Development of a Solid State IR Analyser System Using Chemometrics

Angela Jane Griffin

A thesis submitted to the University of Hull in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Sponsoring Establishment:

University of Hull
Cottingham Road,
Hull
HU6 7RS

Table of Contents

Abstract
Abbreviations

1 Introduction page

1.1 Gas and Vapour Analytical Techniques 1
1.1.1 Introduction 1
1.1.2 Gas Chromatography 2
1.1.3 Gas Array Detectors 2
1.1.3.1 Catalytic Filaments and Pellistors 3
1.1.3.2 Metal Oxide and Semiconductors 4
1.1.3.3 Electrochemical Sensors 5
1.1.3.4 Piezoelectric and Surface Acoustic Wave Detectors 5
1.1.3.5 Electronic Noses 6
1.1.4 Spectrometry 6
1.1.4.1 Photoacoustic Spectroscopy 7
1.1.4.2 Absorption Spectrometry 8
1.1.4.2.1 General IR and Raman Applications 9
1.1.4.2.2 Laser Spectroscopy and Light Emitting Diodes 10
1.1.4.2.3 FTIR, Wavelengths and Filters 11
1.1.4.2.4 Photodetectors 13
1.1.5 Fibre Optics

1.2 Sample Preparation Techniques 15
1.2.1 Static Methods of Sample Preparation 15
1.2.2 Dynamic Methods of Sample Preparation 18

1.3 Chemometric Techniques Employed In Vapour and Gas Analysis 20
1.3.1 Experimentsl Design 20
1.3.2 Approaches to Instrumental Variations and Background Drifts 21
1.3.3 Data Pre-treatment 22
1.3.4 Pattern Recognition, Trends and Identification 22
1.3.4.1 Principal Component Analysis 23
1.3.4.2 Cluster Analysis and Dendrograms 26
1.3.4.3 ANOVA 27
1.3.5 Calibration Models 27
1.3.5.1 Multivariate Linear Regression 27
1.3.5.2 Principal Component Regression 27
1.3.5.3 Partial Least Squares 29
1.3.4 Variable Selection 31

1.4 Aims of the Research 33

2 Experimental Section 36

2.1 Data Treatment and Chemometric Analysis 36

2.1.1 File Formats 36
2.1.2 Software Employed for Analysis 37
2.1.3 Data Pre-treatments 37
2.1.4 Methods and Programs Employed to Analyse Data 38

2.2 Collection of IR Spectra for Multivariate Models 43

2.2.1 Single Component Samples 43
2.2.2 Collection of Samples with Reference Peaks 44
2.2.3 Mixtures of Four Gases plus SF_6 as Internal Standard 45

2.3 Multivariate Modelling with IR Spectra 48

2.3.1 Optimisation of the VS-MLR Modelling Parameters 48
2.3.2 PLS Modelling 48
2.3.3 Conversion of Wavenumbers to Wavelengths 49

2.4 Development of Dilution System 50

2.4.1 General Schematic 50
2.4.2 LabVIEW™ Programming 51
2.4.3 Optimising Gas Chromatograph Conditions and Recording of Data 54
2.4.4 FTIR 55
2.4.5 Different Dilution Set-ups 55
2.4.6 Correlation between GC and IR Measurements 57

2.5 Characterisation of Narrowband Filters by FTIR 58

2.5.1 Characterisation of 3.25 and 4.35 μm Filters 58
2.5.2 Using Filters 2.9, 3.25, 4.35, 4.73, and 10.62 μm 58

2.6 Investigations into Optical Layouts and Signal Processing 62

2.7 The Purgable Box 68

2.7.1 Design and Development of a Purgable Box 68
2.7.2 Effect of Purging 69
2.7.3 Testing the Box 69
2.7.4 Reproducibility 70
2.7.5 Five level Experimental Design Calibrations 70
2.7.6 Investigating the Reproducibility of CH_4 Responses 72
2.7.7 Repeat of Five Level Calibrations with Higher CH_4 Concentrations 72
2.7.8 Sample Preparation using a Sample Loop 72

2.8 Volatile Organic Compound Analysis 74

2.8.1 Samples Collected by Vapour Sampling 74
2.8.2 Improvement of Spectra by Removal of Water Vapour 76
2.8.3 Vapour Samples Prepared by Evaporation 76
2.8.4 Preparation of Mixtures of Heptane, Cyclohexane, Ethanol and O-xylene 78

2.8.5 Investigations into Different Sample Preparation Techniques 79
2.8.6 PLS Investigstions 80

3 Results and Discussion 81

3.1 Gas sample spectra and the effect of pre-treatments on data 81

3.1.1 Comparison of data treatments on spectra for single gas samples 81

3.1.1.1 Spectra collected 81
3.1.1.2 Savitsky-Golay Smoothing 83
3.1.1.3 PCA reconstruction of data 88
3.1.1.4 Work using reference peaks 94
3.1.1.5 Conclusion 99

3.1.2 Mixtures of the four gases plus SF_6 as an internal standard 100

3.1.2.1 Appearance of spectra and the effect of division by reference peak 100
3.1.2.2 Score plots 103
3.1.2.3 Analysis of independent validation data 103
3.1.2.4 Effect of resolution 104
3.1.2.5 Reproducibility of spectra 106
3.1.2.6 Conclusion 108

3.2 Multivariate modelling with IR spectra 110

3.2.1 VS-MLR 110

3.2.1.1 Optimisation of the variable selection process 110
3.2.1.2 Errors in prediction for different wavenumber sets 114
3.2.1.3 Prediction of new validation data 117
3.2.1.4 Conclusion 122

3.2.2 PLS models 122

3.2.2.1 PLS models employing entire spectum 122
3.2.2.2 PLS models for selected wavenumbers 125
3.2.2.3 Conclusion 126

3.2.3 Conversion of wavenumber measurements to wavelengths 127

3.2.4 Conclusion 130

3.3 Dilution system 132

3.3.1 Trial run of dilution system 132

3.3.2 Optimisation of GC conditions 135

3.3.3 Optimisation of GC data recorded in LabVIEW™ 137

3.3.4 Reproducibility of injections 139

3.3.5 Dilution using 1 % mixture with set-up 2 141

3.3.6 The effect of bleeding SF_6 into the system 144

3.3.7 Correlation between GC and IR measurements 144

3.3.8 Conclusion 147

3.4 Narrow band filters with FTIR 148

3.4.1 Characterisation of 3.25 μm and 4.35 μm filters 148

3.4.2 Characterisation of 2.9, 3.25, 4.35, 4.73, and 10.62 μm filters 151

3.4.3 Conclusion 156

3.5 Filter work independent of FTIR 157
3.5.1 General set-up for IR detectors 157
3.5.2 Different detectors 157
3.5.3 Different set-ups 159
3.5.4 Optimisation of the lock-in amplifier settings 163
3.5.5 Signal drift 164
3.5.6 Responses obtained for dilution runs 165
3.5.7 Conclusion 167

3.6 The purgeable box 169
3.6.1 Replacement of MCT detector with the DTGS detector 169
3.6.2 Effect of purging 169
3.6.3 Testing the box 171
3.6.3.1 Static samples 171
3.6.3.2 Dilution samples 172
3.6.4 Five level experimental design calibrations 181
3.6.5 Conclusion 188

3.7 Results and Discussion for Vapour Analysis 191
3.7.1 Spectra collected by sampling vapour with syringe 191
3.7.1.1 Principal Component Analysis (PCA) 191
3.7.1.2 Mean centring pre-treatment of data 197
3.7.1.3 Standard Normal Variate pre-treatment of data 198
3.7.1.4 Cluster analysis 199
3.7.1.5 Conclusion 201
3.7.2 Spectral Correction 201
3.7.2.1 PCA on corrected spectra 203
3.7.2.2 Cluster analysis 205
3.7.2.3 Conclusion 206
3.7.3 Vapour samples prepared from injection of liquid samples 207
3.7.3.1 Calibrations for individual compounds 207
3.7.3.2 PCA on all the spectra 207
3.7.3.3 Reduction of the numbers of spectra analysed 208
3.7.3.4 Analysis of alcohols 209
3.7.3.5 Analysis of mixtures 212
3.7.3.6 Analysis of xylene spectra 214
3.7.3.7 Analysis of acetone, propanoic and tert-butyl alcohol 215
3.7.3.8 Analysis of alkanes 217
3.7.3.9 Analysis of cyclohexane, ethanol, heptane and xylene 219
3.7.3.10 Conclusion 220
3.7.4 Analysis of mixtures of hexane, cyclohexane, ethanol and o-xylene 221
3.7.4.1 PLS analysis 221
3.7.4.2 Investigation into the effect of number of latent variables 223
3.7.4.3 Comparison of different data treatments and models 226
3.7.4.4 VS-MLR 228
3.7.4.5 The effect of preparing mixtures in liquid form before injections 230
3.7.4.6 Conclusion 231

4 Final conclusions and Future Work 232
4.1 Conclusions 232
4.2 Future Work 238
4.3 Comment about the Application of Chemometrics 239

5 References 241

6 Appendices
6.1 Plot of wavenumbers versus wavelengths i
6.2 Reference Spectra ii
6.3 Wiring for MCT Detector iv
6.4 Characteristic IR Absorptions v
6.5 Vapour Spectra vi
6.6 Calibration Plots for Vapour Samples Prepared by Injecting µl Amounts into Gas Cell vii
6.7 LabVIEW™ viii

Abstract

There are many applications for gas and vapour measuring systems in environmental and industrial analysis. Different analytical techniques and detectors employed include; chromatography, catalytic filaments and pellistors, metal oxide and semiconductors, electrochemical sensors, piezoelectric and surface acoustic wave detectors, and spectrometry. The technique employed in this work is infrared spectrometry and it is part of a joint project between the Analytical Science and Optoelectronic Groups at the University of Hull. The aim of the project is to develop a robust multi-component infrared gas analyser using a fixed array of quantum well intersubband photodetectors QWIPs in which the photodetectors are tailored to measure at specific wavelengths.

In order to provide a low cost portable detector, only a limited number of detectors can be incorporated into the array, therefore a method of selecting suitable wavelengths to measure is needed. To achieve this, FTIR spectra were collected for mixtures of CH_4, CO, CO_2 and N_2O. A variable selection algorithm based on multivariate linear regression (VS-MLR) was applied to the spectra to reduce the number of measurements required down to five. It was found that it was possible to build prediction models using low resolution and that baseline variations could be compensated for by using an internal standard. Based on the wavenumber selection models, the concentrations of CH_4, CO, CO_2 and N_2O in new validation data could be predicted using wavenumbers 2129, 2284, 3071, 3148 and 3488 cm^{-1}, with prediction errors averaging less than 14 %. This gave comparable results to PLS models built from the entire spectra.

An exponential dilution system was constructed to enable different concentrations of a gas standard to be prepared and a GC was included into the system to act as a reference method for IR measurements. Good correlation was achieved between IR and GC responses for CH_4, CO_2 and N_2O with R^2 values greater than 0.988.

A prototype system was built (which incorporated the dilution system), utilising narrowband filters (at wavelengths 2.9, 3.25, 4.35, 4.73 and 10.62 μm) combined with general IR detectors, based on the results of the wavelength selection. This was shown to give promising results and future work will be to replace the IR detectors with QWIPs.

Also explored in this work was the feasibility of developing a system to measure volatile organic compounds (VOCs). This was investigated by collecting FTIR spectra for a variety of VOCs and employing PCA and cluster analysis to see if it was possible

to distinguish between different types of compounds. Partial least squares (PLS) and VS-MLR was performed on a simple four component vapour mixture to demonstrate that the same methods employed on the CH_4, CO, CO_2 and N_2O mix could also be applied to other mixtures.

In conclusion, the project demonstrated the role chemometrics can play in the development of a solid state gas analyser, from experimental design to data treatment and analysis. It was proved the VS-MLR could be successfully utilised to determine suitable wavelengths to measure in the array for gas and vapour applications and a prototype system was constructed using narrow band filters to demonstrate the operation of such a system.

Abbreviations

cl	convergence level
DTGS	Deuterated Trigylcine Sulphate
FTIR	Fourier Transform Infra-red
GC	Gas Chromatography
IR	Infra-red
LED	Light Emitting Diode
maxvars	maximum number of variables
MCT	Mercury Cadmium Telluride
MS	Mass Spectrometry
PC	principal component
PCA	Principal Component Analysis
PLS	Partial Least Squares
ppb	parts per billion (N.B. for gases this is by volume)
ppm	parts per million
PRESS	Predicted Residual Error Sum of Squares
QWIP	Quantum Well Infrared Photodetector
RSD	Relative Standard Deviation
SEP	Standard Error in Prediction
VS-MLR	Variable Selection employing Multivariate Linear Regression

1 Introduction

1.1 Gas and Vapour Analysis: Analytical techniques

1.1.1 Introduction

Gas and vapour analysis has importance in a number of different situations, including atmospheric sensing, monitoring of stacks, vehicle emissions, landfill sites and process control. Systems need to be capable of measuring the levels and concentrations of the components of interest, or in some cases identify unknown contaminants. Where samples are analysed in a laboratory, the size and mobility of the instrumentation is unimportant and equipment such as Mass Spectrometers, FTIR Instruments, and Gas Chromatographs can be utilised. [1,2]

Rather than collecting samples and returning them to a lab for analysis, which can result in significant sampling errors, it is more convenient to be able to take in-field or on-line measurements. For these situations, smaller, robust, low cost, more portable monitoring devices are required. The restraints on portable devices i.e. size limitations and physical robustness, result in higher limits of detection compared to laboratory based instruments; however, their advantage lies in the ability to make on the spot measurements thus reducing sampling error as well as allowing an increased number of samples to be measured.

The next section is a brief summary of the main techniques of analysis, focusing on portable devices, with examples of some the current research in these areas.

1.1.2 Gas Chromatography

Chromatography is one of the most common methods of analysis. Advantages include the ability to separate gas mixtures thus allowing components to be identified by their retention times. Columns can be tailored using different lengths and packing materials to obtain improved separation for gases and vapours of interest. The equipment tends to be bulky; a main limitation to portability is the requirement of a supply of carrier gas.

The development of portable gas chromatographs has been slow due to the problems with column efficiency, quality, cost and convenience. Compared to laboratory-based equipment, the resolution of portable Gas Chromatographs (GCs) can be poor due to the limitations in the size of columns. Commercial portable GCs are available, the size of which is decreasing with advancing technology. One of the smallest gas chromatographs being developed is the size of a telephone developed by the Lawrence Livermore Laboratory[3] and uses a

microcapillary on a silicon wafer with the target aim of a resolution of 1 ppb and an identification time within two minutes.

A drawback with decreasing column size is the sample volume also has to be decreased, which leads to higher detection limits. Vapour and contaminant levels in air are quite dilute, often with levels less than a few ppb, so sampling and pre-concentration are important considerations.

