this study would be the Autolab PSTAT10 laboratory instrument (primary) and the PDA field instrument (secondary).

The concept of *piecewise direct standardisation* (PDS) is to permit the model built with calibration data on one day to be able to predict data measured on a different day, and even by a different operator (Herrero & Ortiz, 1998). Another procedure similar to PDS is *global calibration transfer* method (GCT) (Herrero & Cruz Ortiz, 1999). PDS and GCT standardisation methods can also be employed to model the matrix interferences of iron on copper, when measured with DPASV (Herrero & Cruz

Ortiz, 1999). PLS regression to determine the concentrations of both iron and copper in the same solution was also employed. It had been reported that utilisation of the standardisation methods could lead to a 75% reduction in the use of calibration samples.

It has already been stated that the weighting functions developed in this study require improvements. It was also suggested that the PDA datasets should also be weighted in accordance with the laboratory-acquired datasets. However, in doing so one must be certain to ensure that the weighting occurs at the appropriate indices since the PDA datasets contain fewer variables than the laboratory datasets. This would also lead to the reduction in amplitude of the repetitive wave cycles observed in the PDA datasets for low and zero concentration. It is unclear what the cause of this phenomenon is but it can severely affect the determination of low concentration analytes.

The last several years have seen a number of portable devices developed involving the miniaturisation of key components with the aim of performing field-measurements. Examples have included the speciation of arsenic in drinking water by acquisition of voltammograms on solid-state gold working electrodes (Huang & Dasgupta, 1999); the detection of uranium via capillary electrophoresis (Collins, 2002); a multi-enzymatic electrochemical sensor for the detection of heavy metals and pesticides in potato and cabbage saps, which involved cholinesterase, urease, and glucose oxidase coupled "with silicon-nitride ion-sensitive layers" (Starodub *et al.,* 1999); the development of a planar poly(methyl methacrylate) chip for the speciation of arsenic by conductivity measurements (Prest *et al.,* 2003).

6.3.4 Fluorescence Spectroscopy

Fluorescence spectroscopy has become an important and powerful analytical technique especially when coupled to a host of chemometric tools. The application of weights had not led to improved root mean square error of prediction (RMSEP) values as had been observed for the heavy metals. This implies that a number of improvements to the weighting function are required, for example, in enabling the weighting of an individual compound, as opposed to simultaneous weighting of all three compounds. This is adequate for the heavy metals such as cadmium, lead and

copper which do not overlap (on a carbon-ink screen-printed electrode). However the circumstances with regards to anthracene, phenanthrene and naphthalene are different mostly due to the severe overlap observed. The same was observed for mercury and copper on the gold-ink screen-printed electrodes in Section 3.4.2.

The same improvements would also be required on the "intelligent" weighting function with regards to the weighting of spectra and voltammograms of samples with unknown concentrations. Cluster analysis should also be employed to assist this function. As already stated, the weighting function scans the relevant index ranges where target peaks are likely to appear. This is sufficient for the voltammograms containing peaks pertaining to cadmium, lead and copper. An initial analysis of the sample with cluster analysis would indicate the likelihood of the presence of a target component; if the likelihood was above a pre-defined limit, the high weight would thus be applied. The cluster analysis would therefore be applicable to all data types.

It has been shown in this study that better RMSEP values were attained from the 3D fluorescence spectra than the 2D emission spectra. This was due to being able to manually select the most characteristic portion of the EEM. The software binning of the samples led to much improved RMSEP values since prominent regions pertaining to the presence of the given target analyte(s) would be amplified whilst regions of noise would be minimised resulting in an overall improvement of the signal-to-noise ratio. However, the number of spectra that are to be binned still requires optimisation. An automated software procedure would thus be beneficial. The synchronous spectra extracted from the EEMs yielded superior RMSEP values. However, further improvements could be obtained by selective binning of the synchronous spectra, for example, with a $\Delta\lambda$ value ranging from 35 to 45nm.

The concatenation of the binned excitation and emission spectra from the respective EEMs prior to PCA in which the first 30 components were retained did not yield satisfactory RMSEP values compared to binning of emission spectra. This was partly attributed to the loss of spectral information due to employing a smaller region of the EEM. A feasible suggestion could be to also combine the synchronous spectra with the emission and excitation spectra in order to produce a better "fingerprint". This would be followed by PCA on the resulting concatenated vectors.

The EEMs acquired in this study were not subjected to multi-way chemometric analysis such as parallel factor analysis (PARAFAC), self-weighted alternating trilinear decomposition (SWATLD) or *N-way* PLS ((Bro, 1997; Chen *et al.,* 2000; Espinosa-Mansilla *et al.,* 2005)) because of the relative small area of interest as indicated by the region of interest. Furthermore, it had been reported that tools such as PARAFAC required further refining due to the disadvantage of the exact number of components in the mixture being known (Espinosa-Mansilla *et al.,* 2005). This would be most evident if a number of sample EEMs were acquired containing an unknown number of PAH components.

In addition to the application of fluorescence spectroscopy for the detection and quantitative determination of PAHs and other fluorescent analytes, the technique can be applied to the analysis of metal ions (Skoog *et al.,* 1996; Rouessac & Rouessac, 2000). This results from the complexation of a target metal ion with a fluorescing organic compound. It was reported that uranium could be detected in the presence of other toxic heavy metals such as mercury, lead, cadmium and copper with a portable capillary zone electrophoresis "lab-on-chip" device using complexing ligands attached to fluorescing dyes such as rhodamine B and calix-6-arene (Collins, 2002).

The application of disposable carbon-ink screen-printed electrodes for the electrochemical detection and determination of target metal analytes has been discussed in Section 6.3.2. However, it was recently reported that the same technology can be applied to PAHs, such as phenanthrene (Fahnrich *et al.,* 2003). The literature has reported devices containing multiple working electrodes, each coated with a different component that is specific for a given target analyte (Arkhypova *et al.,* 2001; Bachmann & Schmid, 1999; Bachmann *et al.,* 2000). This thus leads to the application of chemometrics and good experimental design for the quantitative determination of these target analytes.

6.4 Overall Conclusions

This study has highlighted the importance and difficulty in developing the correct data analysis tools for the successful monitoring for the production of safe drinking water. The quantitative determinations of a select number of heavy metal ions and polynuclear aromatic hydrocarbons via specialised data pre-treatment techniques followed by multivariate calibration regression models have been reported. As one of the ARTDEMO project goals was to develop on-line and at-line monitoring tools, this study has seen the preliminary attempts at achieving the latter.

A custom-built data analysis package has been developed in both the Matlab and LabVIEW environments. It is able to import the raw DPASV voltammograms and fluorescence spectra and construct the appropriate datasets. A range of data pre-treatment options are available such as the omission of outliers via PCA, the omission of rogue variables, and the application of specific weights. Instrument standardisation in the form of piecewise direct standardisation is available. Other techniques include range-scaling and mean-centring coupled with Savitzky-Golay smoothing, moving average smoothing and fast Fourier transform smoothing. A final pre-treatment option is orthogonal signal correction (OSC). Multivariate calibration models are constructed via either the PCR, NIPALS, SIMPLS or ANNs algorithm followed by leave-one-out cross-validation. All parameters can be saved in appropriate files for retrieval at a later time. Predictions can thus be performed and exported into text files. The package can also import voltammograms acquired on the accompanying PDA application, standardise the datasets and predict the respective concentrations of the target analytes.

The data analysis package has been successfully employed for the quantitative and simultaneous determination of cadmium, lead and copper on carbon-ink screen-printed electrodes; arsenic, mercury and copper on gold-ink screen-printed electrodes; anthracene, phenanthrene and naphthalene via fluorescence emission spectra and synchronous spectra extracted from excitation-emission matrix (EEM) spectra. These were all modelled with the SIMPLS algorithm. The overall better root mean square error of prediction (RMSEP) values was seen with the latter; the worse values were

seen with the gold-ink screen-printed electrodes. This was mostly due to the electrode difficulties as has been explained previously.

The majority of the predictions for each of the target analytes were performed at high latent variables ranging from 14 to 20. This would cause one to infer that the noise was also being modelled. However, given the nature of the voltammograms and the fluorescence spectra in addition to the number of variables which range from 180 to 300, the number of latent variables stated above is sufficient. This implies that any non-linear interactions are taken into account. For example, with regards to the DPASV voltammograms, any inter-metallic complexes can be modelled such as Cu-Hg and Pb-Cu (Zen *et al.*, 2000). Furthermore, the number of latent variables will describe the variability observed in such "complex models" (Thennadil & Martin, 2005).

A PDA application has been successfully developed and employed which leads to the ability to perform "chemometrics on a PDA". Although a number of issues were raised with regards to incompatible results attained between the desktop and PDA versions, this is mostly attributable to the PDA Module Add-On supplied by National Instruments for LabVIEW 7.1. It is evident that the actual programs do work seeing as the same code is employed in both versions and that the desktop version agrees with the Matlab version. This issue can thus be overcome by further investigations of the current code, upgrading to the next available version of the Add-On or re-writing the entire application in a different language such as Visual C#.

Overall, the aims and objectives of this study have been met. A custom-built data analysis package has been developed leading to the successful qualitative and quantitative determination of target analytes via the acquisition of DPASV voltammograms and fluorescence spectra via multivariate calibration. A PDA application has been developed which can acquire and process DPASV voltammograms and thus communicate with the main data analysis package. The potential for the application of automation via the development of a program to simultaneously control a sample preparation unit and screen-printed electrode dispenser has been demonstrated. The specific aims as set out by the ARTDEMO project has also been met.

6.5 Recommendations for Future Work

A number of issues were raised during the course of this project and the bullet points below suggest a number of recommendations for future work.

- The development of a database for storage of voltammograms (or spectra), regression models, RMSEP values, etc in place of the vast amounts of text files that are currently generated
- The further study of gold-ink screen-printed electrodes including:
    - The effects of inter-metallic bonding between mercury and copper
    - The quantitative determination of cadmium and lead
    - The application of PCA and ANNs over a wider range of concentrations
- The quantitative determination of chromium and zinc on carbon-ink screen-printed electrodes
- Incorporation of hierarchal cluster analysis to the main data analysis package to determine whether weights are applied to samples with unknown concentrations of target analytes
- Merging the multi-channel potentiostat data acquisition program on the PDA with the LabVIEW PDA application for instant processing and prediction via pre-loaded models
- The application of multi-way techniques such as PARAFAC to EEM spectra
- The development of an overall data analysis package able to prepare samples, acquire datasets, process, model and predict the amounts of the target analytes
- The application of different experimental design techniques so that larger concentration ranges can be employed but without the need to perform numerous measurements
- Applications and adaptation of the main data analysis package for the analysis of pharmaceutical data

The above tasks should be applicable for not only the purposes of the ARTDEMO project, but for any field whether environmental, industrial or pharmaceutical.

REFERENCES

1.   Agüí, L., Peńa, L., Pedrero, M., Yáńez-Sedeńo, P. and Pingarrón, J.M. (2002) Determination of Disulfiram by Adsorptive Stripping Voltammetry at Gold Disk Microelectrodes. *Electroanalysis*, 14 (7-8), 486-492.

2.   Alloway, B.J. (1990) *Heavy Metals in Soils*. Glasgow, Scotland, UK: Blackie & Son Ltd.

3.   Alpizar, J., Cladera, A., Cerda, V., Lastres, E., Garcia, L. and Catasus, M. (1997) Simultaneous flow injection analysis of cadmium and lead with differential pulse voltammetric detection. *Analytica Chimica Acta*, 340, 149-158.

4.   Amankwah, S.A. and Fasching, J.L. (1985) Separation and Determination of Arsenic (V) and Arsenic (III) in Sea-Water by Solvent Extraction and Atomic Absorption Spectrophotometry by the Hydride Generation Technique. *Talanta*, 32 (2), 111-114.

5.   Antunes, M.C.G. and Esteves da Silva, C.G. (2005) Multivariate curve resolution analysis excitation-emission matrices of fluorescence of humic substances. *Analytica Chimica Acta*, 546, 52-59.

6.   Arkhypova, V.N., Dzyadevych, S.V., Soldatkin, A.P., El'skaya, A.V., Jaffrezic-Renault, N., Jaffrezic, H. and Martelet, C. (2001) Multibiosensor based on enzyme inhibition analysis for determination of different toxic substances. *Talanta*, 55, 919-927.

7.   Atkins, P.W. (1992) *Physical Chemistry*. England: Oxford Press.

8.   ATSDR (Agency for Toxic Substances and Disease Registry) (1989) Toxicological Profile for Cadmium.

9.   ATSDR (Agency for Toxic Substances and Disease Registry) (1990) Toxicological Profile for Polycyclic Aromatic Hydrocarbons.

10.   ATSDR (Agency for Toxic Substances and Disease Registry) (1993) Toxicological Profile for Lead.

11.   Avdikos, E.M., Prodromidis, M.I. and Efstathiou, C.E. (2005) Construction and

analytical applications of a palm-sized microcontroller-based amperometric analyser. *Sensors and Actuators B*, 107 (372-378),

12. Bachmann, T.T., Leca, B., Vilatte, F., Marty, J.-L., Fournier, D. and Schmid, R.D. (2000) Improved multianalyte detection of organophosphates and carbamates with disposable multielectrode biosensors using recombinant mutants of Drosophila acetylcholinesterase and artificial neural networks. *Biosensors and Bioelectronics*, 15 (3-4), 193-201.

13. Bachmann, T.T. and Schmid, R.D. (1999) A disposable multielectrode biosensor for rapid simultaneous detection of the insecticides paraoxon and carbofuran at high resolution. *Analytica Chimica Acta*, 401 (1-2), 95-103.

14. Ben-Bassat, A.H.I. and Azrad, A. (1978) Intermetallic Compounds Formed in Mixed (Complex) Amalgams - I. The Systems: Copper-Mercury, Zinc-Mercury and Copper-Zinc-Mercury. *Electrochimica Acta*, 23, 63-69.

15. Bessant, C. and Saini, S. (1999) Simultaneous Determination of Ethanol, Fructose, and Glucose at an Unmodified Platinum Electrode Using Artificial Neural Networks. *Analytical Chemistry*, 71, 2806-2813.

16. Bessant, C. and Saini, S. (2000) A chemometric analysis of dual pulse staircase voltammograms obtained in mixtures of ethanol, fructose and glucose. *Journal of Electroanalytical Chemistry*, 489 (1-2), 76-83.

17. Bhattacharya, S. (2003) Homeopathy Reduces Arsenic Poisoning in Mice. *New Scientist.com*,

18. Billa, E., Pastou, A., Monties, B., Romero, J. and Koukios, E.G. (2000) Multivariate chemometric analysis of the fluorescence spectra of eucalyptus wood. *Industrial Crops and Products*, 11, 187-196.

19. Borba da Cunha, A.C., Lopez de Alda, M.J., Barcelo, D., Pizzolato, T.M. and dos Santos, J.H.Z. (2004) Multianalyte determination of different classes of pesticides (acidic, triazines, phenyl ureas, anilines, organophosphates, molinate and propanil) by liquid chromatography-electrospray-tandem mass spectrometry. *Analytical and Bioanalytical Chemistry*, 378, 940-954.

20. Bott, A.W. (1995) Voltammetric Determination of Trace Concentrations of Metals in the Environment. *Current Separations*, 14 (1), 24-30.

21. Bouveresse, E., Hartmann, C., Massart, D.L., Last, I.R. and Prebble, K.A.
         (1996) Standardisation of Near-Infrared Spectrometric Instruments.
         Analytical Chemistry , 68, 982-990.

22. Bouveresse, E. and Massart, D.L. (1996) Improvement of the piecewise direct
         standardisation procedure for the transfer of NIR spectra for multivariate
         calibration. *Chemometrics and Intelligent Laboratory Systems*, 32, 201-
         213.

23. Bowen, H.G.M. (1979) *Environmental Chemistry of the Elements*. USA:
         Academic Press Inc.

24. Brainina, Kh.Z. (1995) Sensors and sample preparation in stripping
         voltammetry. *Analytica Chimica Acta*, 305 (1-3), 146-153.

25. Brereton, R.G. (2003) *Chemometrics: Data Analysis for the Laboratory and
         Chemical Plant*. Chichester, UK: John Wiley & Sons.

26. Bro, R. (1997) PARAFAC. Tutorial and applications. *Chemometrics and
         Intelligent Laboratory Systems*, 38, 149-171.

27. Buchberger, W., Niessner, G. and Bakry, R. (1998) Determination of
         nifuroxazide with polarography and adsorptive stripping voltammetry at
         mercury and carbon paste electrodes. *Fresenius Journal of Analytical
         Chemistry*, 362, 205-208.

28. Budavari, S., O'Neill, M.J., Smith, A. & Heckelman, P.E. (1989) *The Merck
         Index*. New Jersey: Merck & Co Inc.

29. Cabanillas, A.G., Díaz, T.G., Díez, N.M.M., Salinas, F., Burguillos, J.M.O. and
         Viré , J.-C. (2000a) Resolution by polarographic techniques of atrazine-
         simazine and terbutryn-prometryn binary mixtures by using PLS
         calibration and artificial neural networks . *The Analyst*, 125 (5), 909-914.

30. Cabanillas, A.G., Díaz, T.G., Díez, N.M.M., Salinas, F., Burguillos, J.M.O. and
         Viré , J.-C. (2000b) Resolution by polarographic techniques of atrazine-
         simazine and terbutryn-prometryn binary mixtures by using PLS
         calibration and artificial neural networks . *The Analyst*, 125 (5), 909-914.

31. Capitan-Vallvey, L.F., Navas, N., del Olmo, M., Consonni, V. and Todeschini,

R. (2000) Resolution of mixtures of three nonsteroidal anti-inflammatory drugs by fluorescence using partial least squares multivariate calibration with previous wavelength selection by Kohonen artificial neural networks. *Talanta*, 52, 1069-1079.

32.   Chen, Z.-P., Wu, H.-L., Jiang, J.-H., Li, Y. and Yu, R.-Q. (2000) A novel trilinear decomposition algorithm for second-order linear calibration. *Chemometrics and Intelligent Laboratory Systems*, 52, 75-86.

33.   Christensen, J.H. , Hansen, A.B., Karlson, U., Mortensen, J. and Andersen, O. (2005) Multivariate statistical methods for evaluating biodegradation of mineral oil. *Journal of Chromatography A*, 1090, 133-145.

34.   Colin, R., Le Fur, E., Charreteur, C., Dufau, C. and Peron, J.-J. (2000) Determination of glyphosate herbicide and (aminomethyl)phosphonic acid (AMPA) in water by liquid chromatography and fluorescence detection. Part II: Direct determination using pre-column derivatisation with NBD-Cl. *Analusis*, 28, 819-824.

35.   Collins, G.E. Lu, Q., (Ed.) (2002) Metal Ion Analysis using Near-Infrared Dyes and the Laboratory-on-a-Chip. USA: US Department of Energy (Project Number: 64982).

36.   Cooper, J. (2004) Electrochemical Method for the Rapid In-Situ Screening of Heavy Metals in Soil and Water Samples.   Cranfield University at Silsoe.

37.   Copeland, T.R., Osteryoung, R.A. and Skogerboe, R.K. (1974) Elimination of Copper-Zinc Intermetallic Interferences in Anodic Stripping Voltammetry. *Analytical Chemistry*, 46 (14), 2093-2097.

38.   Crosmun, S.T., Dean, J.A. and Stokely, J.R. (1975) Pulsed Anodic Stripping Voltammetry of Zinc, Cadmium and Lead with a Mercury-Coated Wax-Impregnated Graphite Electrode. *Analytica Chimica Acta*, 75, 421-430.

39.   Datta, S. (2001) Exploring Relationships in Gene Expressions: A Partial Least Squares Approach. *Gene Expression*, 9 (6), 249-255.

40.   Datta, S. (2003a) Statistical Techniques for Microarray Data: A Partial Overview. *Communications in Statistics*, 32 (1), 263-280.

41.   Datta, S. (2003b) Statistical Techniques for Microarray Data: A Partial
         Overview. *Communications in Statistics*, 32 (1), 263-280.