Another limitation of chromatography is that it is not continuous; the interval between sampling is dependent on the time taken to run samples, which in turn, is dependent on retention times. A practical example of on site equipment is an automated field instrument for the determination of acidic gases (HCl, $HONO^{-3}$, HNO_3 and SO_2,) in air, giving detection limits of $\mu g/m^3$, but with the analysis time taking 20 minutes.[4] For the measurement of volatile organic vapours, portable GC's can be combined with solid phase microextraction (SPME) techniques to increase concentration levels.[5] Separation can be achieved in less than five minutes, however this is far slower than other analytical methods such as Infrared and Raman Spectroscopy, where analysis can be performed within less than a second.

1.1.3 Gas Arrays Detectors

Many of the current gas sensors are based on arrays of various catalytic, semi-conducting, electrochemical, piezoelectric and optical type detectors.[6] These types of detectors tend to respond to numerous different gases, so a single detector on its own gives very limited information. By creating an array of detectors, where each detector has a different response to different analytes, unique response patterns can be obtained for various analytes. Pattern recognition techniques such as neural networks and principal component analysis, and other multivariate techniques can be applied to derive useful information from array responses; these are discussed later in section 1.3.

1.1.3.1 Catalytic Filaments and Pellistors

For personal monitoring devices, a common type of detector is the catalytic filament. The advantage of these detectors is that they are small and easily incorporated into portable devices that can be worn in pockets or on belts.

Catalytic filaments are particularly useful for monitoring flammable vapours; an example is the platinum resistance thermometer.[7] Combustion of flammable vapours is catalysed by a heated filament, causing the filament temperature and hence its resistance to change. The increase in resistance is proportional to vapour concentration, so measuring the cell

resistance can be used as a means to determine vapour concentrations. Problems can arise due to catalyst inhibitors and poisoning, for example from silicone (found in grease and polishes), which forms silicon oxides on the catalyst surface, or inhibition due to absorption of halogens from compounds containing halides. One method to overcome the effects of poisoning is the use of coatings and capsules. Zeolite structures can be utilised as ion exchange materials and the small pore sizes can prevent larger poison molecules such as silicates from reaching the catalyst surface.[6] Pellistors are well known catalytic detectors and consist of a platinum wire coil contained within an alumina bead treated with catalyst. A major drawback with these devices is the lack of selectivity, responses arise from a number of flammable gases and distinguishing between them can be difficult, if not impossible. The signal response can vary widely depending on the molecular weight, functional groups and lower explosive limits of the vapours present, and this variation is used to build arrays. For CO monitoring, the Hopcalite catalyst, which gives very little response to other vapours at low temperatures, can be employed.[6]

Catalytic devices require oxygen for combustion to occur thus can fail in oxygen deficient situations, also as detection is based on the combustion of gases, gases in the immediate vicinity of the detectors are depleted hence good air circulation is required.

1.1.3.2 Metal Oxide and Semiconductors

Metal oxide detectors consist of a ceramic tube or plate coated with a layer of metal oxide. The response arises due to oxygen in the air reacting with vacancies in the metal oxide lattice. Changes in the composition and materials used in the detectors cause different sensitivities to different compounds. As with filament detectors, these sensors also suffer from problems with poisoning, and poor selectivity. Dampness can also lead to problems and increasing amounts of water vapour can lead to instability and decrease signal detection. To overcome cross-sensitivity to water, higher operating temperatures are required. Metal oxide detectors use porous alumina beads to improve stability and resistance to poisoning.

Some of the recent work includes combinations of palladium and platinum metal oxide semiconductor field effect transistors (MOSFET) sensors and palladium and platinum catalysts. [8] A well-known type of detector is the Taguchi sensor, which is based on a ceramic SnO_2 bead. Simultaneous determinations of ethanol, toluene, and o-xylene can be made, employing commercially available Taguchi sensors, with detections in the 20 to 100 ppm concentration range.[9] An array of six Taguchi sensors combined with a neural network system was applied for the quantification of organic vapours (butanol, toluene and xylene).[10]

A large amount of research has been investigating the effects of composition and surfaces of such sensors, the addition of metals such as Bi and Pd to act as dopants or blending different oxides or including metal porphyrins can all decrease the operating temperatures required. Devices can also be coated with catalytic layers to improve selectivity.[10] Work by Park *et al.* employed SnO_2 doped with Pt and Nb_2O_5 to tailor detectors that were sensitive to alcohol gas but with negligible responses to ethyl alcohol and cigarette smoke.[11]

Other detectors employ conducting polymers such as polyacetate, polpyrrole, phthalocyanines and polysiloxanes. A film of conducting polymer is used to coat detectors, and different responses arise due to the adsorption of gases and vapour, which cause the conductivity to change. Such detectors can have good sensitivity to particular analytes; however, they tend to have high temperature dependence and a limited lifetime. Work with semi-conducting polymers includes the development of sensor systems for alcohol and chlorinated hydrocarbon vapours utilising poly-2,5-furylene-vinylene (PFV) derivatives.[12] Fabrication techniques such as sputtering and chemical vapour deposition make it possible to design microsystems consisting of multiple sensor elements. An example of this is work by Althainz *et al.*, where a system consisting of 39 metal oxide detectors, 2 temperature detectors and 2 heaters in area of 8 x 9 mm^2 was produced. Different responses for various vapours were achieved by applying a temperature gradient across the array.[13]

1.1.3.3 Electrochemical Sensors

Electrochemical sensors are common, especially for monitoring oxygen and the use of microelectrodes allow small compact sensors to be built. Different developments in this area include: chemically modified electrode sensors, microelectrodes, voltametric and potentiometric methods (such as stripping voltametry). Electrodes are tailored for specific analytes by the use of membranes or polymer coatings allowing even anions that are not electrochemically oxidizable or reducible to be analysed. Using a Nafion(R) cation exchange membrane amperometric gas sensors can be designed to detect gases such as sulphur dioxide, ethylene, ethanol, acetylaldehyde and acetylene in the ppb range.[14]

Microelectrode arrays can be fabricated using photolithology employing various materials and modification techniques to make individual elements more selective, making it possible to construct hand held electronic systems.[1] Zirconia based sensors for potentiometric and amperometric measurements can be utilised for measuring hydrocarbons, nitrogen oxides and oxygen.[15]

A drawback of electrochemical methods can be the lack of selectivity that often requires a large amount of data processing and modelling, and that they can be adversely affected by humidity.

1.1.3.4 Piezoelectric and Surface Acoustic Wave detectors

Piezoelectric detectors consist of oscillating crystals. The absorption of compounds on the crystal surface causes a change in vibration frequencies, allowing the detectors to act as chemical mass balances. Like the previously mentioned detectors they are non-selective, but can be made more selective by coating with selective absorbates such as plasma polymer films. These can be employed as moisture detectors with sensitivities of 0.1 ppm. Another example of their application is for CO monitoring with detection limits down to 10 ppm.[16] An example is the multi-component gas analysis of a mixture of chloroform, octane, and toluene using a piezoelectric quartz crystal sensor array, where an array of four crystals was designed to calibrate vapour mixtures between 100 and 2000 ppm, predictions errors varied from +28 % to -8 %.[17] Other applications have included classification of organic compounds into groups, odour analysis and the identification of anaesthetics.

Surface acoustic wave sensors (SAW) detect changes in mass of surface coated films. Changing the polymer used can result in different responses to different vapours thus allowing characteristic response patterns than can be used to identify different compounds. Flexural plate waves (FPW) sensors are similar to SAW sensors; both methods are based on the employment of acoustic waves generated from within a piezoelectric substrate. In FPW the waves travel in a membrane that has a thickness smaller than the length of the acoustic wave and as a result the membrane undergoes mechanical flexure.[18]

An example of this is work by Groves *et al.*[19] who developed an instrument based on an array of SAW sensors to analyse organic vapours present in exhaled breath and in ambient air.

1.1.3.5 Electronic Noses

Electronic noses are an attempt to imitate the olfactory function of the human nose. Rather than identify each individual compound and its concentration, the overall effect is looked at. A definition given by Gardner and Bartlett is: "An electronic nose is an instrument, which comprises an array of electronic chemical sensors with partial specificity and appropriate pattern-recognition system, capable of recognising simple or complex odours".[20]

Applications of this include odour analysis, where a particular smell is not due to one particular compound but can arise from a complex mixture of many different chemicals for example perfumes, cheese, coffee, etc. Other situations where they are useful are

environmental odour problems such as livestock waters, pig slurries, tainted water supplies and sewage treatment works.[21] Electronic noses can also be applied to gas and vapour detection.

Electronic noses based on the aforementioned detectors and also on optical detectors (discussed later in section 1.1.4) have been developed for numerous applications, so any advantages or disadvantages are those associated with the type of detectors employed.

Work by Natale *et al.*[22] looked at the application of an electronic nose based on an array of shear mode quartz resonators to analyse peaches. The resonators were coated with thin layers of metal complexes of porphyrins, this included cobalt, manganese, rhodium, tin, vanadium, ruthenium and copper, allowing different sensitivities to aromatics, alcohols and aldehydes.

Another example includes work by Maekawa *et al.* utilising an array of SnO_2 sensors for the identification of odours from citreous fruit and different types of coffee beans[23].

Electronic noses developed to identify different organic compounds include an array based on SAW detectors for odour identification of ethanol, ether, acetone and ethylacetate developed by Yang *et al.*[24], and an array of conducting polymer sensors based on polypyrrole polymers to identify and quantify benzene, toluene, ethyl benzene and xylene in work by Barisci *et al.*[25]

1.1.4 Spectrometry

1.1.4.1 Photoacoustic Spectroscopy

Photoacoustic spectroscopy is based on optically exciting samples with radiation and detecting the energy decay as an acoustic wave.

The sensitivity and detection is dependent on the power of the radiation source.

High sensitivity can be achieved to detect absorption coefficients down to 10^{-8} cm^{-1} by modulating the radiation source at an acoustic resonance frequency of the absorption cell.

CO_2 lasers can be used that have a wavelength range of 9 to 11 μm, which is advantageous as many characteristic absorption bands of gases occur in this region.

Acoustic devices can be made from optically transparent flat materials laid in parallel with each other to form an array of Fabry-Perot optical interferromatic cavities. As one tyne vibrates the light transmittance oscillates, and this oscillation can be employed to determine the strength of the resonance. The tynes represent certain frequency components and this can be related to different absorbing species. Problems can arise due to varying water vapour and CO_2 levels which can affect the amplitude of the detected acoustic signal.

Work in this area includes the development of a mobile CO_2 laser photoacoustic spectrometer and its potential for analysis of fruit storage chambers and ambient air measurements in field studies.[26] By applying least square fitting algorithms such as a Levenberg-Marquardt fitting algorithm to the resulting signal it is possible to analyse different gas mixtures in the ppb to ppm range. Lasers are usually employed due to the output signal being proportional to the energy of the source. External cavity diode lasers (ECDL) can also be used as excitation sources, the lack of sensitivity can be compensated for by tuning the lasers to the peak absorptions, as demonstrated by a system using ECDLs capable of detecting NH_3 down to 8 ng/l.[27]

Gas lasers have the disadvantage of being relatively large due to the requirements of gas flow, water or liquid nitrogen cooling, sophisticated electronics and power stabilisation, thus making them less suitable for field instruments. Diode lasers, which are more compact, can also be utilised as radiation sources, although they have less power and therefore are less sensitive they can be used to detect gases in the low ppm range.[28]

1.1.4.2 Absorption Spectrometry

Absorption spectrometry is based on measuring the amount of light of a specific wavelength that is absorbed by the analyte(s) to be measured. Infrared spectrometry concerns the range of electromagnetic radiation with wavelengths from 0.8 μm in the near IR region to 1000 μm in the far IR region. Mainly measurements are made in the mid IR region, which has a wavelength range from 2.5 to 25 μm. Absorption of IR light is due to the interaction of light with various stretches, bends and vibration movements of molecular bonds. When the vibration causes a change in the dipole moment of the molecule, the molecule can be excited by IR radiation. When the wavelength of light (and hence its energy) matches the vibration energy it is absorbed. Non-polar molecules or bonds such as N_2 and O_2 do not have a change in dipole moment so are not IR active, whereas asymmetric and ionic molecules (or group contained within a molecule) like H_2O and CO_2 exhibit strong IR absorptions. Particular molecular bonds have characteristic absorptions (or bands) at specific frequencies making it possible to identify some compounds by the characteristic spectra produced.[29]

Absorption energies can be explained using Hooke's law, which is based on a simple harmonic oscillator. A common analogy is the comparison of the molecular bond to a spring. The potential energy is at its lowest when the bond is in equilibrium, and energy increases when the bond expands or contracts in the same way that pulling or pushing a spring increases its recoil energy.

Raman spectrometry also employs infrared radiation; however, it is based on measuring light from inelastic scattering.

Advantages with spectrometry are that continuous measurements can be made, it is a non-destructive technique, and that it can be employed for remote sensing.

1.1.4.2.1 General IR and Raman Applications

Both IR and Raman spectroscopy have been successfully employed to analyse various gas species simultaneously, as shown by studies measuring air quality and pollution levels.[30] Examples include the simultaneous analysis of CH_4, CO_2, N_2O and CO in air by FTIR using a classical least squares algorithm for calibration.[31] Although high resolution is normally used for gas analysis, it is also possible to analyse gas mixtures with low resolution as demonstrated in the analysis of stack gas pollutants.[32]

The major problem in the application of Raman spectroscopy is the signal-to-noise ratio, which is related to the extremely low scattering cross section. It is possible to utilise surface enhanced Raman scattering (SERS) for the analysis of vapours using metals with organic coatings. Trace organic compounds can identified by the location of the bands. SERS can be limited by chemical reactivity and contamination by adsorption; if contaminants and the analyte of interest are adsorbed by the same host molecules, then competition for adsorption places can affect the results giving lower values then actually present.[33]

Spectrometers basically consist of a radiation source, some means of wavelength selection and a detector. Developments in absorption spectrometry can be broadly split into three areas based on which component is focused on.

a) Laser spectrometry and the use of light emitting diodes focus on refining the radiation source. Wavelength selection is achieved due to the emission of very narrow wavebands of radiation. Since the wavelength of light measured is determined by the source, simple non-selective light detectors can be employed. Lasers and LEDs are further discussed in section 1.1.4.2.2.

b) Fourier transform (FT) IR, and wavelength modulation techniques employ general sources and detectors, and rely on various methods to select wavelengths. In FTIR, interferometers are used to produce interferrograms, which can be resolved mathematically using Fourier transform. Alternatively, diffractive grating and filters can be utilised to separate wavelengths or restrict the light passing through the sample or reaching the detectors. Work in this area is described in section 1.1.4.2.3.

c) The third alternative is to employ detectors that only measure specific wavelengths. Work in this area includes quantum well photodetectors (QWIPs), which are discussed in section 1.1.4.2.4.