42.   de Campos, R.C., Grinberg, P., Takase, I. and Luna, A.S. (2002) Minimisation
         of Cu and Ni interferences in the determination of Sb by hydride
         generation atomic absorption spectrometry: the use of picolinic acid as
         masking agent and the influence of L-cysteine. *Spectrochimica Acta Part
         B*, 57, 463-472.

43.   de Carvalho, R.M., Mello, C. and Kubota, L.T. (2000) Simultaneous
         determination of phenol isomers in binary mixtures by differential pulse
         voltammetry using carbon fibre electrode and neural network with
         pruning as a multivariate calibration tool. *Analytica Chimica Acta*, 420
         (1), 109-121.

44.   de Jong, S. (1993) SIMPLS: An alternative approach to partial least squares
         regression. *Chemometrics and Intelligent Laboratory Systems*, 18 (3),
         251-263.

45.   de Moraes Flores, E.M., da Silva, F.E.B., dos Santos, E.P., Paula, F.R., Barin,
         J.S., Zanella, R., Dressler, V.L. and Bittencourt, C.F. (2002)
         Determination of total arsenic by batch hydride generation atomic
         absorption spectrometry in injectable drugs containing high levels of
         Sb(V) as N-methylglucamine antimonate. *Spectrochimica Acta Part B*,
         57, 2095-2102.

46.   del Olmo, M., Diez, C., Molina, A., de Orbe, I. and Vilchez, J.L. (1996)
         Resolution of phenol, o-cresol, m-cresol and p-cresol mixtures by
         excitation fluorescence using partial least-squares (PLS) multivariate
         calibration. *Analytica Chimica Acta*, 335, 23-33.

47.   Desmond, D., Lane, B., Alderman, J., Glennon, J.D., Dermot Diamond and
         Arrigan, D.W.M. (1997) Evaluation of miniaturised solid state reference
         electrodes on a silicon based component. *Sensors and Actuators B:
         Chemical*, 44 (1-3), 389-396.

48.   Desmond, D., Lane, B., Alderman, J., Hall, G., Alvarez-Icaza, M., Garde, A.,
         Ryan, J., Barry, L., Svehla, G., Arrigan, D.W.M. and Schniffner, L.
         (1996) An ASIC-based system for stripping voltammetric determination
         of trace metals. *Sensors and Actuators B: Chemical*, 34 (1-3), 466-470.

49.   Despagne, F. and Massart, D.L. (1998) Neural Networks in Multivariate
         Calibration. *The Analyst*, 123 (11), 157-178.

50.   DiabeteSuffolk (2006) *Blood Glucose Meters*. Available at:
      www.diabetesuffolk.com/Managing%20Diabetes/Meters/Table%20of%
      20meters.asp.

51.   Diaz-Cruz, M.S., Lopez de Alda, M.J., Lopez, R. and Barcelo, D. (2003)
      Determination of estrogens and progestogens by mass spectrometric
      techniques (GC/MS, LC/MS and LC/MS/MS). *Journal of Mass
      Spectrometry*, 38, 917-923.

52.   Donachie, A., Walmsley, A.D. and Haswell, S.J. (1999) Application and
      comparisons of chemometric techniques for calibration modelling using
      electrochemical/ICP-MS data for trace elements in UHQ water and
      humic acid matrices. *Analytica Chimica Acta*, 378 (1-3), 235-243.

53.   Doong, R.-A., Shih, H.-M. and Lee, S.-H. (2005) Sol-gel-derived array DNA
      biosensor for the detection of polycyclic aromatic hydrocarbons in water
      and biological samples. *Sensors and Actuators B*, 111-112, 323-330.

54.   Ek, P. and Hulden, S.-G. (1987) A Continuous Hydride-Generation System for
      Direct Current Plasma Atomic Emission Spectrometry (DCP-AES):
      Determination of Arsenic and Selenium. *Talanta*, 34 (5), 495-502.

55.   Epperson, P.M. and Denton, M.B. (1989) Binning Spectral Images in a Charge-
      Coupled Device. *Analytical Chemistry*, 61, 1513-1519.

56.   Ergon, R. (2003) Compression Into Two-Component PLS Factorisations.
      Journal of Chemometrics, 17, 303-312.

57.   Erickson, C.L., Lysaght, M.J. and Callis, J.B. (1992) Relationship between
      digital filtering and multivariate regression in quantitative analysis.
      Analytical Chemistry, 64 (24), 1155A-1163A.

58.   Eriksen, R.S., Nowak, B. and van Dam, R.A. (2001) Copper speciation and
      toxicity in a contaminated estuary. *Supervising Scientist Report*, 1-23.
      1325-1554.

59.   Eriksson, L., Trygg, J., Johansson, E., Bro, R. and Wold, S. (2000) Orthogonal
      signal correction, wavelet analysis, and multivariate calibration of
      complicated process fluorescence data. *Analytica Chimica Acta*, 420,
      181-195.

60.   Espinosa-Mansilla, A., de la Pena, A.M., Salinas, F. and Zamoro, A. (1992a)
           Simultaneous determination of pesticides by multivariate spectral
           analysis and derivative spectrophotometry. *Analytica Chimica Acta*, 258
           (1), 47-53.

61.   Espinosa-Mansilla, A., de la Pena, A.M., Salinas, F. and Zamoro, A. (1992b)
           Simultaneous determination of pesticides by multivariate spectral
           analysis and derivative spectrophotometry. *Analytica Chimica Acta*, 258
           (1), 47-53.

62.   Espinosa-Mansilla, A., Munoz de la Pena, A., Gomez, D.G. and Salinas, F.
           (2005) Photoinduced spectrofluorimetric determination of
           fluoroquinolones in human urine by using three- and two-way
           spectroscopic data and multivariate calibration. *Analytica Chimica Acta*,
           531, 257-266.

63.   Espinosa-Mansilla, A., Munoz de la Pena, A., Salinas, F. and Gomez, D.G.
           (2004) Partial least squares multicomponent fluorimetric determination
           of fluoroquinolones in human urine samples. *Talanta* , 62, 853-860.

64.   Estela, J.M., Tomas, C., Cladera, A. and Cerda, V. (1995) Potentiometric
           Stripping Analysis: A Review. *Critical Reviews in Analytical Chemistry*,
           25 (2), 91-141.

65.   Fahnrich, K.A., Pravda, M. and Guilbault, G.G. (2003) Disposable
           amperometric immunosensor for the detection of polycyclic aromatic
           hydrocarbons (PAHs) using screen-printed electrodes. *Biosensors and
           Bioelectronics*, 18, 73-82.

66.   Fearn, T. (2000) On orthogonal signal correction. *Chemometrics and Intelligent
           Laboratory Systems*, 50 (1), 47-52.

67.   Feeney, R. and Kounaves, S.P. (2002) Voltammetric measurement of arsenic in
           natural waters. *Talanta*, 58, 23-31.

68.   Ferguson, J.E. (1990) *The Heavy Elements: Chemisty, Environmental Impact
           and Health Effects*. UK: Pergamon Press.

69.   Fernandez-Bobes, C., Fernandez-Abedul, M.T. and Costa-Garcia, A. (1998)
           Anodic Stripping of Heavy Metals Using a Hanging Mercury Drop
           Electrode in a Flow System. *Electroanalysis*, 10 (10), 701-706.

70.   Fernandez Pierna, J.A., Baeten, V., Michotte Renier, A., Cogdill, R.P. and Dardenne, P. (2004) Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *Journal of Chemometrics*, 18, 341-349.

71.   Ferrer, R., Beltran, J.L. and Guiteras, B.J. (1998) Multivariate calibration applied to synchronous fluorescence sprectrometry. Simultaneous determination of polycyclic aromatic hydrocarbons in water samples. *Talanta*, 45, 1073-1080.

72.   Ferrer, R., Guiteras, J. and Beltran, J.L. (1999) Artificial neural networks (ANNs) in the analysis of polycyclic aromatic hydrocarbons in water samples by synchronous fluorescence. *Analytica Chimica Acta*, 384, 261-269.

73.   Forina, M., Casolino, C. and Almansa, E.M. (2003) The Refinement of PLS models by iterative weighting of predictor variables and objects. *Chemometrics and Intelligent Laboratory Systems*, 68, 29-40.

74.   Forshed, J., Torgrip, R.J.O., Aberg, K.M., Karlberg, B., Lindberg, J. and Jacobsson, S.P. (2005) A comparison of methods for alignment of NMR peaks in the context of cluster analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 38, 824-832.

75.   Frank, I.E. (1995) Modern non-linear regression methods. *Chemometrics and Intelligent Laboratory Systems*, 27, 1-9.

76.   Galvao, R.K.H., Jose, G.E., Filho, H.A.D., Araujo, M.C.U., da Silva, E.C., Paiva, H.M., Saldanha, T.C.B. and Nunes de Souza, E.S.O. (2004) Optimal wavelet filter construction using X and Y data. *Chemometrics and Intelligent Laboratory Systems*, 70 (1), 1-10.

77.   Goldwater, L.J. (1972) *Mercury: A History of Quicksilver*. Baltimore, MD, USA: York Press.

78.   Goto, M. and Shimada, K. (1986) Rapid-Scanning Electrochemical Detector with Micro Working Electrode for Micro High Performance Liquid Chromatography. *Chromatographia*, 21 (11), 631-634.

79.   Goyer, R. (1991) *Toxic Effects of Metals*. New York, US: Pergamon Press.

80.   Greenwood, N.N. & Earnshaw, A. (1990) *Chemistry of the Elements*. UK:
           Pergamon Press plc.

81.   Guiberteau, A., Diaz, T.G., Salinas, F. and Ortiz, J.M. (1995) Indirect
           voltammetric determination of carbaryl and carbofuran using partial least
           squares calibration. *Analytica Chimica Acta*, 305 (1-3), 219-226.

82.   Gutsche, C.D. and Muthukrishnan, R. (1978) Calixarenes: 1. Analysis of the
           Product Mixtures Produced by the Base-Catalysed Condensation of
           Formaldehyde and Para-Substituted Phenols. *Journal of Organic
           Chemistry*, 43 (25), 4905-4906.

83.   Haaland, D.M. and Thomas, E.V. (1988a) Partial Least-Squares Methods for
           Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods
           and the Extraction of Qualitative Information. *Analytical Chemistry*, 60,
           1193-1202.