1.1.4.2.2 Laser Spectroscopy and Light Emitting Diodes

The main advantage of laser absorption spectroscopy (LAS) is its selectivity (lead salt lasers have a linewidth of 10 MHz ($3x10^{-4}$ cm^{-1}) that makes it an ideal technique for looking at small specific regions of spectra where absorption lines are close together. Speed of analysis depends on the range; several wavelengths can be scanned within milliseconds, whereas scanning an entire spectrum greatly increases analysis time.[34] The use of a high-power coherent source and the lack of dispersive gratings mean that greater sensitivity and higher signal-to-noise ratio can be achieved compared to spectrometric techniques that employ conventional sources. This allows the detection of low concentrations (ppb) and weak spectral features can be analysed. The wavelength at which light is emitted relies on the band transmissions of the bulk material.

Lead salt lasers are good for mid-IR applications but the drawback is that they normally require cryogenic cooling. Alternatives include antimonide lasers and quantum cascade lasers.[35] An example of a portable gas analyser using a tuneable lead salt diode laser is the development of TRISTAR (a tracer in-situ TDLAS (tuneable diode laser atomic spectroscopy) for atmospheric research), an airborne platform consisting of three lasers, two detectors, a cryostat and a multi-path cell for the analysis of N_2O, CO and CH_4. Reproducibility of calibrations was within 2 % and accuracy within 2.8 % with detection levels down to ppb (noise levels at <10 ppbv, <20 ppbv, and 30 ppbv for N_2O, CO, and CH_4 respectively).[36] Along similar lines, is work by G.Toci *et al.*, designing an airborne diode laser spectrometer for measuring H_2O and HNO_3.[37]

Room temperature lasers using bulk $LiNbO_3$ with Nd:YAG can be employed giving detections in ppb range for gases CH_4, C_2H_6, H_2CO, NO_2, N_2O, HCl, CO and OCS, with on-line measurements between 200ppb and 1% for methane.[38]

Semiconductor lasers are normally monochromatic, but it is possible to design lasers that emit several simultaneous wavelengths by employing separated multiple resonators or multi-section injection devices. In effect this produces a cross between a superlattice with continuous energy minibands and multiple quantum wells, resulting in a multi-wavelength semiconductor laser.[39]

The use of light emitting diodes (LEDs) for gas analysis has been ongoing for several years and has been growing in popularity. LEDs can increase the portability of sensors due to the small size and lack of moving parts.

LEDs can be realised from a number of different materials, electroluminescent organic polymers such as binaphthyl-based polymers are used for LEDs in the visible light region. For infrared LEDs, various materials and structures have been based on InGaAs, InAsSbP/InAs and other InAs alloys. These materials are utilised to grow quantum well structures with the energy difference between excited and ground states corresponding to characteristic infrared absorption bands of gases. Quantum wells can be grown by a number of different techniques, among them are; LPE (liquid phase epitaxy), MBE (molecular beam epitaxy), and MOVPE (metalorganic vapour phase epitaxy). These techniques allow for a high degree of control over the thickness of layers even down to monolayers. Work by Krier *et al.* employed rapid slider LPE for the growth of InAs material systems. The thickness of layers was controlled using a graphite boat with an adjustable slit. Using LPE, layers as thin as 20 A can be grown, and produces material with high crystalline perfection with few point defects and impurities. Work comparing different types of LEDs found InAsSbP/InAs double heterojunctions to be the most efficient structures with emissions between 3.3 and 3.7 μm.[40,41] LEDs can increase the portability of sensors due to the small size and lack of moving parts. A drawback with LEDs is the limited optical output power due to effects from Auger recombination.

An EPSRC Link Photonics project known as ASPIRE (atmospheric sensing of pollutants by infrared emitters) was set up to examine LED performance shows that there are clear advantages for using lasers and LEDs, but that problems with power output, temperature stability and emitter lifetime still need to be overcome.[42]

1.1.4.2.3 FTIR, Wavelength Modulation (derivative, or frequency modulation (FM)) and filters

FTIR instruments are the most commonly used type of instrumentation for infrared analysis. They employ interferometers so that an interferrogram signal is produced, which can be processed using Fourier transformation. The instrumentation tends to be bulky and requires complex signal processing, but smaller portable instruments do exist. Portable FTIR gas analysers that are built for field measurements are commercially available from companies such as GASMET™ that can simultaneously analyse up to 30 different components with an analysis time in less than 180 seconds.[43]

Wavelength selection can also be achieved using diffraction gratings and filters. Since only common sources and detectors are required these methods have the advantages of requiring little maintenance and low power supply. The use of optical filters can be an advantage for regions where no low cost laser sources are available as they allow LEDs or thermal sources to be used. Filters can either be placed to filter the radiation source, or to filter the light reaching the detectors.

Methods include, using a tuneable Fabry-Perot interferometer and vibrating mirror inserted into a grating monochromator. Although the wavelength selectivity for this method is not as high as the employment of laser diodes due to effects such as line shape distortions, it does have the advantage of being able to cover a wide range of wavelengths from UV to far infrared.

Immersion gratings from silicon and germanium can be utilised, the high refraction of these materials allows the design of compact spectrometers with no moving parts.

It is possible to design hand held sensors using a thin vibrating diffraction grating fabricated on a silicon chip as demonstrated by Chen in the development of a CO_2 and hydrocarbon detector capable of detecting levels down to ppm.[44] A drawback is that sensitivity is affected by temperature changes of the radiation sources which influence the signal detected.

Filters can be integrated to produce a filter array, which can be coupled to the detectors as demonstrated in work by Lopez *et al.* A filter array was coupled to PbSe detectors to produce a spectral retina, forming signal channels allowing the identification and quantification of SO_2 and CO.[45,46,47]

The detection is limited by the detector performance, and coupling filters with more sensitive detectors can increase the sensitivity. Another example is the employment of micromachined optical tuneable filters in work by Alause *et al.*[48]

The approach used by Myrick *et al.* is to employ interference filters referred to as multivariate optical elements (MOEs). Optical interference coatings are produced whose transmission spectra incorporate features corresponding to the target analytes.[49,50,51]

A practical example of portable infrared analysers based on filters is for measurements of hydrocarbon emission levels. The analyser used photoacoustic detectors combined with an optical carousel containing filters at 6 discrete wavelengths, and was found to be reliable for field analysis to monitor hydrocarbon emissions.[52]

1.1.4.2.4. Photodetectors

Photodetectors include PbSe and HgCdTe, the main disadvantage with these materials is that they require cryogenic cooling in order to reduce noise and increase sensitivity. Research

has been carried out to produce photodiodes that can operate at room temperature, thus reducing the power requirements needed for thermoelectric cooling.

Photodiodes are in effect the reverse of LEDs; detection is based on the excitation of electrons to conduction bands by light, compared to LEDs where electrons fall to lower energy levels so light energy is emitted.

Quantum well intersubband photodetectors (QWIPs) can be grown as described for photodiodes.[53] The devices have similar problems with quantum efficiency due to Auger recombination and the effects of dark current (current not due to light detection). At the moment wavelength detection is limited to shorter wavelengths with a cut off around 2.5 μm, but improvements on the design and materials used have reduced the amount of dark current and the response time. Compared to HgCdTe (sometimes referred to as MCT) detectors, the growth and processing technologies give high uniformity and reproducibility which leads to large area low cost arrays. The most basic wells consist of layers of alternating materials such as GaAs and AlGaAs, or Si and SiGe.[54,55] By varying the thickness of layers, the materials used and the amount of doping, the size and shape of the wells can be tailored, thus changing the energy of transitions and hence the wavelength that can be detected.

Asymmetric wells are achieved by the addition of an extra barrier layer that can allow transitions normally forbidden in symmetric wells. Figure 1.1.1 shows an example of a simplified diagram for an asymmetric well.

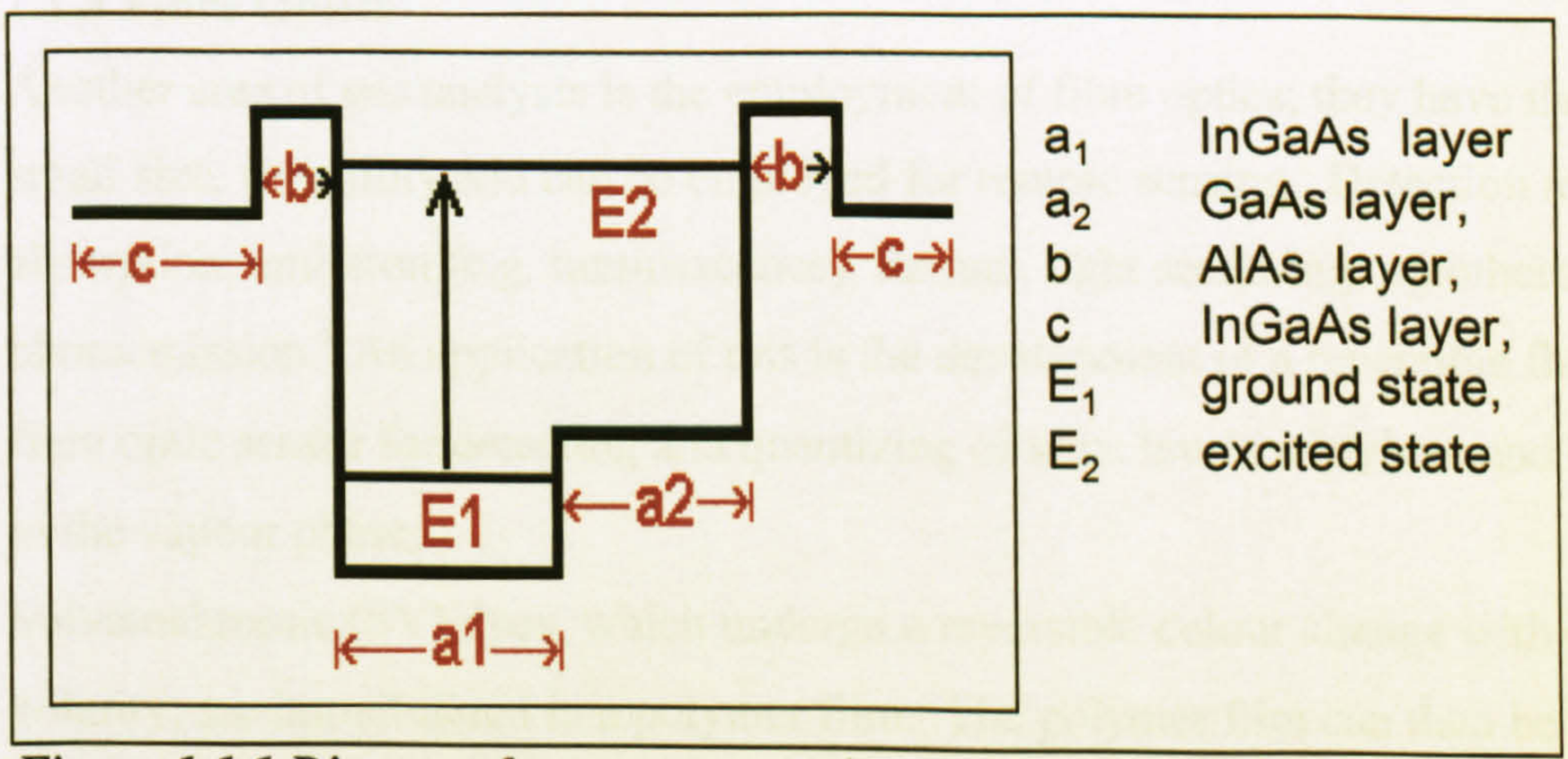


Figure 1.1.1 Diagram for an asymmetric quantum well. N.B. Electrons are excited from E_1 (the ground state) to E_2 when the wavelength of light matches the energy difference between the two levels.

The additional barrier also reduces the amount of dark current, leading to increased sensitivity. Work in this area includes uncooled InAsSbP/InGaAs photodiodes (grown by

liquid phase epitaxy) which can operate in the 1.8- 3.4 μm range, they compare favourably with HgCdTe detectors, and can be employed for methane and HF gas detection.[56]

The detectors being developed at Hull University by the optoelectronic group are asymmetric strained QWIPs based on GaAs and InP.[57,58] Figure 1.1.2 shows an absorption plot for one of the unprocessed samples (NB. The sample is not yet processed into a detector; hence no voltage is applied across the sample and light is absorbed by the sample).

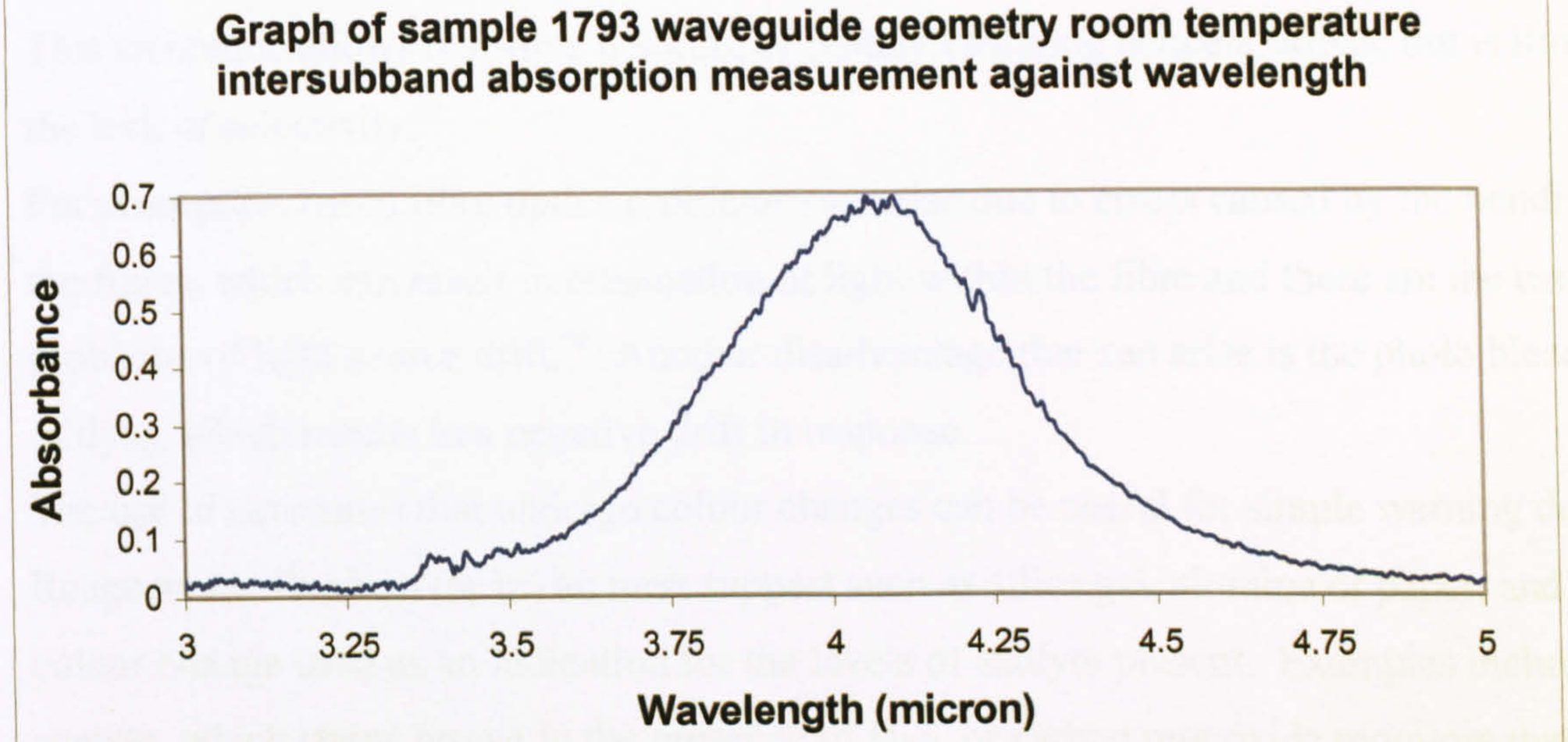


Figure 1.1.2 Absorption plot for unprocessed sample designed by Electronic group at Hull

1.1.5 Fibre Optics

Another area of gas analysis is the employment of fibre optics; they have the advantages of small size, versatility and can be employed for remote sensing. Detection methods include; absorption, emission (e.g. luminescence), Raman, light scattering, optothermal and photoemission.[5] An application of this is the development of a reversible fluorescence-based fibre optic sensor for detecting and quantizing of trace levels of xylene and dichloromethane in the vapour phase.