84.   Haaland, D.M. and Thomas, E.V. (1988b) Partial Least-Squares Methods for
           Spectral Analyses. 2. Application to Simulated and Glass Spectral Data.
           Analytical Chemistry, 60, 1202-1208.

85.   Hagan, M.T., Demuth, H.B. & Beale, M. (1996) *Neural Network Design*.
           Boston: International Thompson Publishing.

86.   Hassan, H.N.A., Hassouna, M.E.M. and Habib, I.H.I. (1998) Multivariate
           analysis of Cd(II), In(III), Tl(I) and Pb(II) in mixtures using square wave
           anodic stripping voltammetry. *Talanta*, 46 (5), 1195-1203.

87.   Hawley, G.G. (1987) *The Condensed Chemical Dictionary*. New York: Van
           Norstrand Reinhold Company.

88.   Herrero, A. and Ortiz, M.C. (1997) Multivariate calibration transfer applied to
           the routine polarographic determination of copper, lead, cadmium and
           zinc. *Analytica Chimica Acta*, 348, 51-59.

89.   Herrero, A. and Cruz Ortiz, M. (1999) Modelling the matrix interference of iron
           in the multivariate determination of copper by stripping voltammetry:
           Instrument standardization. *Talanta*, 49 (4), 801-811.

90.   Herrero, A. and Ortiz, M.C. (1998) Modelling the background current with
           partial least squares regression and transference of the calibration models

in the simultaneous determination of Tl and Pb by stripping voltammetry. *Talanta*, 46 (1), 129-138.

91.   Hirst, D.M. (1991) *Mathematics for Chemists*. Hong Kong: Macmillan Education Ltd.

92.   Hocevar, S.B., Wang, J., Deo, R.P. and Ogorevc, B. (2002) Potentiometric Stripping Analysis at Bismuth-Film Electrode. *Electroanalysis*, 14 (2), 112-115.

93.   Hoffman, L.C. and Nakayama, T. (1968) Screen Printed Capacitor Dielectrics. *Microelectronics and Reliability*, 7, 131-135.

94.   Honeychurch, K.C. , Hart, J.P., Cowell, D.C. and Arrigan, D.W.M. (2001) Voltammetric studies of lead at calixarene modified screen-printed carbon electrodes and its trace determination in water by stripping voltammetry. *Sensors and Actuators B: Chemical*, 77 (3), 642-652.

95.   Hosohara, K., Kozuma, H., Kawasuki, K. and Tsurata, T. (1961) *J. Chem Soc Japan*, 82, 1107

96.   Huang, D., Liao, F., Molesa, S., Redinger, D. and Subramanian, V. (2003) Plastic-Compatible Low Resistance Printable Gold Nanoparticle Conductors for Flexible Electronics. *Journal of the Electrochemical Society*, 150 (7), G412-G417

97.   Huang, H. and Dasgupta, P.K. (1999) A field-deployable instrument for the measurement and speciation of arsenic in potable water. *Analytica Chimica Acta*, 380, 27-37.

98.   ISA (2004) Anonymous, *FluoroMax-2 with DataMax for Windows: Hardware Operation Manual*. USA: ISA Jobin Yvon-Spex

99.   Jaenicke, S., Sabarathinam, R.M., Fleet, B. and Gunasingham, H. (1998) Determination of lead in blood by hydrodynamic voltammetry in a flow injection system with wall-jet detector. *Talanta*, 45, 703-711.

100. Jagner, D., Renman, L. and Stefansdottir, S.H. (1994) Determination of Arsenic by Stripping Potentiometry on Gold Electrodes Using Partial Least Squares (PLS) Regression Calibration. *Electroanalysis*, 6, 201-208.

101. Janska, M., Tomaniova, M., Hajslova, J. and Kocourek, V. (2004) Appraisal of "classic" and "novel" extraction procedure efficiencies for the isolation of polycyclic aromatic hydrocarbons and their derivatives from biotic matrices. *Analytica Chimica Acta*, 520, 93-103.

102. Johansson, J., Cauchi, M. and Sundgren, M. (2002) Multiple fiber-optic dual-beam UV/Vis system with application to dissolution testing. *Journal of Pharmaceutical and Biomedical Analysis*, 29, 469-476.

103. Keller, O.C. and Buffle, J. (2000a) Voltammetric and Reference Microelectrodes with Integrated Microchannels for Flow through Microvoltammetry. 1. The Microcell. *Analytical Chemistry*, 72, 936-942.

104. Keller, O.C. and Buffle, J. (2000b) Voltammetric and Reference Microelectrodes with Integrated Microchannels for Flow through Microvoltammetry. 2. Coupling the Microcell to a Supported Liquid Membrane Preconcentration Technique. *Analytical Chemistry*, 72, 943-948.

105. Kemp, W. (1991) *Organic Spectroscopy*. Edinburgh, UK: Macmillan.

106. Kennard, R.W. and Stone, L.A. (1969) Computer Aided Design of Experiments. *Technometrics*, 11 (1), 137-148.

107. Kennedy, J.H. (1990) *Analytical Chemistry*. New York, USA: Saunders College Publishing.

108. Kroger, S. and Turner, A.P.F. (1997) Solvent-Resistant Carbon Electrodes Screen Printed onto Plastic for Use in Biosensors. *Analytica Chimica Acta*, 347, 9-18.

109. Kuster, M., Lopez da Alda, M.J. and Barcelo, D. (2004) Analysis and distribution of estrogens and progestogens in sewage sludge, soils and sediments. *Trends in Analytical Chemistry*, 23 (10-11), 790-798.

110. Lajunen, L.H.J. (1992) *Spectrochemical Analysis by Atomic Absorption Emission*. England: RSC.

111. Lam, M.H.-W., Tjia, A.Y.-W., Chan, C.-C., Chan, W.-P. and Lee, W.-S. (1997) Speciation Study of Chromium, Copper and Nickel in Coastal Estuarine Sediments Polluted by Domestic and Industrial Effluents. *Marine*

Pollution Bulletin, 34 (11), 949-959.

112. Lambert, J. (1997) *Contaminated Land and its Reclammation*. Cambridge, UK: RSC.

113. Lastres, E., de Armas, G., Catasús, M., García, L. and Cerda, V. (1997) Use of Neural Networks in Solving Interferences Caused by Formation of Intermetallic Compounds in Anodic Stripping Voltammetry. *Electroanalysis*, 9 (3), 251-254.

114. Li, H. and Smart, R.B. (1996) Determination of sub-nanomolar concentration of arsenic (III) in natural waters by square wave cathodic stripping voltammetry. *Analytica Chimica Acta*, 325, 25-32.

115. Lindgren, F., Geladi, P. and Wold, S. (1993) The Kernal Algorithm for PLS. *Journal of Chemometrics*, 7, 45-60.

116. Luis, M.L., Fraga, J.M.G., Jimenez, A.I., Jimenez, F., Hernandez, O. and Arias, J.J. (2004) Application of PLS regression to fluorimetric data for the determination of furosemide and triamterene in pharmaceutical preparations and triamterene in urine. *Talanta*, 62, 307-316.

117. Lukaszewski, Z., Pawlak, M.K. and Ciszewski, A. (1980) Determination of thallium and lead in cadmium salts by anodic stripping voltammetry with addition of surfactants to suppress the cadmium peaks. *Talanta*, 27 (2), 181-185.

118. Lundh, M., Engblom, K. & Holstrom, E. (2003) Gothenburg, Sweden: The Gothenburg Region Association of Local Authorities (GR).

119. Manahan, S.E. (1994) *Environmental Chemistry*. UK: Lewis Publisher.

120. Marhaba, T.F., Bengraine, K., Pu, Y. and Arago, J. (2003) Spectral fluorescence signatures and partial least squares regression: model to predict dissolved organic carbon in water. *Journal of Hazardous Materials*, B97, 83-97.

121. Martens, H. & Naes, T. (2001) *Multivariate Calibration*. Guildford, UK: John Wiley & Sons Ltd.

122. Martin, M.Z., Labbe, N., Rials, T.G. and Wullschleger, S.D. (2005) Analysis of

preservative-treated wood by multivariate analysis of laser-induced breakdown spectroscopy spectra. *Spectrochimica Acta Part B*, 60, 1179-1185.

123. Martos, N.R., Diaz, A.M., Navalon, A., Paya, I.D.O. and Capitan-Vallvey, L.F. (2000) Simultaneous spectrofluorimetric determination of (acetyl)salicylic acid, codeine and pyridoxine in pharmaceutical preparations using partial least-squares multivariate calibration. *Journal of Pharmaceutical and Biomedical Analysis*, 23, 837-844.

124. Mathworks (2002) *MATLAB: Getting Started with MATLAB*. MA, USA: MathWorks Incorporated.

125. Matsumoto, H., Inukai, J. and Ito, M. (1994) Structures of copper and halides on Pt(111), Pt(100) and Au(111) electrode surfaces studied by in-situ scanning tunneling microscopy. *Journal of Electroanalytical Chemistry*, 379, 223-231.

126. Meras, I.D., Diaz, T.G. and Franco, M.A. (2005) Simultaneous fluorimetric determination of glyphosate and its metabolite, aminomethylphosphonic acid, in water, previous derivatisation with NBD-Cl and by partial least squares calibration (PLS). *Talanta*, 65, 7-14.

127. Moreira, A.B., Dias, I.L.T., Neto, G.O., Zagatto, E.A.G., Ferreira, M.M.C. and Kubota, L.T. (2005) Solid-phase spectrofluorimetric determination of acetylsalicylic acid and caffeine in pharmaceutical preparations using partial least-squares multivariate calibration. *Talanta*, 67, 65-69.

128. Morrison, R.T. & Boyd, R.N. (1992) *Organic Chemistry*. New Jersey, USA: Prentice Hall International Inc..

129. Mueller-Hoecker, J., Meyer, U. and Wiebecke, B. (1988) Copper storage disease of the liver and chronic dietary copper intoxication in two further German infants mimicking Indian childhood cirrhosis . *Pathology - Research and Practice*, 183, 39-45.

130. Navalon, A., Blanc, R., del Olmo, M. and Vilchez, J.L. (1999) Simultaneous determination of naproxen, salicylic acid and acetylsalicylic acid by spectrofluorimetry using partial least-squares (PLS) multivariate calibration. *Talanta*, 48, 469-475.