Solvatochromic (SV) dyes, which undergo a reversible colour change with changes in polarity, are immobilized in a polymer film. The polymer film can then be coated on to the sides or end of an optical fibre allowing concentrations of analytes to be measured by measuring the colour change.[1] Detection using fluorescent compounds is widespread and high sensitivity can be achieved. It is possible to detect polyaromatic hydrocarbons (PAHs) down to ng m^{-3}.[59]

Albert *et al.* demonstrated that fibre optic beads arrays can be employed for the discrimination of odours such as toluene, acetone and dinitrotoluene, as well as more complex odours such as coffee beans.[60]

Oxygen sensors based on luminescence quenching can be applied for monitoring oxygen in gas flows, as well as in situ within biological systems. The probes use inorganic films based on metal complexes and porphyrin systems, when compared to amperometric sensors they are faster, do not consume the analyte and lack electrical connections.

This technique allows real-time tracking of rapidly changing concentrations, but is limited by the lack of selectivity.[61]

For absorption based fibre optics problems can arise due to errors caused by the bending of the fibres, which can result in attenuation of light within the fibre and there are the usual problems of light source drift.[62] Another disadvantage that can arise is the photo bleaching of dyes, which results in a negative drift in response.

The use of substrates that undergo colour changes can be useful for simple warning devices. Reagents are fixed on (or in) an inert support such as silica gel, alumina or paper, and the colour change used as an indication for the levels of analyte present. Examples include lead acetate, which stains brown in the presence of H_2S, or carbon monoxide monitors that change colour when exposure limits are reached.[1]

1.2 Sample Preparation Techniques for Gas and Vapour Standards

It is very important to be able to prepare reliable gas and vapour standards to validate and calibrate gas detection systems. A number of different strategies can be employed to generate gas and vapour samples; these techniques can be split into two areas, static and dynamic. Factors to consider include, time, expense, equipment available and the required accuracy. Table 1.2.1, taken from literature[63], gives a comparison between the most common methods of sample preparation.

Method	Usual range of concentration	Volume of test mixture (L)	Average best accuracy (%)	Operating pressure (atm)	Degree of wall loss
Static systems					
Single vessel	10ppm to 5%	10% of vessel	3-5, 1	≤1	Medium
Multiple vessel	10ppm to 5%	Vessel volume	5-10, 3	≤1	Medium
Non-rigid vessel	10ppm to 5%	1-1000	3, 1	1	High
Gravimetric	Low ppm to 50%	1-5000	0.1-0.002	1-150	Medium
Partial pressure	100ppm to 50%	0.5-5000	5-10, 1	1-150	Medium
Volumetric	Low ppm to 50%	0.5-5000	3-5, 2	1-150	Medium
Partial pressure	Low ppm to 50%	0.1-2	1-5, 1	0.001-1	Low-medium
Dynamic systems: gas stream mixing					
Single dilution	100ppm to 50%	>5000	1-3, 1	≥1	Low
Double dilution	Low ppb to 1%	>5000	5-10, 2	≥1	Low
Dynamic systems: injection					
Syringe drive	1ppm to 0.1%	5-1000	2-5. 1	≥1	Low
Liquid pump	1ppm to 5%	Very large	5, 1	1	Low
Gravity feed	100ppm to 5%	Large	5, 3	≥1	Low
Pulse diluter	0.1ppb to 0.1%	Very large	3-5, 1	≥1	Low-medium
Diffusion	0.1 to 500 ppm	Very large	2-6, 1	1	Low
Permeation	0.05 to 200ppm	Large	1-3, 0.5	1	Low
Evaporation	50 ppm to 10%	Large	3-15, 2	≥1	Low
Electrolysis	Low ppm to 10%	Ltd. by soln. vol.	2-5, 1	1	Low
Chemical reaction	Low ppm to 1%	Ltd. by soln. vol.	3-10, 1	1	Low-medium

Table 1.2.1 Comparison of common methods of sample preparation[63]

1.2.1 Static Methods of Sample Preparation

Static methods involve the addition of a known quantity of solvent or gas into a container of known volume. These methods are quite simple with accuracy depending on the volume of the system, the purity and volume of analyte. The method of addition can be via syringe. Table 1.2.2 gives a comparison of the accuracy of different injection methods.[60]

For gas mixtures, concentrations are calculated from volumes as shown by equation 1.1.

$$C_{ppm} = \frac{V_c}{(V_c + V_d)} \times 10^6 \, ppm \qquad \text{Equation 1.1}$$

Where;

V_c = target gas volume

V_d = dilutant volume

For liquids added to a closed system the concentration in ppm can be calculated from equation 1.2.

$$C = \frac{22.4\times10^6\left(\frac{T}{273^\circ K}\right)\left(\frac{760mm}{P}\right)W}{VM} \qquad \text{Equation 1.2}$$

Where;

C = concentration (ppm),

T = temperature (^oK),

P = system pressure (mmHg),

W = weight (g),

V = volume (l) of the system

M is the molecular weight (g/mol)

At normal temp (25°C and 760 mmHg) this reduces to

$$C = \frac{24.5\times10^6 W}{VM} \qquad \text{Equation 1.3}$$

Where;

V= volume of liquid,

W = weight (g),

V_a = volume of dilutant,

M = the molecular weight of the liquid

Injector	Phase injected	Plunger material	Capacity	Accuracy at capacity,%
Microlitre syringe	Liquid	Wire	1-5 µL	2
Microlitre syringe	Liquid	Wire	5-500 µL	1
Gas-tight syringe	Liquid,gas	Teflon tip	0.05-2.5 mL	1
Gas-tight syringe	Liquid, gas	Teflon tip	5-50 mL	1
Micropipette	Liquid	Air	2-4 µL	1.8
Micropipette	Liquid	Air	5-25 µL	0.5
Micropipette	Liquid	Air	35-150 µL	0.3
Microm burette	Liquid	Teflon	0.2-2 mL	0.5
Microm burette	Liquid	Glass	0.25-2.5 mL	0.04
Volumetric pipette	Liquid	Air	0.5-200 mL	1

Table 1.2.2 Accuracy of different injection methods[63]

The most accurate technique of preparation is by gravimetric means. An example of this is the collection of Infrared Spectra for the NIST (National Institute of Standards and Technology) database, where VOC standards were prepared using evacuated, preweighed

cylinders fitted with septums. A known weight of analyte is injected into the cylinder, which is then filled with nitrogen at a calculated pressure and then finally weighed.[64]

In work by Groves and Zellers, breath samples were collected in Tedlar® bags with known volumes of liquid solvents injected into the bag and samples were mixed by massaging the bag.[18]

Another method of preparation is by employing pressure techniques; figure 1.2.1 is an example of a schematic for such a system. The vessel/cell to be filled with sample is evacuated using a vacuum pump. For preparing vapours the liquid sample is cooled (normally with liquid N_2) and the air above the liquid sample is pumped off, then the liquid sample is heated allowing vapour saturation. The valve for vessel/cell is then opened allowing gas/vapour to flow in.[63]

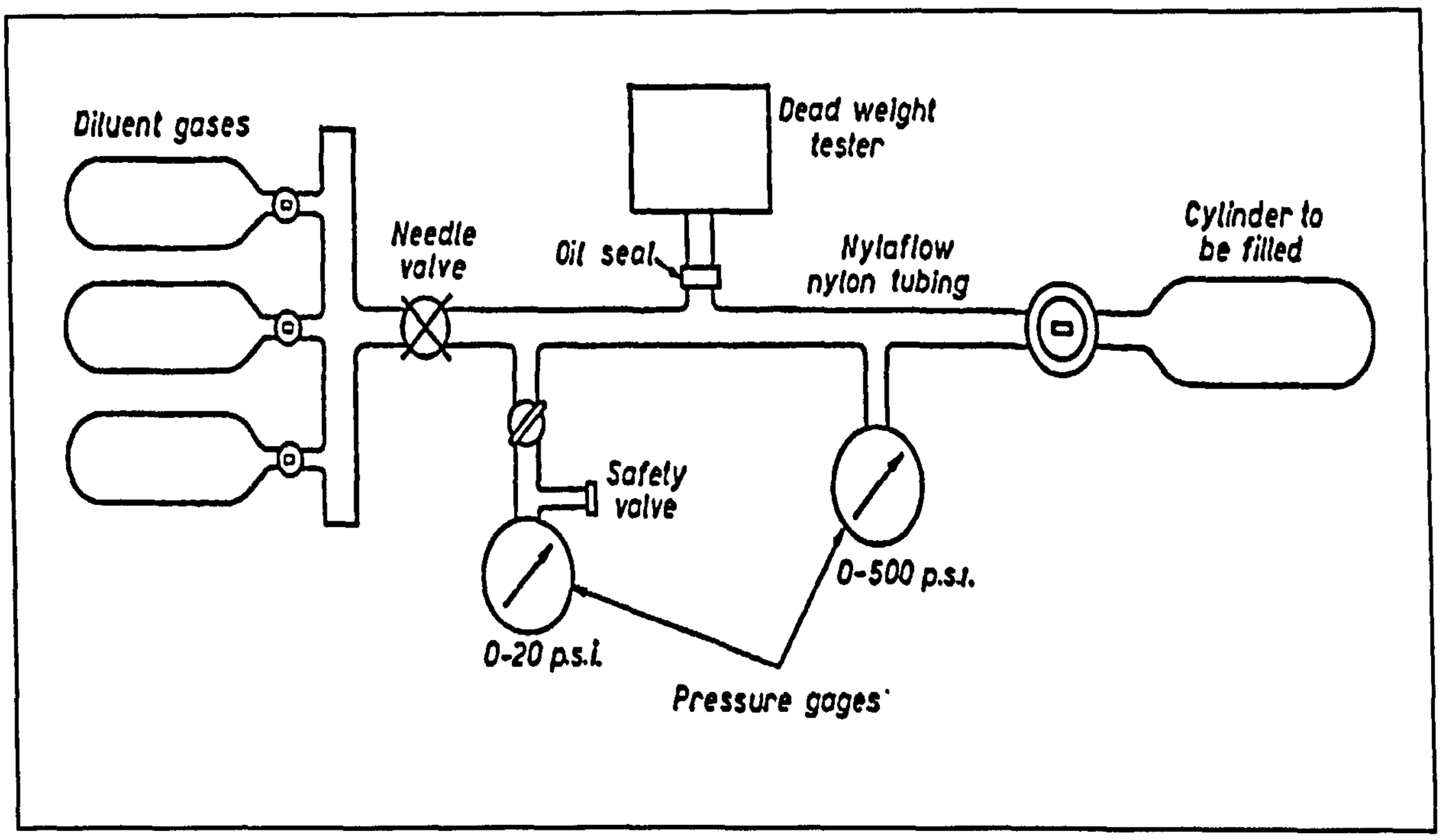


Figure 1.2.1 High Pressure manifold for the preparation of gas mixtures in cylinders[63]

A system based on this method was engaged for generating quantitative vapour-phase infrared reference spectra in work by Richardson and Griffiths[65] and was found to give accuracy within 2% when compared with certified standards.

The main disadvantage associated with static systems is adsorption on the container walls. This varies for different compounds, and can be a particularly significant at low levels especially with strong oxidising and reducing agents such as ozone, nitrogen dioxide, hydrogen fluoride, and nitrated compounds. Adsorption can be reduced by using unreactive

coatings and by increasing the volume size of the chamber (to decrease surface to volume ratio).

1.2.2 Dynamic Methods of Sample Preparation

Dynamic methods involve moving gases. Gases are either pumped or metered using controlled flow rates.

Exponential dilution systems were first described by Lovelock in 1960[66], they basically start with a fixed amount of calibrant, which is mixed with a carrier gas (flowing at a fixed rate) so that the concentration is gradually diluted over time, resulting in a decay curve. The advantage with this method is that a fixed amount of calibrant can be gradually diluted to the required concentrations. The rate of dilution is controlled by the rate of flow of the dilutent gas; by using slow flow rates this method can be an accurate calibration technique. Any unwanted gases can be swept away and adsorption becomes negligible once an equilibrium is established.

Mixing can be achieved by simple flow through, passive mixing (dependant on diffusion) or turbulent eddy devices (paddle or vanes), for small systems (of about 1 L) magnetic stirrers can be used. Assuming perfect mixing the concentration can be calculated by equation 1.4

$$C_t = C_o e^{\frac{-rt}{V}} \qquad \text{Equation 1.4}$$

Where;

C_t = concentration at time t,

C_0 = the original concentration,

V = the volume of the system,

r = the flow rate.

For an ideal system, a plot of log instrument response versus time should give a straight line with a gradient dependent on the flow rate. Another method of sample preparation is by gas stream mixing, an example of a set-up for 3 component mixing is shown in figure 1.2.2, different concentrations can be prepared by using different flow rates for the different gases.

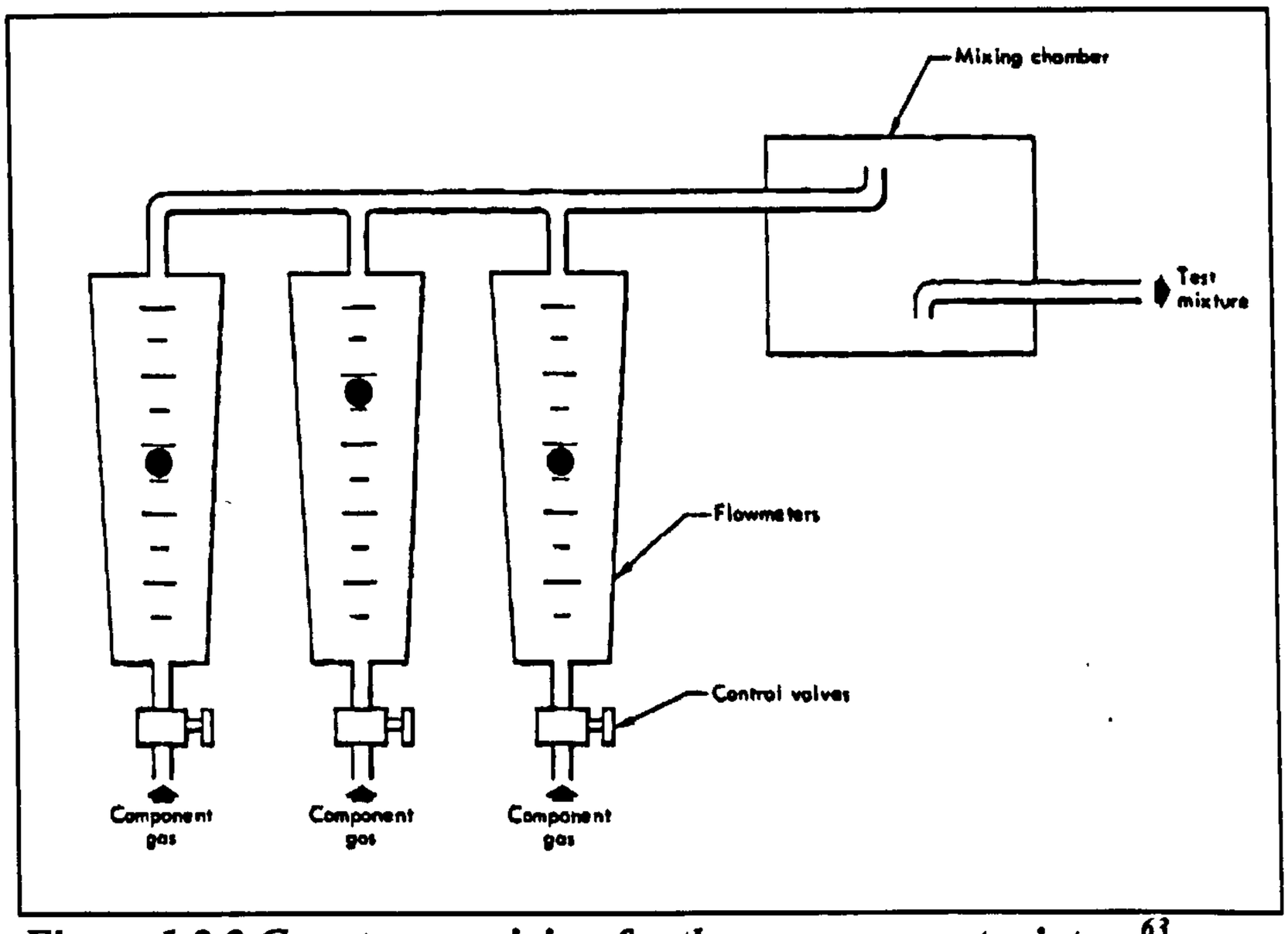


Figure 1.2.2 Gas stream mixing for three-component mixture[63]

Permeation tubes are commonly employed for preparing vapour mixtures. Permeation tubes consist of a tube made of Teflon or some other unreactive polymeric material filled with the compound of interest. The tubes are sealed at both ends and diffusion of the gas occurs through the polymer walls and into the surrounding atmosphere. The permeation rate is dependent on a number of factors including the compound used, the size and thickness of the tubing, and temperature.

Concentrations can be calculated from equation 1.5

$$C_{ppmv} = \frac{P \times (2446/mw)}{F_c} \qquad \text{Equation 1.5}$$

Where;

C = concentration,

P = the permeation rate (ng/min),

mw = molecular weight,

24.46 = (molar volume at reference conditions),

F_c = the flow rate of mixture (ml/min).

An example of the employment of permeation tubes is in investigations of atmospheric levels of NO and NO_2.[67]

1.3 Chemometric Techniques Employed in Vapour and Gas Analysis

Chemometrics refers to mathematical, statistical and logical approaches applied to chemistry. Many techniques can be employed for gas and vapour analysis ranging from methods of improving data such as smoothing, background corrections and deconvolution to building multivariate calibration models. Other methods looking at trends and patterns in data include flow control charts, principal component analysis, dendrograms and neural networks. This section looks at some of the different techniques and their applications, in particular ones relating to the work carried out in this project.

1.3.1 Experimental Design

Experimental design is a statistical approach to planning experimental work. This takes into consideration the different variables, referred to as factors, and their corresponding levels. Examples of different factors can include components present in a mixture, temperature, time, etc. The corresponding levels can be qualitative i.e. present or not present, on or off, etc, or can be quantitative i.e. have various magnitudes e.g. concentration.

Other considerations are replication which allows errors to be determined, and randomisation which prevent biased results and minimises the effect of uncontrolled factors on results.

Most experimental designs are based on factorial design whereby all the possible combinations of factors and different levels are determined. Table 1.3.1 shows the design matrix for a two level (represented by '-' and '+'), three factor (A, B and C) experiment. For F number of factors with N number of levels the total number of experiments would be N^F.

Run	A	B	C
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

Table 1.3.1 Design Matrix for a two level experiment with three factors

The advantage with this method is that it allows all the interactive effects to be estimated. A main disadvantage is that the number experiments required increases dramatically as the number of factors and/or levels increases.

To reduce the number of experimental runs, various partial factorial designs such as partial and central composite can be employed. In full factorial design, many of the runs can be regarded as redundant as the number of higher order interactions increases (e.g. for an experiment containing six factors with two levels each, there are 64 possible runs: one average run, six main effects, 15 two-factor interactions, 20 three-factor interactions, 15 four-factor interactions, six five-factor interactions and one six-factor interaction) and the main effects can still be calculated with fewer runs.[68]

1.3.2 Approaches to Instrumental Variations and Background Drifts

Common problems in analytical measurements are noise and instrumental drifts. One way to compensate for this is with simultaneous reference measurements, for example in thermal conductivity detection in gas chromatography two columns are used, one containing the carrier gas into which sample is injected, and the other just the carrier gas. The detector measures the conductivity for each and the final measurement is a differential signal of the two, thus removing electrical noise. In many cases background and instrument variations are compensated for by repeating the analysis under the same conditions without the analyte of interest, for example the collection of a background scan in FTIR that is ratioed against sample scans to obtain absorption spectra. In detector arrays having duplicates of the same detectors can also compensate for abnormalities and drift.

Methods of increasing the signal to noise ratio includes signal averaging (improvement of signal to noise by using repetitive samples, based on the principal that noise is random so eventually will cancel itself out), signal smoothing and filtering.[69] Different types of smoothing include averaging; a frequently applied technique is Savistky-Golay smoothing, which is based on a least squares procedure.[70] Moving average smoothing and filtering in the Fourier domain are both methods used for treating Raman Spectroscopic data as described in work by Estienne *et al.*[71]

Calibration transfer methods look at ways of dealing with variations that arise between different instruments. Methods include signal processing, where the application of digital filters can be used to isolate target signals from the remaining data and pattern recognition techniques to characterise analyte signals. Koehlar *et al.* describes a calibration transfer algorithm based on piecewise linear discriminant analysis (PLDA) and back-propagation neural networks (BNN) used for passive Fourier Transform Spectroscopy.[72] In passive FTIR, naturally occurring IR radiation is used as the light source thus allowing open-air measurements, but the technique is hindered by lack of stable, reproducible reference spectra

for background correction. The sample data was left as interferrograms and spectral features were either kept or rejected dependant on their widths. Digital filters were then used to enhance signals before using either linear discriminant analysis or the automated detection algorithm, a back-propagation neural network.[73]

To correct for background drift Vogt *et al.* used an artificial pseudo principal component.[74]

1.3.3 Data Pre-treatment

Prior to further analysis, various pre-treatment methods may be applied to the data. This can include the previously mentioned smoothing and filtering.

Other data treatments that can be adopted include centring and scaling.[75]

In mean centring, the average value is subtracted from the data so that the origin of the data becomes the centre of the data (equation 1.6).

$$x_{ik}^{*} = x_{ik} - \bar{x}_k \qquad \text{Equation 1.6}$$

Where;

i is the row index

k is the column index

$\bar{x}_k$ the column mean

Auto scaling employs both the mean and standard deviations as a means of scaling the data i.e. the mean is subtracted and the data divided by the standard deviation (equation 1.7).

$$x_{ik}^{*} = \frac{x_{ik} - \bar{x}_k}{S_k} \qquad \text{Equation 1.7}$$

Where;

$$S_k = \sqrt{\frac{\sum_{i=1}^{N}(x_{ik} - \bar{x}_k)}{N-1}}$$

Normalisation of the data vector to length 1 can also be applied (equation 1.8).

$$x_{ik}^{*} = \frac{x_{ik}}{\|\mathbf{x}_k\|} \qquad \text{Equation 1.8}$$

Where;

$$\|\mathbf{x}_k\| = \sqrt{x_{1k}^2 + x_{2k}^2 + \ldots x_{Nk}^2}$$

1.3.4 Pattern Recognition, Trends and Identification

Various work has been carried out in identification of compounds using FTIR. Different approaches include; pattern recognition, factor analysis, clustering and expert systems. These

methods are employed to identify structural properties and classify data accordingly. Decisions are based on information on peak position, uniqueness and intensity.

As the size and number of IR spectra libraries increases most spectra for single compounds can be identified using routine library searches to identify the closest match. Knowledge based systems use spectroscopic and chemical expertise to assist interpretation of spectra. Mostly this involves comparing spectra with existing libraries of IR spectra. This can be achieved using least square fit procedures as demonstrated by Hong-kui X. *et al.*[76] In this case, samples containing 5 to 11 compounds at approximately 2 ppm were analysed using least squared fit methods. A library of 47 compounds, split into 3 categories: aliphatic hydrocarbons, aromatic compounds and oxygenated compounds were used for identification. IR windows where selected for each compound that were either narrow specifically chosen for that compound or wider more general windows.

One method was to use these windows to eliminate compounds not present in the sample, the other method to build a set of possible compounds by fitting reference spectra one at a time to the sample spectrum.

Optimal least squares fits have also been applied to the quantitative analysis of airborne vapours, using mixtures in ambient air (from 50 ppm to 100 ppb). At 50 ppm the average error was 8.2 %, at 10 ppm 14 % and at 1 ppm the average error was 40 %.[77]

1.3.4.1 Principal Component Analysis (PCA)

Principal Component Analysis is one of the most commonly employed techniques for examining patterns and trends within data. PCA is a form of factor analysis in which the data is split into further sets of data, known as principal components. The principal components are used to describe variations within the data.

Each principal component consists of a loading (also referred to as an eigenvector) and a score (or eigenvalue), which relates the loading to each sample. This is shown in matrix form in figure 1.3.1.

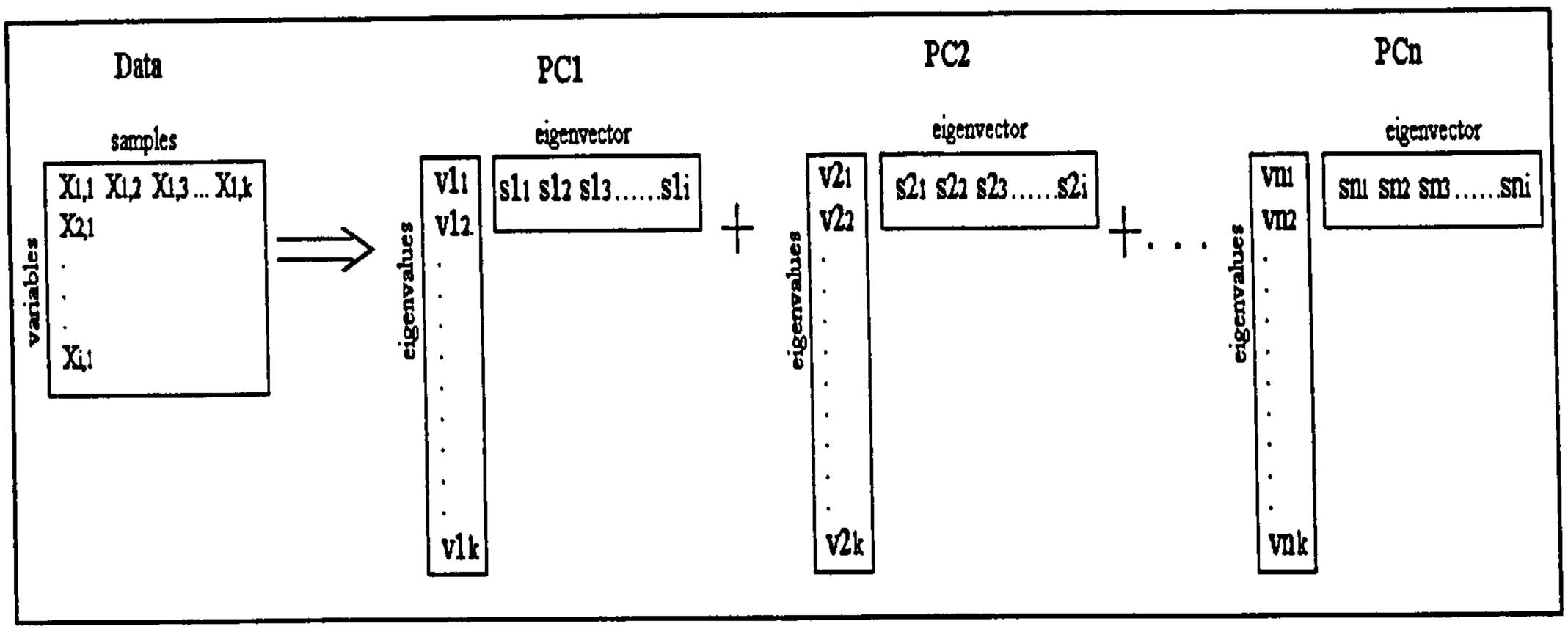


Figure 1.3.1 Breakdown of data into principal components represented by matrices

An equation for this is:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{T} + \mathbf{E} \qquad \text{Equation 1.9}$$

Where;

X = data matrix ($N \times K$)

T = matrix of scores ($N \times A$)

P^T = matrix of loadings ($A \times K$)

E = matrix of residuals ($N \times K$)

(With A the number of PCs used to form the model, N the number of samples, and K the number of variables).