131. Ni, Y. and Gong, X. (1997) Simultaneous spectrophotometric determination of

mixtures of food colorants. *Analytica Chimica Acta*, 354 (1-3), 163-171.

132. Ni, Y. and Jin, L. (1999) Simultaneous polarographic chemometric determination of lead, copper, vanadium, cadmium and nickel. *Chemometrics and Intelligent Laboratory Systems*, 45 (1-2), 105-111.

133. Nordberg, G.F., Kjelstrom, T. & Nordberg, M. (1985) *Kinetics and Metabolism*. US: CRC Press.

134. Nordqvist, M., Thakker, D.R. and Vyas, K.P. (1981) Metabolism of Chrysene and Phenanthrene to Bay-Region Diol Epoxides by Rat Liver Enzymes. *Molecular Pharmacology*, 19, 168-178.

135. O'Neil, P. (1993) *Environmental Chemistry* . UK: Chapman & Hall.

136. Ortega-Algar, S. , Ramos-Martos, N. and Molina-Diaz, A. (2003) A flow-through fluorimetric sensing device for determination of Alpha- and Beta-Naphthol mixtures using a partial least-squares multivariate calibration approach. *Talanta*, 60, 313-323.

137. Osteryoung, J.G. and Osteryoung, R.A. (1985) Square Wave Voltammetry. *Analytical Chemistry*, 57 (1), 101A-110A.

138. Otto, M. (1999a) *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*. Germany: Wiley-VCH.

139. Otto, M. (1999b) *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*. Germany: Wiley-VCH.

140. Palchetti, I., Cagnini, A., Mascini, M. and Turner, A.P.F. (1999) Characterisation of Screen-Printed Electrodes for Detection of Heavy Metals. *Mikrochimica Acta*, 131, 65-73.

141. Palchetti, I., Marrazza, G. and Mascini, M. (2001) New Procedures to Obtain Electrochemical Sensors for Heavy Metal Detection. *Analytical Letters*, 34 (6), 813-824.

142. Patra, D. and Mishra, A.K. (2002) Recent Developments in Multi-Component Synchronous Fluorescence Scan Analysis. *Trends in Analytical Chemistry*, 21 (12), 787-798.

143. Pearce, F. (2003) Arsenic's Fatal Legacy Grows Worldwide. *New Scientist.com*,

144. Persson, T. and Wedborg, M. (2001) Multivariate evaluation of the fluorescene of aquatic organic matter. *Analytica Chimica Acta*, 434, 179-192.

145. Prado, C., Wilkins, S.J., Marken, F. and Compton, R.G. (2002) Simultaneous Electrochemical Detection and Determination of Lead and Copper at Boron-Doped Diamond Film Electrodes. *Electroanalysis*, 14 (4), 262-272.

146. Prest, J.E., Baldock, S.J., Fielden, P.R., Goddard, N.J. and Treves-Brown, B.J. (2003) Miniaturised isotachophoretic analysis of inorganic arsenic speciation using a planar polymer chip with integrated conductivity detection. *Journal of Chromatography A*, 990, 325-334.

147. Relfe, S. (2003) Arsenic Toxicity. *www.relfe.com*,

148. Ribeiro, F.A.L. and Ferreira, M.M.C. (2005) QSAR model of the phototoxicity of polycyclic aromatic hydrocarbons. *Journal of Molecular Structure: THEOCHEM*, 719, 191-200.

149. Richards, E., Bessant, C. and Saini, S. (2002) Multivariate Data Analysis in Electroanalytical Chemistry. *Electroanalysis*, 14 (22), 1533-1542.

150. Richards, E., Bessant, C. and Saini, S. (2003) A liquid handling system for the automated acquisition of data for training, validating and testing calibration models. *Sensors and Actuators B: Chemical*, 88 (2), 149-154.

151. Richards, E.C.L. (2003) A Generic Method for Optimal Multivariate Calibration of Voltammetric Data. Cranfield University at Silsoe.

152. Rodriguez-Mozaz, S., Lopez da Alda, M.J. and Barcelo, D. (2004) Monitoring of estrogens, pesticides and bisphenol A in natural waters and drinking water treatment plants by solid-phase extraction-liquid chromatography-mass spectrometry. *Journal of Chromatography A*, 1045, 85-92.

153. Rouessac, F. & Rouessac, A. (2000) *Chemical Analysis: Instrumentation Methods and Techniques*. Chichester, UK: John Wiley & Sons Ltd.

154. Sadana, R.S. (1983) Determination of Arsenic in the Presence of Copper by Differential Pulse Cathodic Stripping Voltammetry at a Hanging

Mercury Drop Electrode. *Analytical Chemistry*, 55, 304-307.

155. Samek, O., Beddows, D.C.S., Telle, H.H., Kaiser, J., Liska, M., Caceres, J.O. and Urena, A.G. (2001) Quantitative laser-induced breakdown spectroscopy analysis of calcified tissue samples. *Spectrochimica Acta Part B*, 56, 865-875.

156. Sandmeyer, E.E. (1981) *Aromatic Hydrocarbons*. New York: John Wiley & Sons.

157. Saurina, J., Leal, C., Compano, R., Granados, M., Tauler, R. and Prat, M.D. (2000) Determination of triphenyltin in sea-water by excitation-emission matrix fluorescence and multivariate curve resolution. *Analytica Chimica Acta*, 409, 237-245.

158. Savitzky, A. and Golay, M.J.E. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36 (8), 1627-1639.

159. Sax, N.I. & Lewis, R.L., (Eds.) (1987) *Hawley's Condensed Chemical Dictionary*.New York: Van Nostrand Reinhold Company.

160. Serradell, M., Izquierdo, S., Moreno, L., Merkoçi, A. and Alegret, S. (2002) Mercury-Free PSA of Heavy Metals Using Graphite-Epoxy Composite Electrodes. *Electroanalysis*, 14 (18), 1281-1287.

161. Sharp, D.W.A. (1990) *Dictionary of Chemistry*. London, England: Penguin Books.

162. Sjostrom, M., Wold, S., Lindberg, W., Persson, J.-A. and Martens, H. (1983) A Multivariate Calibration Problem in Analytical Chemistry Solved by Partial Least-Squares Models in Latent Variables. *Analytica Chimica Acta*, 150, 61-70.

163. Skoog, D.A. (1985) *Principles of Instrumental Analysis*. USA: Saunders College Publishing International.

164. Skoog, D.A., West, D.M. & Holler, F.J. (1996) *Fundamentals of Analytical Chemistry*. Fort Worth, Texas, USA: Saunders College Publishing.

165. Sorouraddin, M.-H., Rashidi, M.-R., Ghorbani-Kalhor, E. and Asadpour-Zeynali, K. (2005) Simultaneous spectrofluorimetric and

spectrophotometric determination of melatonin and pyridoxine in pharmaceutical preparations by multivariate calibration methods. *Il Farmaco*, 60, 451-458.

166. Starodub, N.F., Kanjuk, N.I., Kukla, A.L. and Shirshov, Y.M. (1999) Multi-enzymatic electrochemical sensor: field measurements and their optimisation. *Analytica Chimica Acta*, 385, 461-466.

167. Stollery, B.T., Broadbent, D.E., Banks, H.A. and Lee, W.R. (1991) Short-term prospective study of cognitive functioning in lead workers. *British Journal of Industrial Medicine*, 48, 739-749.

168. Storer, J.S., DeLeon, I. and Millikan, L.E. (1984) Human Absorption of Crude Coal Tar Products. *Archives of Dermatology*, 120, 874-877.

169. Stuyfzand, P.J. (2004) Quantifying the hydrogeochemical impact and sustainability of artificial recharge systems. *Artificial Recharge Demonstration Project (EVK1-CT2002-00114)*, 1-10.

170. Svancara, I., Vytras, K., Bobrowski, A. and Kalcher, K. (2002) Determination of arsenic at a gold-plated carbon paste electrode using constant current stripping analysis. *Talanta*, 58, 45-55.

171. Sweedler, J.V., Shear, J.B., Fishman, H.A., Zare, R.N. and Scheller, R.H. (1991) Fluorescence Detection in Capillary Zone Electrophoresis Using a Charge-Coupled Device with Time-Delayed Integration. *Analytical Chemistry*, 63, 496-502.

172. Thennadil, S.N. and Martin, E.B. (2005) Empirical preprocessing methods and their impact on NIR calibrations: a simulation study. *Journal of Chemometrics*, 19 (2), 77-89.

173. Thissen, U., Pepers, M., Ustun, B., Melssen, W.J. and Buydens, L.M.C. (2004) Comparing support vector machines to PLS for spectral regression applications. *Chemometrics and Intelligent Laboratory Systems*, 73, 169-179.

174. Thomas, P. (2003) *Metal Analysis in Chemical Analysis of Contaminated Land*. Oxford, UK: Blackwell Publishing.

175. Thompson, J.E. and Pardue, H.L. (1983) Synchronous Fluorescence

Spectroscopy with a Silicon-Intensified Target Vidicon. *Analytica Chimica Acta*, 152, 73-82.

176. Todeschini, R., Galvagni, D., Vilchez, J.L., del Olmo, M. and Navas, N. (1999) Kohonen artificial neural networks as a tool for wavelength selection in multicomponent spectrofluorimetric PLS modelling: application to phenol, o-cresol, m-cresol and p-cresol mixtures. *Trends in Analytical Chemistry*, 18 (2), 93-98.

177. Tomaniova, M., Hasjlova, J., Pavelka (Jnr), J., Kocourek, V., Holadova, K. and Klimova, I. (1998) Microwave-assisted solvent extraction - a new method for isolation of polynuclear aromatic hydrocarbons from plants. *Journal of Chromatography A*, 827, 21-29.

178. Torralba, R., Bonilla, M., Perez-Arribas, L.V., Palacios, A. and Camara, C. (1994) ANALYTICAL NOTE: Speciation and simultaneous determination of arsenic (III), arsenic (V), monomethylarsonate and dimethylarsinate by atomic absorption using inverse least sqaures multivariate calibration. *Spectrochimica Acta*, 49B (9), 893-899.

179. Trojanowicz, M., Szewczynska, M. and Wcislo, M. (2003) Electroanalytical Flow Measurements - Recent Advances. *Electroanalysis*, 15 (5-6), 347-365.