Figure 1.3.2 is graphical representation of the determination of PCs for an example of two-dimensional data and subsequent projection of data onto those PCs. The first principal component is calculated according to a least squares principal (to minimize the residual variation) effectively calculating a 'vector of best fit' passing through the center of the data. The variation accounted for by this vector is subtracted from the data, and the next principal component is calculated so that it describes the maximum residual variation and is orthoganol to the first. This process is repeated till all the variation in the data is accounted for, with each principal component orthogonal to the others.

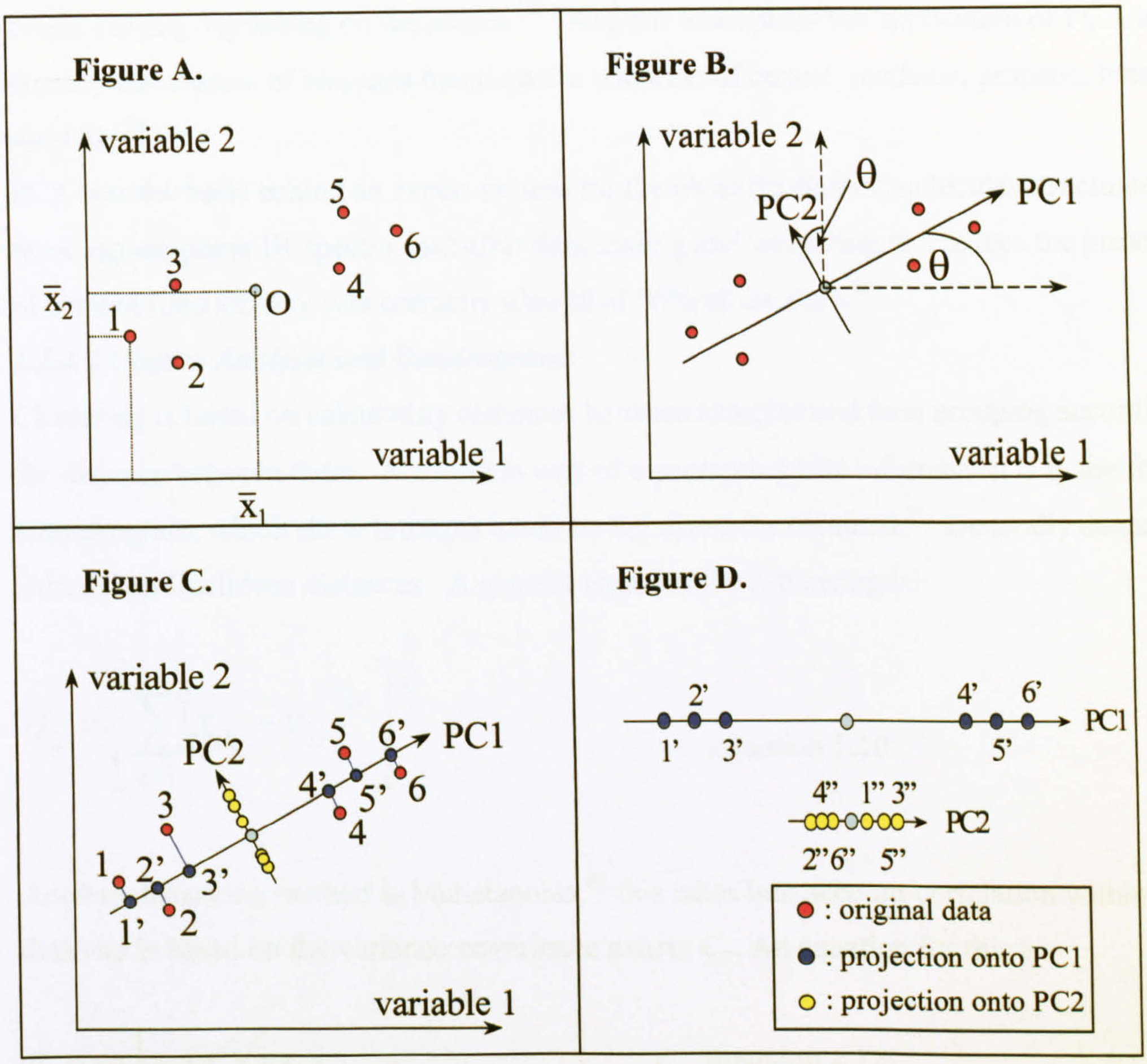


Figure 1.3.2 Graphical representation of the determination of PCs for an example of two-dimensional data
Figure A. Two dimensional data
Figure B. Principal Components (vectors) determined to describe the maximum variation in the data, with PC 2 orthogonal to PC 1
Figure C. Projection of Data points onto PCs
Figure D. Data points projected on PCs
(N.B. O represents the centroid of the data)

By decomposing the data in this manner it is possible to then recombine the data in a way that separates information from noise.

Trends and variations between samples can be investigated by plotting the sample scores for different principal components against each other. An example of the application of PCA is the identification of trends in polynuclear aromatic hydrocarbons (PAHs) in air samples. PAHs are formed by incomplete combustion of organic materials and various sources include traffic, diesel markers, gasoline engine samples, lubricant oil, oven emissions, wood smoke emissions, forest fire smoke and charcoal smoke, with the type and levels of different

University
Library
Hull

PAHs varying depending on the source.[78] Another example is the application of PCA to identify the sources of biogases based on the variation of ethane, methane, propane, butane and CO_2.[79]

PCA was the basis behind an expert system for the identification of molecular structures from vapour-phase IR spectra, and after data scaling and weighting techniques the presence of alcohol functionality was correctly identified 96% of the time.[80]

1.3.4.2 Cluster Analysis and Dendrograms

Clustering is based on calculating distances between samples and then grouping according to the distance between them. A common way of representing this information is in the form of a dendrogram, which show linkages based on the distances obtained.[81] Generally distancing is based on Euclidean distances. A general equation for distancing is:

$$d_{ij} = \left[\sum_{k=1}^{NV} \left(x_{ik} - x_{jk} \right)^N \right]^{\frac{1}{N}} \qquad \text{Equation 1.10}$$

Another distancing method is Mahalanobis,[82] this takes into account correlation within the data and is based on the variance–covariance matrix $\mathbf{C}_x$. An equation for this is:

$$C_x = \frac{1}{(n-1)} (X_c)^T (X_c) \qquad \text{Equation 1.11}$$

where;

$\mathbf{X}$ = the data matrix containing n objects in the rows measured for p variables

$\mathbf{X}_c$ is the column-centered data matrix ($X - \overline{X}$).

The MD (Mahalanobis distance) for each object x_i is then:

$$MD_i = \sqrt{(x_i - \overline{x}) C_x^{-1} (x_i - \overline{x})^T} \qquad \text{Equation 1.12}$$

Different clustering methods can be used; this can be based on the distance to the nearest neighbour or can be the distance to the nearest group (based on the centre of gravity for points in a cluster). An example of the application of cluster analysis is work by Walmsley *et al.* for the determination of suitable sensors to incorporate into a gas sensor.[83]

1.3.4.3 ANOVA (Analysis of Variance)

ANOVA and discriminant function analysis are other methods that can be applied to classification. Work on optical multibead arrays for odour discrimination used these techniques to classify the response obtained for six types of sensor.

Groves and Zeller compared the distance between mean centred responses against vectors derived from the calibration of known samples in their work analysing solvent vapours and also employed neural networks for the identification of individual vapours .[19]

1.3.5 Calibration Models

The main aim of calibration models is to find a relationship between predictor and response variables. Examples of different methods utilised include; Multivariate Linear Regression, Principal Component Regression, and Partial Least Squares.[84]

1.3.5.1 Multivariate Linear Regression(MLR)

The most basic models are Multivariate-linear Regression models. These are based on the premise that there is a linear relationship between variables. Assuming variable y is an observed response that corresponds to variable x (the predictor variable), a linear, first order model relating y to x can be written as:

$$y = \beta_0 + \beta_1 x + \in \qquad \text{Equation 1.13}$$

Where;

β_1 = the regression coefficient,

β_0 = the intercept

$\in$ = the increment by which any individual y differs from the regression line

When multiple variables are used, coefficients relating each response to each predictor variable needs to be modelled, resulting in a series of simultaneous equations.

For a set of i number of y responses from n number of x predictor variables;

$$y_1 = \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1j}x_n + \in \qquad \text{Equation 1.14}$$

$$y_i = \beta_{i1}x_1 + \beta_{i2}x_2 + \beta_{ij}x_n + \ldots + \in \qquad \text{Equation 1.15}$$

In matrix form this can be written as:

$$Y = \beta X + \in \qquad \text{Equation 1.16}$$

Using a least squares estimate, the $\in$ term assumed to be random error is minimised, hence this can be expressed by the normal equations:

$$X'Xb = X'Y \qquad \text{Equation 1.17}$$

Giving the solution:

$$b = (X'X)^{-1}X'Y \qquad \text{Equation 1.18}$$

In spectroscopy, models relating absorbance to concentration can be based on Beer's law,

$$A_{\lambda i} = \varepsilon_{\lambda i} c l \qquad \text{Equation 1.19}$$

Where;

A = the absorption at a specific wavelength,

c = the concentration of the absorbing constituent,

l = the path length of light,

ε = the absorptivity coefficient that relates absorption to concentration

When there is more than one constituent present, the total absorbance is a sum of the absorbance due to each constituent

at wavelength λi,

$$A_{\lambda i} = \varepsilon_{1\lambda i} c_1 + \varepsilon_{2\lambda i} c_1 \ldots + \varepsilon_{n\lambda i} c_n \qquad \text{Equation 1.20}$$

Where;

c_n = the concentration of component n,

$\varepsilon_{n\lambda i}$ = the molar absorptivity for component n at wavelength λi.

The absorptivity coefficients for each constituent vary for different wavenumbers (or wavelengths), hence a spectra consisting of n measurements can be regarded as a series of n simultaneous equations. This allows a MLR model to be employed, where X is a matrix of concentrations of components in the sample and Y the corresponding absorption measurements (various names for this approach include, K-matrix, inverse prediction, reverse calibration or classic least squares). The resulting matrix of coefficients can then be used to predict the concentrations for new spectra. In theory, the coefficients b calculated from the model are inversely related to ε.

Alternatively, inverse models can be built where X is the matrix of absorption measurements and Y the concentration matrix. This method is sometimes referred to as the P-matrix, inverse least squares, forward calibration or ordinary least squares.

The main advantages of using MLR are that it is simple to compute and models can be built from small calibration sets. However, the coefficients are calculated using the inverse of (X'X), which often results in problems with singularity and co-linearity. The effect of

singularity results in unstable prediction models with small variations in the spectra producing large variations in the predicted concentrations. Co-linearity in spectral data arises due to the symmetrical nature of absorption peaks. With variable selection, the disadvantages of co-linearity can be removed. Due to the Gaussian nature of peaks only one measurement per peak need be used, by employing an algorithm to identify variables that contain important information, redundant data as well as regions with large amounts of noise can be ignored. An example of the employment of MLR is in work by Jaakola *et al* in on-line analysis of stack gas compositions, where a modified MLR algorithm was used to fit unknown spectra by using a set of single component calibration spectra.[85, 86]

1.3.5.2 Principal Component Regression (PCR)

One way to reduce the problem of co-linearity in data is to break down the X data (i.e. spectra) into principal components. Each principal component is orthogonal to the others; hence the problem of co-linearity is eliminated. As described for PCA (section 1.3.4) these components consist of scores and loadings. Regression coefficient are calculated that relate the principal components to the predicted data (i.e.) concentrations.

Combining equations 1.9, 1.15 and 1.17, gives equation 1.21 for calculating the regression coefficients.

$$b = (T'T)^{-1}TY \qquad \text{Equation 1.21}$$

Where;

T = matrix of scores

Giving the PCR model equation as:

$$Y = bX(P^T)' + \in \qquad \text{Equation 1.22}$$

A drawback with PCR is that the principal components (that determined from the decomposition of X) may not necessarily relate to the components of interest (Y data).

1.3.5.3 Partial Least Squares (PLS)

Partial Least Squares models (also referred to as Projection to Latent Structures) are the methods most commonly applied to spectra calibration models. The advantages with PLS models are that they can employed on noisy, strongly collinear data and large data sets. [87] As in PCR and PCA, the X data is broken down into scores and loadings, but an additional set of vectors (weights) are calculated. When there is more then one predicted variable (Y), scores and loadings are also calculated for the Y data.

This allows the model to calculate the covariance between X and Y data (the latent variables) and from this determine regression coefficients. Figure 1.3.3 displays a graphical representation of PLS.

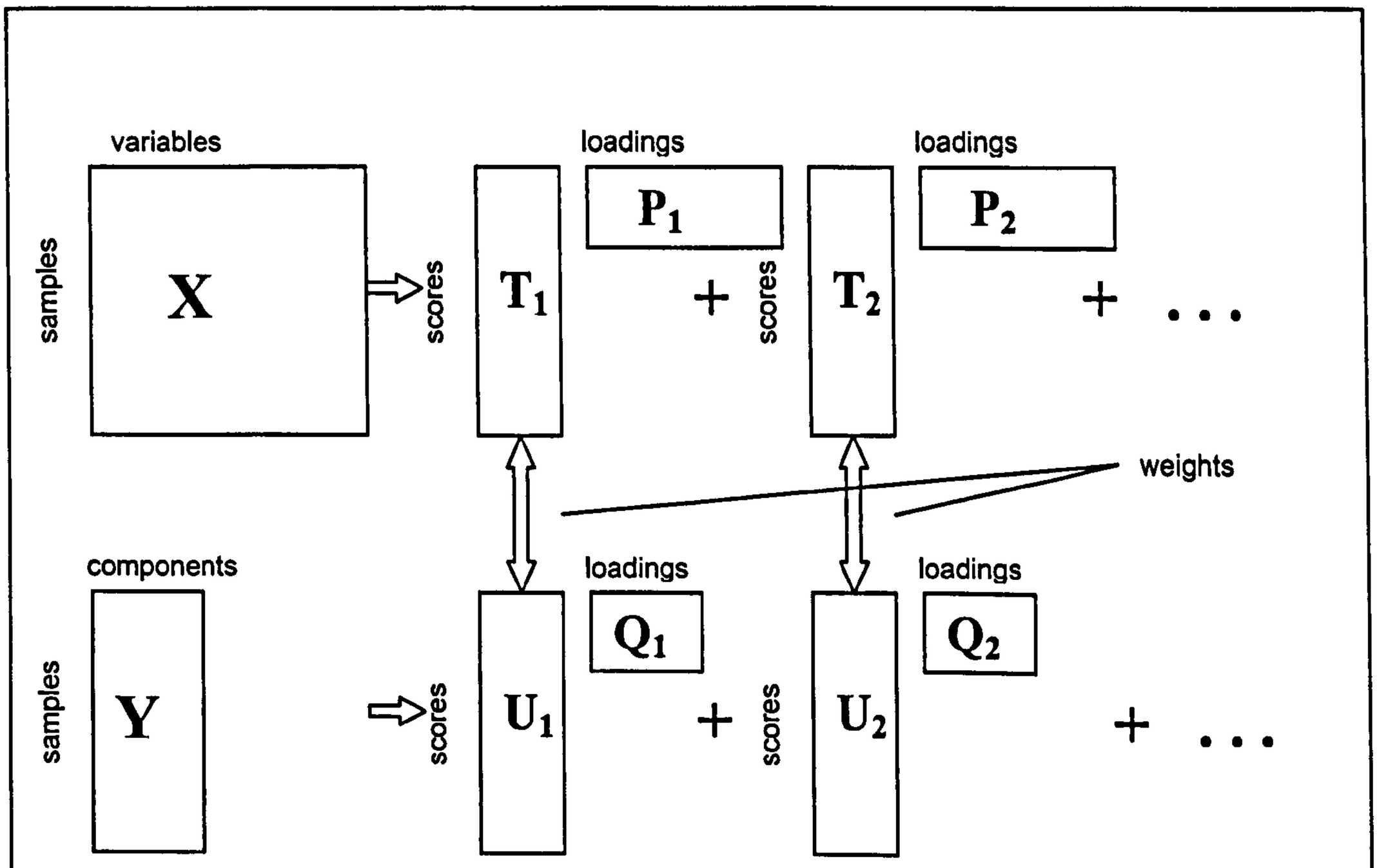


Figure 1.3.3 Breakdown of data by PLS

Equations for PLS can be written as;

$$X = TP' + E \qquad \text{Equation 1.23}$$

$$Y = UQ' + F \qquad \text{Equation 1.24}$$

A set of weights, W*, are determined that relate the X and Y scores, (T and U), so that;

$$T = XW^* \qquad \text{Equation 1.24}$$

and;

$$Y = TQ' + G \qquad \text{Equation 1.26}$$

So that the model can be expressed as;

$$Y = XW^*C' + F = XB + F \qquad \text{Equation 1.27}$$

Where the regression coefficients can be written as;

$$B = W^*C' \qquad \text{Equation 1.28}$$

(N.B. E, F and G are residuals).[82]

Different methods of computation can be employed for this, including; NIPALS (non-iterative Partial Least Squares) and SIMPLS.

An example of PLS applied to FT-IR spectra is work by Johansen *et al.,* where it was employed in the calibration of a FT-IR spectrometer for ambient air monitoring.[88]

Griffiths *et al.* investigated the feasibility of using unattended open path FTIR with PLS models. The conclusion reached was that better accuracy for VOC models could be obtained at 8 cm^{-1} resolution as opposed to higher resolution due to the increased effect of pressure broadening. With the PLS modelling, background spectra of shorter pathlengths could be utilised.[89]

1.3.5 Variable Selection

Variable selection refers to the process of choosing variables from a larger set of variables, in spectrometry where the variables are measurements made at different wavelengths this is commonly referred to as wavelength selection. Manually selecting variables is difficult, particularly with large amounts of data; one approach for spectrometric data would be to choose wavelengths corresponding to regions of maximum absorption thus giving the advantage of high signal to noise; however, these may not necessarily give the best prediction results, particularly in mixtures where overlap of absorption bands may occur. Building models for every single combination of variables would be extremely time consuming and unfeasible for large amounts of data. A better approach is to apply a variable selection algorithm.

Different approaches to algorithms can be applied to variation selection, these are combined with a calibration model technique, such as MLR, PLS or neural networks, and the model results can be compared to determine which variables give the optimum results. Approaches include, forward selection, where variables are added one at a time to a model and are either kept in or removed depending on whether the model is improved, the reverse of this is backward elimination where models start with all the variables or they are then removed one at a time. Stepwise models are a combination of the two and include both addition and removal stages. Genetic algorithms are a more complicated approach, based on the idea of natural selection where strings of variables are crossed and mutated, with strings that give good results given dominance till eventually an optimal set is chosen.

Work by Xu and Zhang compared these approaches, plus a leap and bound method based on comparing results of subsets and another method based on orthogonal descriptors. In their

work, the first three approaches did not give quite as good models as the other methods, and results tended to favour the genetic and leap and bound strategies.[90]

The type of modelling employed also has a bearing on the ability of the variable selection process, work by Spiegelman *et al.* employed a method based on PLS calibration. Variables were ranked in order according to a tuneable signal to noise with variables added in order one at a time and variables giving the lowest cross-validated mean-squared error (CVMSE) for the specified number of latent variables were selected. [91]

In work by Estienne *et al* investigating at Raman spectra, different models were applied including Classical Least Squares, Inverse Least Squares, Stepwise Variable Selection with Multivariate Linear Regression, Genetic Algorithm with Multivariate Linear Regression, PCR with variable selection and PLS. In this case (for synthetic Raman spectra ranging from single to 5 constituent mixtures) stepwise MLR methods were found to give the best models.[71]

Research by Jouan-Rimbaud *et al.* indicated that genetic algorithms could still result in the selection of irrelevant variables and found selection results were improved by applying a forward selection to the chosen subsets. [92] MLR was favoured over PLS and PCR due to its ease of computation, variables were related directly to the original data allowing direct interpretation, and previous work had shown MLR to give slightly better results than PLS. Work by Broadhurst *et al.* applying VS-MLR and VS-PLS to pyrolysis mass spectrometry found that both the MLR and PLS models performed to the same standard.[93]

The variable selection procedure employed in this research utilised MLR and employed an addition and a removal mode whereby variables were added or removed from the selected set depending on whether the PRESS (predicted residual error of the sum of the squares (see equation 1.29)) improved.[94]

$$PRESS = \sum_{i=1}^{n} (\tilde{y}_i - y_i)^2 \qquad \text{Equation 1.29}$$

Models built from the variables selected were then assessed by comparing the standard errors in prediction (SEP (see equation 1.30)).

$$SEP = \sqrt{\frac{\sum (y - \hat{y})^2}{N - 1}} \qquad \text{Equation 1.30}$$

The VS-MLR procedure and its application in this work are discussed in more detail in section 3.2 and 4.2.

1.4 Aims of this Research

The aim of this work was to develop a portable gas analyser system with the aid of chemometrics.

The system would be based on a novel infrared detector array with each detector in the array measuring a different wavelength.

Research by the Optoelectronic group at Hull University into the development of QWIP detectors indicated it would be possible to design detectors to measure at specific wavelengths in the mid IR range. As discussed earlier, IR analysis has the advantage of being a fast, non-destructive and selective means of analysis; combining this with the advantages of detector arrays (i.e. small, portable devices) can lead to the development of useful, portable analysers that can be tailored for a number of different situations. Examples of work developing IR detector arrays include the previously mentioned work by López *et al.* [45,46,47] A major difference between the research carried out here and previous work developing IR detectors is the approach taken to determining which wavelengths to measure. Usually the wavelengths are decided based on the main absorption bands of the gases of interest (discussed in section 1.3.4), however in this case models were built from spectra obtained from real mixtures and a variable selection algorithm based on MLR was utilised to determine suitable wavelength sets.

The objectives in this research were:

- To collect IR spectra for multivariate models
- Apply variable selection to the IR spectra to identify suitable wavelengths to measure in the array
- Develop a Dilution System
- Imitate the effect of detectors using narrowband filters by FTIR
- Measure mixtures of CH_4, CO, CO_2 and N_2O
- Investigate different optical layouts and signal processing
- The development of a prototype system
- Investigate feasibility of developing a system to measure volatile organic compounds

Collection IR Spectra for Multivariate Models

The first stage, the collection of IR spectra for multivariate models (described in section 2.2), was the preparation of different gas samples of CH_4, CO, CO_2 and N_2O, and collecting the

corresponding FTIR spectra from them. The goal of this work was to collect suitable spectra for building wavenumber selection models.

The gas samples were prepared using a simple static system whereby different amounts of gases were injected into a gas cell. This method of preparation was employed as it was a simple, cost effective means of producing samples with mixtures of different concentrations and required no complex apparatus. According to reference 61, an accuracy of 3 – 5 % is achievable with this technique, which was comparable to many of the other methods of preparation.

Part of this work included investigating different data treatment methods, as alternatives to the traditional method of ratioing against a background spectrum, to remove the effect of background variations.

Variable Selection and Other Models (described in section 2.1)

This stage of the work employed a variable selection algorithm based on Multivariate Linear Regression in order to determine sets of wavenumbers that would be suitable to incorporate in a detection array. The aim of the variable (wavenumber) selection was to reduce the number of measurements required without increasing the prediction error. Models built from just a few selected wavenumbers were compared to PLS models built from the entire spectra to prove that comparable prediction errors were obtained.

Development of a Dilution System (described in section 2.3)

A dilution system was developed as a means to create different concentrations of samples from a single standard gas mixture. This allowed responses over a wide concentration range to be investigated. A gas chromatograph sampling loop was also included in the system allowing samples to be also measured by GC, thus GC and IR responses could be correlated.

Characterisation of Narrowband Filters by FTIR (described in section 2.4)

Narrowband filters restrict light except at the wavelength of interest. These combined with a general detector (in this case the one in the FTIR) were employed as a means to imitate the effect of detectors that measure only a specific wavelength. The point of this work was to show that it would be feasible to use an array measuring at the selected wavelengths to measure mixtures of CH_4, CO, CO_2, and N_2O.

Investigations into Different Optical Layouts and Signal Processing (section 2.5)

This stage of the work focused on the design and subsequent building of an optical set-up employing the narrowband filters. This looked at the problems of signal processing, alignment of optical components and the effect of different detectors.

Development of Purgeable Box (section 2.6)

The aim of this work was to build a purgeable box, hence a prototype measuring system, using the optimum optical layout determined by the previous work.

Volatile Organic Compound Analysis (section 2.7)

This work looked at the feasibility of developing an IR analyser for identifying/ measuring different volatile organic compounds. This included cluster analysis of IR vapour spectra for different organic compounds and variable selection models on simple four component mixtures.

2 Experimental

2.1 Data treatment and chemometric analyses

This section contains details about the data treatment and chemometric analyses used throughout this work. The preparation of samples, collection of data and different experimental set-ups are described in sections 2.2 – 2.8.

2.1.1 File formats

Three different instruments were utilised to collect data.

The majority of FTIR spectra were collected on a FTS3000MX Excaliber FTIR Spectrometer, with Win-IR Pro version 2.6 software (from Bio-Rad laboratories, Hertfordshire, UK). FTIR spectra for different organic vapour samples (details about this work can be found in section 2.7) were collected on a Perkin Elmer Paragon 1000 FTIR Instrument. The remaining data i.e. non-FTIR data (see section 2.4, 2.6 and 2.7) was collected with a PC-516/DAQ card ™ employing LabVIEW ™ version 5.0.

Spectral data collected with FTS3000MX Excaliber FTIR Spectrometer

The Win-IR software utilised with the Bio-Rad spectrometer saved data as Bio-Rad spectra format files with a .bsp extension, these opened up in the software as a spreadsheet showing the spectra saved within that file, and contained information from the original interferrograms. As well as this, spectra after any additional processing, the scan rate, number of co-adds, time, the spectrum history (including the collection properties) were stored. To open the data in Excel, the spectra of interest were saved as a .csv file which included listing the file type i.e. single beam, display direction, peak direction, wavenumber measurements and the corresponding responses at the wavenumbers for each spectrum. Later on, Excel was installed onto the computer controlling the spectrometer (Win-IR software), so that the data could be cut and pasted directly into an Excel spreadsheet.

For use in MATLAB®, the excess information was deleted to leave the data in the form of a matrix and the file saved as an .asc file.

Spectral data collected with Perkin Elmer Instrument

Spectra collected on the Perkin Elmer instrument, were saved as .asc files which could be opened directly in Excel; these contained the spectra name, the time and the wavenumbers and corresponding responses for a single spectra. For use in MATLAB® the excess information was deleted and the file resaved as an .asc file.

Data collected with LabVIEW™ version 5.0 with a PC-516/DAQ card™ (see sections 2.4,2.6 and 2.7)

Data collected with LabVIEW™ were saved into a text file (.txt), which could be opened directly in Excel.

2.1.2 Software employed for analysis

Microsoft Excel

Plots of data and calibration plots were generated in Microsoft Excel versions 1997 and 2000. Averages, standard deviations, correlations between data and frequency tables were calculated using the statistical functions within Excel.

MATLAB®

Mean centring, Autoscaling, Normalisation, Savitsky Golay Smoothing, PCA, PLS, MLR and VS-MLR were all performed in MATLAB® version 4.2.

Before performing using any of the above programs the IR data had to be transposed so that the columns of data corresponded to different variables (wavenumbers) and the rows to different sample (for Excel, the matrix had to transposed otherwise the spreadsheet width would be exceeded).

2.1.3 Data pre-treatments

Various pre-treatments applied to the data included mean centring, autoscaling, normalisation, Savitsky Golay smoothing and reconstruction using PCA.

Mean centring, autoscaling and normalisation

These treatments were normally applied by employing programs in MATLAB® from PLS Toolbox, Eigenvector Technologies.

Savisky-Golay Smoothing

Savtisky-Golay smoothing was only performed on the spectra for single gases (see section 2.2). This was investigated as a possible technique for removing background variation between different spectra. A program written by Sijmen de Jong, Unilever Research Laboratorium Vlaardingen in 1993 and modified by Barry M. Wise, 1996, was used to perform Savitsky-Golay smoothing and differentiation.

The command used in MATLAB® was;

```
> y_hat = savgol(y,width,order,deriv);
```

Where; "y_hat" is the name of the output matrix of smoothed and differentiated row vectors, "savgol" the name of the program, "y" the name of the matrix of data to be smoothed, "width" the number of points to employ in the filter, "order" the order of the polynomial to use and "deriv" and the number of derivative. For first order, first derivative smoothing the last two inputs were both 1.

Reconstruction of the data employing PCA

Reconstruction of data by employing PCA was also only performed on the spectra for single gases. This was investigated to see if it was possible to reconstruct the data in such a manner that only information relating to the analytes of interest was retained.

The PCA program used was from PLS Toolbox, Eigenvector Technologies.

The command used in MATLAB® was;

>[scores,loads,ssq,res,q,tsq] = pca(data);

where; "scores", "loads", "ssq", "res", "q", "tsq" and "tsqs" are output matrices corresponding to scores, loadings, variance information, residuals, q limit, and t^2 limits. "pca" is the program name, and "data" is the matrix of data that PCA is to be performed on.

The main outputs of interest were the "scores" matrix, which consists of the sample scores (rows) obtained for different principal components (PC) (columns) and the "loads" matrix, which consists of the loadings of different principal components (rows) for different variables (columns).