180. Turner, P., Milne, G., Turner, S., Kubitscheck, M. and Penman, I. (2004) Towards the Wireless Ward: Evaluating a Trial of Networked PDAs in the National Health Service. *Lecture Notes in Computer Science*, 2954, 202-214. 0302-9743.

181. Vvedenskii, A.V. and Grushevskaya, S.N. (2003) Kinetic peculiarities of anodic dissolution of copper and its gold alloys accompanied by the formation of insoluble Cu (I) products. *Corrosion Science*, 45, 2391-2413.

182. Walcerz, I., Glab, S. and Koncki, R. (1998) Potentiometric enzyme electrode in a flow injection system for the determination of urea in human serum samples. *Analytica Chimica Acta*, 369, 129-137.

183. Wang, J. (1985) *Stripping Analysis: Principles, Instrumentation and Applications*. Florida, USA: VCH Publishers.

184. Wang, J. (1994) *Analytical Electrochemistry*. New York, NY, USA: VCH

Publishers.

185. Wang, J. and Chen, Q. (1994) Screen-Printed Glucose Strip Based on Palladium-dispersed Carbon Ink. *Analyst*, 119, 1849-1851.

186. Wang, J. and Tian, B. (1993a) Mercury-Free Disposable Lead Sensors Based on Potentiometric Stripping Analysis at Gold-Coated Screen-Printed Electrodes. *Analytical Chemistry*, 65 (11), 1529-1532.

187. Wang, J. and Tian, B. (1993b) Screen-Printed Electrodes for Stripping Measurements of Trace Mercury. *Analytica Chimica Acta*, 274, 1-6.

188. Wang, J., Tian, B., Wang, J., Lu, J., Olsen, C., Yarnitzky, C., Olsen, K., Hammerstrom, D. and Bennett, W. (1999) Stripping analysis into the 21st century: faster, smaller, cheaper, simpler and better. *Analytica Chimica Acta*, 385, 429-435.

189. Wang, J. and Varughese, K. (1990) Polishable and Robust Biological Electrode Surfaces. *Analytical Chemistry*, 62 (3), 318-320.

190. Wang, Y., Lysaght, M.J. and Kowalski, B.R. (1992) Improvement of Multivariate Calibration through Instrument Standardisation. *Analytical Chemistry*, 64, 562-564.

191. Wang, Y., Veltkamp, D.J. and Kowalski, B.R. (1991) Multivariate Instrument Standardisation. *Analytical Chemistry*, 63, 2750-2756.

192. Wang, Z., Dean, T. and Kowalski, B.R. (1995) Additive Background Correction in Multivariate Instrument Standardisation. *Analytical Chemistry*, 67, 2379-2385.

193. Wiberg, K., Sterner-Molin, A. and Jacobsson, S.P. (2004) Simultaneous determination of albumin and immunoglobulin G with fluorescence spectroscopy and multivariate calibration. *Talanta*, 62, 567-574.

194. Wingfors, H., Sjodin, A., Haglund, P. and Brorstrom-Lunden, E. (2001) Characterisation and determination of profiles of polycyclic aromatic hydrocarbons in a traffic tunnel in Gothenburg, Sweden. *Atmospheric Environment*, 35, 6361-6369.

195. Winquist, F., Bjorklund, R., Krantz-Rulcker, C., Lundstrom, I. , Ostergren, K.

and Skoglund, T. (2005) An electronic tongue in the dairy industry. *Sensors and Actuators B*, 111-112, 299-304.

196. Winquist, F., Holmin, S., Krantz-Rulcker, C., Wide, P. and Lundstrom, I. (2000) A hybrid electronic tongue. *Analytica Chimica Acta*, 406, 147-157.

197. Winquist, F., Wide, P. and Lundstrom, I. (1997) An electronic tongue based on voltammetry. *Analytica Chimica Acta*, 357, 21-31.

198. Wise, B.M. & Gallagher, N.B. (1998a) *PLS_Toolbox 2.0*. WA, USA: Eigenvector Research Incorporated.

199. Wise, B.M. & Gallagher, N.B. (1998b) *PLS_Toolbox 2.0*. WA, USA: Eigenvector Research Incorporated.

200. Wold, S., Antti, H., Lindgren, F. and Ohman, J. (1998) Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 44 (1-2), 175-185.

201. Wold, S., Esbensen, K. and Geladi, P. (1987) Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2 (1-3), 37-52.

202. World Health Organisation (WHO) (2000) *Towards an assessment of the socioeconomic impact of arsenic poisoning in Bangladesh*. Available at: http://www.who.int/water_sanitation_health/dwq/arsenic2/en/print.html. Accessed 2004.

203. Zen, J.-M., Chung, H.-H. and Kumar, A.S. (2000) Determination of lead (II) on a copper/mercury-plated screen-printed electrode. *Analytica Chimica Acta*, 421, 189-197.

204. Zhang, L., Jiang, J.-H., Liu, P., Liang, Y.-Z. and Yu, R.-Q. (1997) Multivariate nonlinear modelling of fluorescence data by neural network with hidden node pruning algorithm. *Analytica Chimica Acta*, 344, 29-39.

205. Zhou, F., Aronson, J.T. and Ruegnitz, M.W. (1997) High-Throughput Fast-Scan Anodic Stripping Voltammetry in a Microflow System. *Analytical Chemistry*, 69, 728-733.

206. Zhu, Y., Coury, L.A., Long, H., Duda, C.T., Kissinger, C.B. and Kissinger, P.T. (2000) Liquid Chromatography with Multi-Channel Detection for the Determination of Resveratrol in Wine, Graper Juice and Grape Seed

Capsules with Automated Solid Phase Extraction. *J. Liq. Chrom. & Rel. Technol.*, 23 (10), 1555-1564.

207. Zima, J. and van den Berg, C.M.G. (1994) Determination of arsenic in sea water by cathodic stripping voltammetry in the presence of pyrrolidine dithiocarbamate. *Analytica Chimica Acta*, 289, 291-298.

APPENDICES

A1: The NIPALS Algorithm

Known as "the classical PLS algorithm", the steps are outlined below (Lindgren *et al.,* 1993; Wise & Gallagher, 1998b) (Wise & Gallagher, 1998a). For each latent variable, the following calculations are performed. Firstly, the score vector for the *Y* matrix is calculated. If there is only one column, the score vector, *u* is set equal to it, else it is set equal to the column with the maximum sum of squares. The *PLS weight* vector is obtained from:

$$\mathbf{w}^T = \frac{\mathbf{u}^T X}{\mathbf{u}^T \mathbf{u}}$$

A1.1

The weight is then *normalised* with the *norm* set to one.

$$w = \frac{w}{\|w\|}$$

A1.2

The corresponding X-score vector is calculated.

$$t = \frac{Xw}{w^T w}$$

A1.3

The Y-loading vector is calculated.

$$q^T = \frac{t^T Y}{t^T t}$$

A1.4

The Y-score vector is computed.

$$u = \frac{Yq}{q^T q}$$

A1.5

The u in Equation A1.5 is substituted into Equation A1.1 until *convergence* has been achieved. This is when the difference between the *new t* vector and the *old t* vector is very small (normally less than 1×10^{-10}). Once convergence has been achieved, the next set of calculations is carried out. Firstly, the X-loading vector for the latent variable is calculated.

$$p^T = \frac{t^T X}{t^T t} \qquad\qquad \textbf{A1.6}$$

The X-score, X-PLS weights and X-loading vectors are respectively calculated by treatment with the norm of p from Equation A1.6. And thus:

$$t = t \times \|p\| \qquad\qquad \textbf{A1.7}$$

$$w = w \times \|p\| \qquad\qquad \textbf{A1.8}$$

$$p = \frac{p}{\|p\|} \qquad\qquad \textbf{A1.9}$$

The *residuals* for X and Y are computed via Equations A1.10 and A1.11 respectively.

$$E = X - tp^T \qquad\qquad \textbf{A1.10}$$

$$F = Y - tq^T \qquad\qquad \textbf{A1.11}$$

For the next latent variable, E and F are substituted into the above equations for X and Y respectively, and the whole procedure is repeated until all of the latent variables selected have been accounted for. The *regression coefficient*, B is then calculated from the W, P and Q matrices.

$$B = W \left(P^T W \right)^{-1} Q^T \qquad\qquad \textbf{A1.12}$$

If the dimensions of X are $m \times n$, Y are $m \times h$, and a latent variables were selected, the corresponding dimensions for B, W, P and Q are $n \times h$, $n \times a$, $n \times a$, and $h \times a$ respectively.

A2: The SIMPLS Algorithm

SIMPLS is computationally faster than NIPALS (de Jong, 1993). Furthermore, the X-scores, *t*, in SIMPLS are *orthonormal* and the X-loadings, *p*, are not normalised. This implies that the information with regards to the variance is contained within the loadings (Wise & Gallagher, 1998b).

The first step is to mean-centre the Y matrix. The cross-product of X and Y_0 are calculated.

$$S = X^T Y_0 \qquad\qquad \textbf{A2.1}$$

For the first latent variable, the initial step is to set the Y-loading vector, *q*, to the most dominant *eigenvector* of the cross-product of S^T and S. The X-weight, *r* is calculated (A2.2) followed by the X-score vector, *t* (A2.3).

$$r = Sq \qquad\qquad \textbf{A2.2}$$
$$t = Xr \qquad\qquad \textbf{A2.3}$$

The scores are mean-centred followed by the normalisation of the scores (A2.4) and adaptation of the weights (A2.5) respectively. The X-block factor loading vector is then calculated (A2.6) followed by the Y-block loading vector (A2.7) and the corresponding Y-block factor score vector (A2.8).

$$t = \frac{t_0}{\|t_0\|} \qquad\qquad \textbf{A2.4}$$

$$r = \frac{r}{\|t_0\|} \qquad\qquad \textbf{A2.5}$$

$$p = X^T t \qquad\qquad \textbf{A2.6}$$

$$q = Y_0^T t \qquad\qquad \textbf{A2.7}$$

$$u = Y_0^T q \qquad\qquad \textbf{A2.8}$$

The orthogonal loadings are initialised by setting $v = p$. Equations A2.9 and A2.10 are calculated for all of the latent variables *except* for the first one.

$$v = v - V\left(V^T p\right) \qquad\qquad \text{A2.9}$$

$$u = u - T\left(T^T u\right) \qquad\qquad \text{A2.10}$$

The orthogonal loadings are normalised (A2.11) followed by "deflation" of the S matrix (A2.12).

$$v = \frac{v}{\|v\|} \qquad\qquad \text{A2.11}$$

$$S = S - v\left(v^T S\right) \qquad\qquad \text{A2.12}$$

All vectors calculated above are stored in their respective matrices and the procedure is repeated until all of the latent variables are accounted for. The regression coefficients matrix, **B**, is calculated below.

$$B = \left(RQ\right)^T \qquad\qquad \text{A2.13}$$

The variances for **X** and **Y** are calculated from the cross-products of **P** and **Q** respectively.

$$\operatorname{var} X = \frac{diag\left(P^T P\right)}{n-1} \qquad\qquad \text{A2.14}$$

$$\operatorname{var} Y = \frac{diag\left(Q^T Q\right)}{n-1} \qquad\qquad \text{A2.15}$$

The *diag* term implies utilization of the elements along the main diagonal of the matrix. The number of observations (rows in **X**) is given as n.

A3: The Data Analysis Package

A3.1 The Main Program

The functionality of the main program is illustrated in the flow diagram below (Figure A3.1 and A3.2).
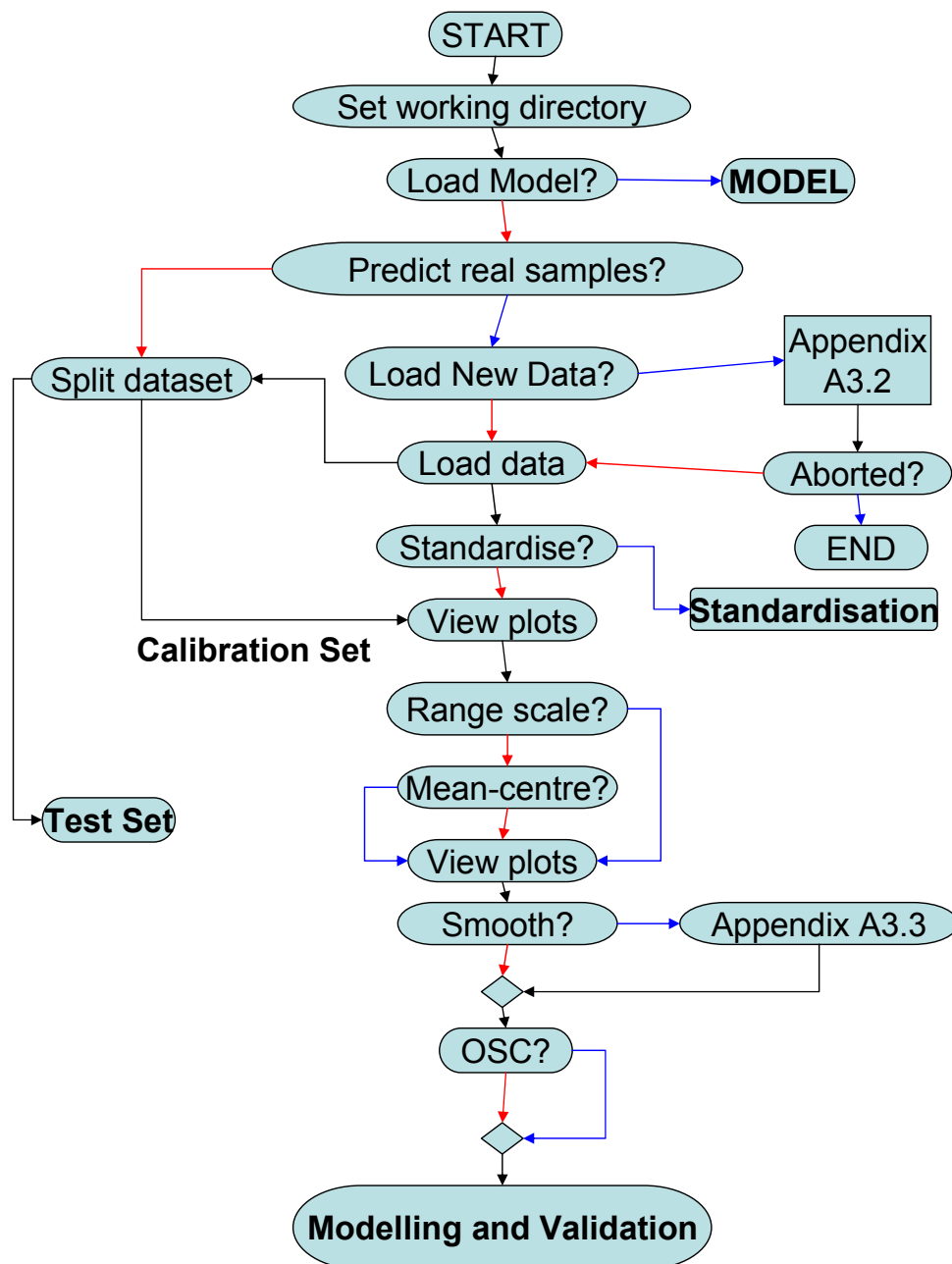

Figure A3.1: Functionality of the main program prior to modelling

Blue arrows denote positive output (YES); Red arrows denote negative output (NO); Black arrows denote normal flow
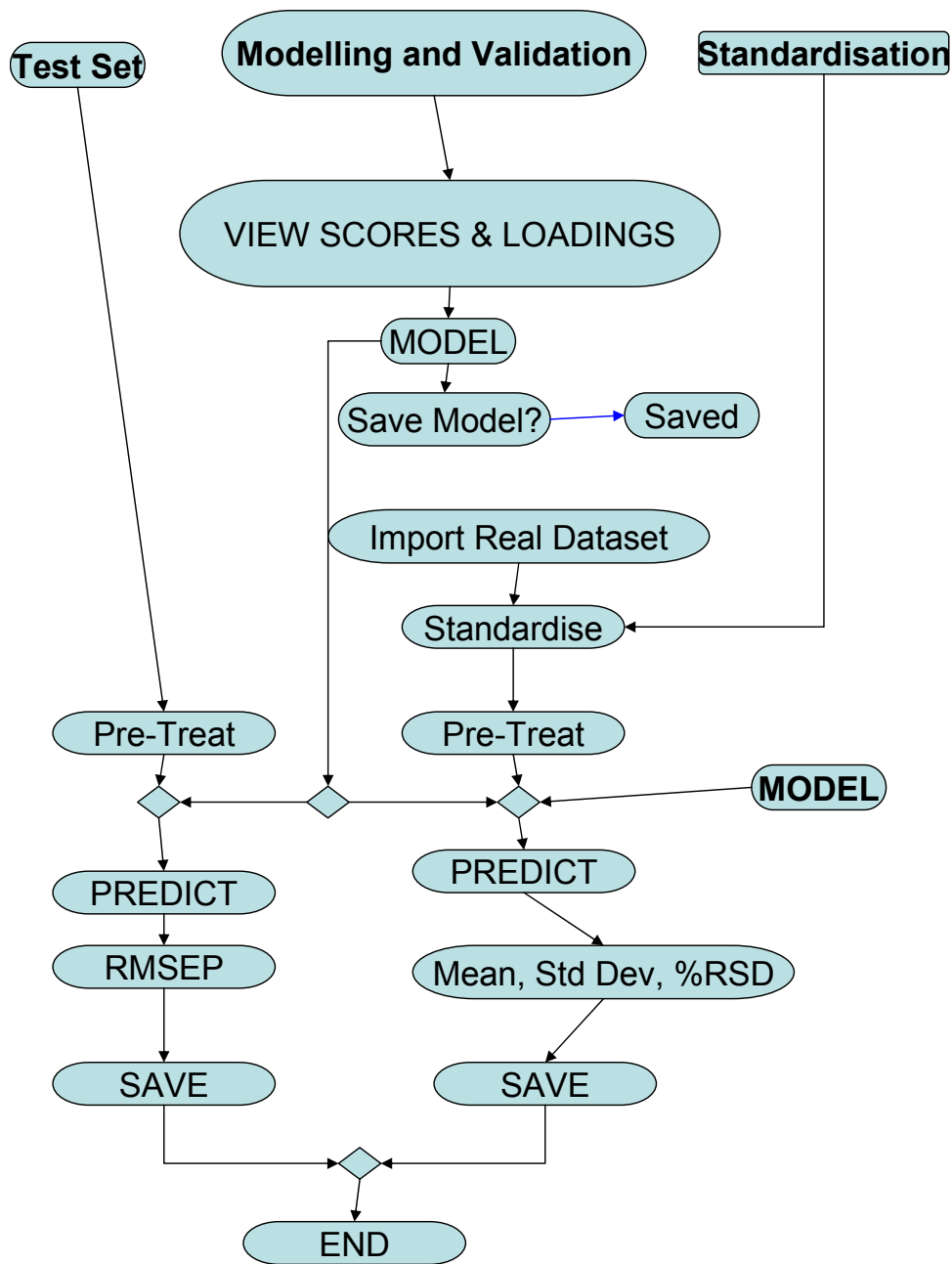
Figure A3.2: Functionality of the main program after modelling.

Blue arrows denote positive output (YES); Red arrows denote negative output (NO); Black

arrows denote normal flow

A3.2: Importation of Datasets

The functionality of the data importation program is illustrated in the flow diagram below (Figure A3.3).
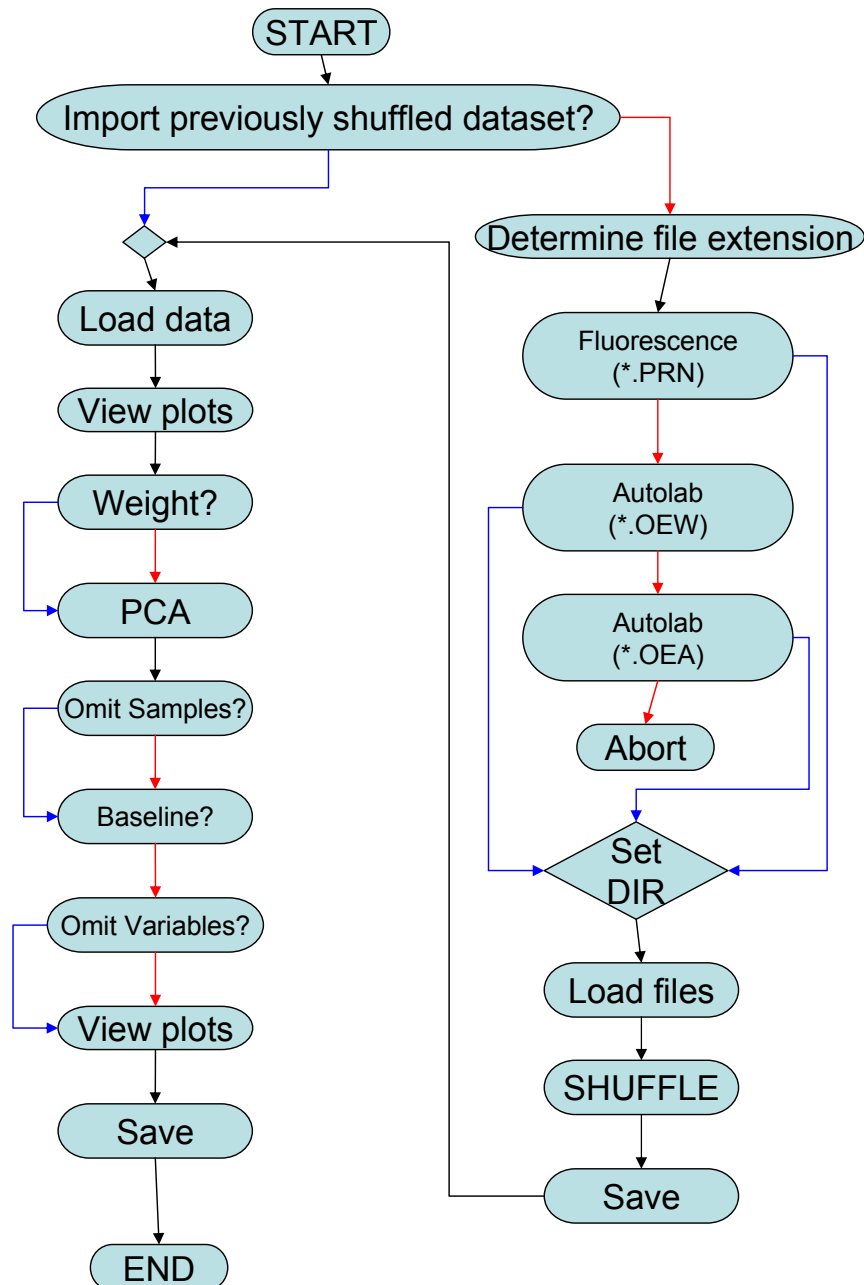

Figure A3.3: Functionality of the dataset importation program referenced in Figure A3.1.

Blue arrows denote positive output (YES); Red arrows denote negative output (NO); Black arrows denote normal flow

A3.3: Application of Smoothing

The functionality of the signal smoothing program is illustrated in the flow diagram below (Figure A3.4).
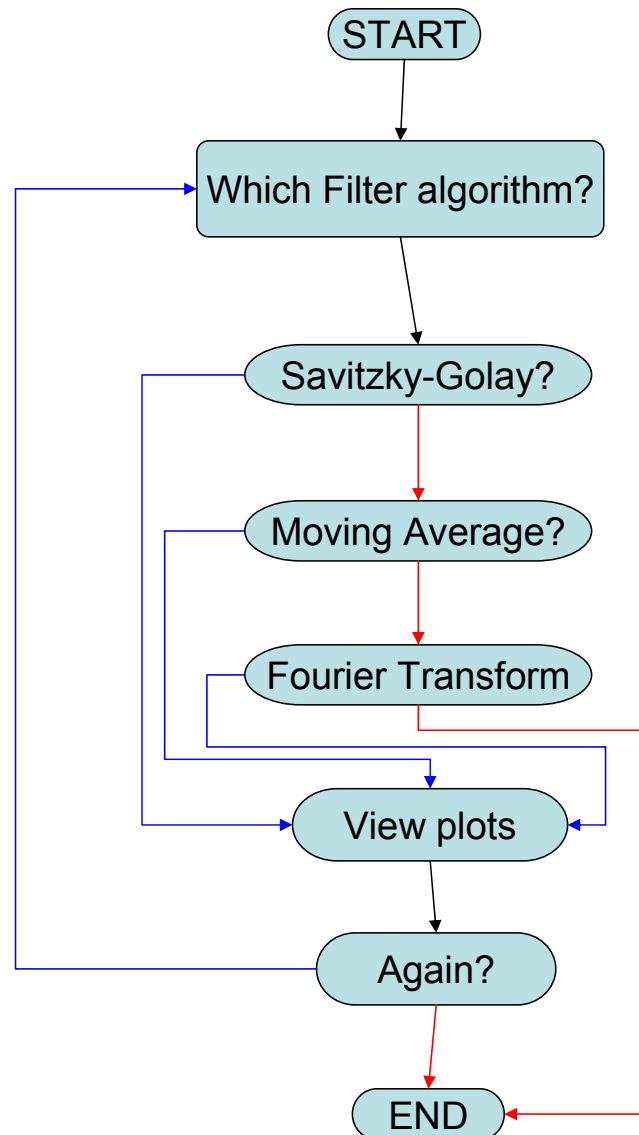

Figure A3.4: Functionality of the smoothing program referenced in Figure A3.1

Blue arrows denote positive output (YES); Red arrows denote negative output (NO); Black arrows denote normal flow

A4: Blood Glucose Meters

Table A4.1 compares five blood glucose meters manufactured for the quantitative determination of glucose in blood in which operational conditions, physical construct and calibration modes are detailed.

Table A4.1: A comparison of five blood glucose meters.

Source:

http://www.diabetesuffolk.com/Managing%20Diabetes/Meters/Table%20of%20meters.asp

Manufacturer	Meter	Strip	Blood Sample	Time (s)	Visual Reading	Range (mmol/l)	Calibration
Hypoguard	Supreme Plus	Hypoguard Supreme	Non-wipe	30-60	Yes	2.0 – 25.0	Coded chip with batch
Lifescan	Pocketscan	Pocketscan	Non-wipe	15	No	1.1 – 33.3	Calibration strip
Medisense	Precision QED	Medisense G2	Non-wipe	20	No	1.1 – 33.3	Calibration strip
Roche Diagnostics	Accu-Chek Active	Accu-Chek Active	Non-wipe	5	Yes	0.6 – 33.3	Coded chip with batch
Roche Diagnostics	Accu-Chek Compact		Non-wipe	15	Yes	0.6 – 33.3	Automatic – no calibration strip required

A5: Portable Multi-Channel Potentiostat

Below is displayed a brochure which details the functionalities of the custom-built in-house multi-channel potentiostat device.
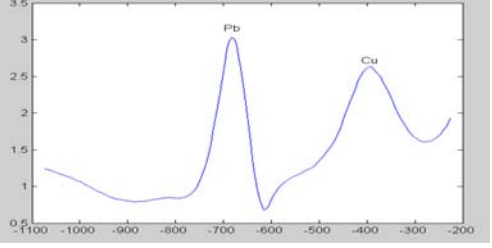

Portable Multi-Channel Potentiostat

Overview:

The four channel potentiostat is a portable and flexible instrument capable of carrying out a wide variety of electrochemical tests in the laboratory and in the field.

The device is battery operated and ruggedised, and can operate in a stand-alone mode or with an attached PC or PDA

The device is shown configured to perform heavy metal detection in an environmental application, but the flexibility afforded by its software configuration makes it suitable for a wide range of other electrochemical applications.

A typical disposable screen printed electrode, and heavy metal voltammagram

Technology:

• Four completely independent potentiostat channels

• Electrochemical drive voltages fully synthesised in software to 1mV, 5usec resolution

• Measures down to picoAmp resolution with full auto ranging.

• Onboard Microcontroller – fast, low current consumption

• Structured software design, implemented in 'C' language

• On board data storage, or real time transmission over RS232 / Bluetooth link to PC / PDA

• Calibration data stored in on-board EEPROM, allows low cost hardware components to be used.

• Battery operated, robust, portable for field use

• On board or PC based data analysis

Heavy Metal Detection Application:

• Anodic Stripping Voltammetry – An electroanalytical method, both quantitative & qualitative

• Rapid testing (4 min. analysis time), sensitive (down to 20 ppb)

• Carbon + gold electrodes allow simultaneous detection of Pb, Cu, Cd, Zn, Hg, As

• Low-cost testing (meter, disposable screen printed electrodes, reagents)

• Rapid, low-cost site screening tool

Medical Sensor Application

• End of line production batch testing device for blood glucose strips for diabetics

• Simple fixed potential application. Wide range of possible currents generated by sensors

• Simple go / no-go indication to operator –signals details also recorded for QA traceability

Contact : Paul Knight or Steve Setford, Cranfield University at Silsoe, Bedford MK45 4DT
Email p.v.knight@cranfield.ac.uk, s.j.setford@cranfield.ac.uk, Tel 01525 863563

A6: Publications

A6.1: Conferences

Poster presentation at the "First SWIFT-WFD Workshop on Validation of Robustness of sensors and Bioassays for Screening Pollutants" Conference, Mao, Menorca, Spain. (02 – 03 December 2004).

Oral presentation at the RSC Chemometrics Committee 'Emerging Chemometricians' meeting at the GMEX Centre in Manchester (28 – 29 March 2006) on "Chemometrics on a PDA". Unfortunately, the event was cancelled.

Attendance of "Informatics and Data Visualisation Conference", hosted by the Chromatography Society, INTECH Centre, Winchester, UK (21 – 22 October, 2004)

Attendance of "4th International Conference for Instrumental Methods of Analysis 2005 (IMA05)", Almedar Knossos Royal Village Hotel, Crete, Greece (02 – 06 October, 2005).

A6.2: Reports

Some of the findings of this thesis are to appear in the overall ARTDEMO final report, which is to be published in book format in the 4th quarter of 2006.

A6.3: Papers

At the time of writing, there are several papers pending.