Plots of the loadings for each principal component were compared with reference absorption spectra (taken from NIST database[64] shown in the appendix (section 6)) to select loadings containing similar information.

Reconstruction of the data was performed by selecting the column of interest from the score matrix and multiplying it by the corresponding row from the load matrix. To reconstruct data from just PC2, column two from the score matrix was multiplied by row two of the loadings matrix.

2.1.4 Methods and programs employed to analyse data.

PCA

The MATLAB® program employed to perform PCA is described in section 2.1.3. PCA was mainly used to generate score plots to check for outliers in samples collected to build calibration models (as in section 2.2) and to compare similarities between samples (as in section 2.7).

Sample score plots could be generated by plotting different columns of the score matrix against each other, i.e. plotting the values in column one against those of column two would give a plot of sample scores for PC1 against scores for PC2.

Clustering

Clustering was mainly employed in the work investigating organic vapour spectra to see whether it was possible to cluster the spectra according to the type of compound.

The program used for clustering and generating dendrograms was from Eigenvector Technologies.

The command used in MATLAB® was;

> cluster(dat,labels)

where; "cluster" is the program name, "dat" the matrix of data to be clustered and "labels" a matrix of identifying labels to be attached to each sample.

The output of the function is a dendrogram showing the distances between the samples. The program gives the option of choosing between KNN and K-means distancing, the type of data scaling (autoscale or mean centring), and an option to do PCA on the data and base the distance measure on the raw scores or on Mahalanobis distances.

PLS

The PLS program used was from PLS Toolbox, Eigenvector Technologies.

The command employed in MATLAB® was;

>[p,q,w,t,u,b,ssqdif] = pls(x,y,lv);

Where, "p", "q", "w", "t", "u", "b" and "ssqdif" are output matrices corresponding to x loadings, y loadings, x weights, x scores, y scores, inner relation coefficient and the fraction of variance used in the x and y matrices. "pls" is the program name, "x" is the matrix of predictor variables (in this case spectra), "y" the matrix of predicted variables (concentrations) and "lv" is the number of latent variables to be calculated. The output matrices were used to predict concentrations (y) from spectra (x) to determine the standard error in prediction for the calibration data (SEPC), as well as predicting concentrations for new spectra not previously employed in PLS to determine the standard error in prediction for validation data (SEPV).

MLR

MLR was performed in MATLAB® using the command

>coeff = x\y

Where; "coeff" is the output matrix of MLR coefficients, "x" is the matrix of predictor variables (spectra) and "y" the matrix of predicted variables (concentrations).

The command calculates coefficients by determining the inverse matrix of x.

VS-MLR

For a more detailed account of the selection algorithm refer to reference 94.

The data analysed consists of four matrices: the training spectra, the training concentrations, the validation spectra, and validation concentrations. The program randomly sorts the variables, and then performs MLR each variable using the training data to build the model and the validation data to evaluate it. The result giving the lowest predicted error sum of squares (PRESS) is stored. The algorithm works in two stages, the first an "addition mode" where variables are added to the model and kept if the PRESS is improved. When completed, the second stage, "a removal mode", removes variables one at a time from the selected set, removing variables that increase the PRESS value; this results in a set of 'best" variables.

Figure 2.1.1 gives a basic schematic of the stages of the algorithm.

For each run; the maximum number of variables, convergence level 1 (cl 1), convergence level 2 (cl 2), and the number of iterations could be set to different values. In order to obtain the best selection results these four values need to be optimised.

The maximum number of variables ('maxvars') limits the number of variables used in the selection process, thus reducing the computation time, this is advantageous for data with a large number of variables many of which may be redundant, however if too few variables are tested in the selection process this can result in the loss of important variables.

The most important factors of the selection process are the convergence levels, as these determine whether variables were retained or removed from the selected sets. The convergence level is the factor that the previous PRESS value is multiplied by, before comparing it to the new PRESS value cl 1 relates to the addition stage and cl 2 to the reduction stage.

In the addition stage,

if new PRESS > lastPRESS × cl 1 , the variable is not added

In the removal stage,

If newPRESS > lastPRESS × cl 2, the variable is not removed

The default values for the cl's are 1.

The order variables are added into models has a major influence on the final set of chosen variables, hence the variables are randomly sorted prior to the addition stage, this requires the process to be reiterated a number of times to allow different random orders of variables to be added to the model.

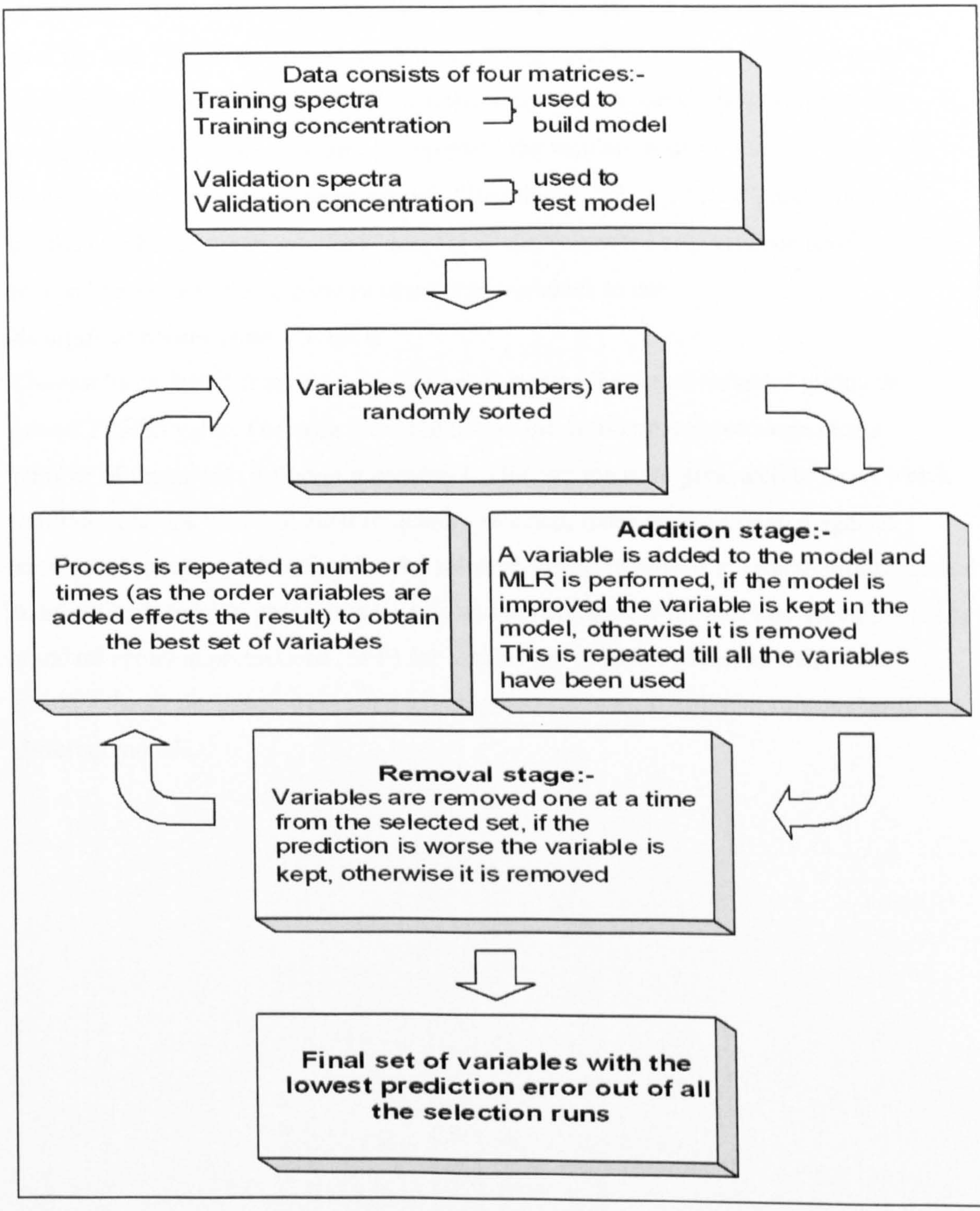


Figure 2.1.1 General Scheme of Variable Selection Algorithm

The command used in MATLAB® was;

[ypred, press, selected, coeffs] = vsmlr(t_spectra, t_conc, v_spectra, v_conc, iterations, cl_1, cl_2, maxvars)

Where; "ypred", "press", "selected", "coeffs" are the output matrices corresponding to the predicted values for y, the best PRESS value obtained, the selected variables that gave the best PRESS value and the corresponding coefficients. "vsmlr" is the name of the program, "t_spectra" the matrix of training spectra, "t_conc" the corresponding matrix of training concentrations, "v_spectra" the validation spectra and "v_conc" the corresponding validation concentrations. "Iterations", "cl _1", "cl_2", and "maxvars" are the number of iterations, "cl 1" and "cl 2" the values for convergence levels one and two and "maxvars" the maximum number of variables to use.

Analysis of results from VS-MLR

The results collected from the selection process were the set of variables giving the lowest PRESS value. For each selection model the entire process was repeated a number of times with different parameters. Histograms were generated to show which variables/regions were the most frequently selected; these were compared against absorption spectra to identify what the selected variables related to (whether they related to specific absorption peaks or were selected to compensate for baseline variations). The standard errors in predictions (SEP) for each of the different gases, as well as the average for all the gases, were used to compare results from different runs and with different models.

2.2 Collection of IR spectra for multivariate models

2.2.1 Single component samples

The spectrometer and scan conditions:

The IR spectra were collected on the Excaliber FTIR Spectrometer (described in section 2.1.1). Scans were collected as single energy beam scans (1.2 filter, sensitivity 1, gain range radius 40, speed 5 kHz, co-addition of 8 scans) from 0 to 9000 cm^{-1} with a resolution of 4 cm^{-1} (taking into account the cut off points of the source and KBr windows, the range was effectively 400 cm^{-1} to 4000 cm^{-1}), using a DTGS detector. The gases used were CH_4, CO, CO_2, N_2O and SF_6 from Intergas, Stafordshire, UK supplied by Engineering and Welding Supplies Ltd., Hull, UK, with purities of 99.995, 99.95, 99.995, 99.95, and 99.9 % respectively.

Samples were prepared in a mini long path gas cell model 1.6 from Infrared Analysis Inc. Anaheim, CA. (7.2 meter path length), with volume of 530 cm^3.

Sample preparation

The gas cell was purged with helium to remove any residual air or previous samples. Samples were then prepared using a 10 ml gas tight syringe to inject different quantities into the cell (performed at normal room temperature and pressure). After injecting mixtures, the cell was given a few minutes to allow the mixture to equilibrate. Between samples the cell was purged with helium till any residual peaks were removed. Residual peaks were checked for by scanning the purged gas cell and ratioing against scans collected prior to sample injections. The scans were then saved as .csv files for import to Excel and MATLAB®.

Initial calibration spectra were prepared by injecting between 15 and 150 ml of the analyte of interest into the gas cell (corresponding to concentrations between 2.8 and 27 %). The concentrations of samples prepared for each gas is shown in table 2.2.1.

Gas	Concentrations of samples prepared
CO	15.1%, 26.4%, 15.1%, 7.5%, 26.4%, 15.1%, 7.5%, and 3.8%
N_2O	8.5%, 4.7%, 2.8%, 8.4%, 8.5%, 15.1%, and 8.5%
CO_2	9.4%, 15.1%, 5.7%, 1.9%, 9.4%, 9.4%, and 8.0%

Table 2.2.1 Concentrations of samples prepared for single component samples

Additional Calibration Spectra

Additional calibration spectra were prepared using lower concentrations (range of 0.18 to 4.8%, corresponding to between 1ml and 20 ml injections). After injections the cell was allowed to equilibrate for at least 10 minutes before scanning.

Gas	Concentrations of samples
N_2O	1888, 18878, 18878, 9434, 9434, 28302, 28302, 37736 ppm
CO_2	47170, 28302, 9434, 28302, 9434, 47170, 18878 ppm
CH_4	28302, 5660, 28302, 15094, 9434, 28302, 18878, 9434, 9434 ppm
CO	9434, 18878, 28302, 5660, 18878 ppm

Table 2.2.2 Concentrations of samples prepared for single component samples

MLR and VS-MLR models were built from the spectra obtained, and the effects of different data pre-treatment including mean centring, autoscaling, Savisky-Golay smoothing and reconstruction of the data employing PCA was investigated.

2.2.2 Collection of samples with reference peaks

CH_4 as an internal standard

Samples of CO were prepared in the long path gas cell using a fixed amount (8 ml) of CH_4 injected into to each sample (equivalent to 1.5 %).

The concentrations of CO used were, 3.77, 3.20, 4.72, 1.32, 2.26, 0.38, 1.89, 0.94, and 0.19 %.

The spectra were divided by the response at measurement number 1565 (corresponding to 3018.6 cm^{-1}) and models then built using both this data and with the untreated data.

Samples with an external reference and N_2O

Scans using an external reference were obtained by inserting a piece of polystyrene film (provided by BioRad for FTIR instrument calibration purposes). Samples of different concentrations of CH_4 with a fixed amount of N_2O (0.94 %) were prepared. After collecting the spectra for the gas cell (plus polymer film) the spectra were divided by the response obtained at the wavenumber of the chosen reference peaks at measurement number 756 (corresponding to 1456.3 cm^{-1}) and measurement 1028 (1980.9 cm^{-1}). The results were compared with results from raw data and from spectra divided by N_2O peaks at measurement number 677 (1303.9 cm^{-1}) and measurement number 1338 (2578.8 cm^{-1}).

CO_2 and N_2O samples using a fixed amount of SF_6

Samples of CO_2 were prepared in the long path gas cell with concentrations corresponding to 1.88, 1.32, 0.94, 2.45, 0.94, 0.38, 2.26, and 0.38 % made up in helium with a fixed amount of SF_6 (0.38 %) added.

Samples of N_2O were prepared in a similar manner using concentrations of 0.38, 0.75, 1.70, 0.19, 1.13, 2.83, 0.19, 1.13, 2.83, 0.38, 1.51 %, again with 0.38 % SF_6.

Models were built from spectra with and without division of the reference SF_6 peak at 941 cm^{-1}.

2.2.3 Mixtures of the four gases plus SF_6 as internal standard

Mixtures of the four gases were prepared for different combinations of low, medium and high amounts for each of the analytes. The % concentration of gases added to each sample is shown in table 2.2.3.

Concentration (%) of gases in each sample				
Sample	**CH_4**	**CO**	**CO_2**	**N_2O**
1	0.19	0.19	0.19	0.19
2	1.89	0.94	0.38	0.19
3	0.38	0.38	1.89	1.89
4	0.38	1.89	1.89	2.08
5	1.89	1.70	0.19	0.38
6	2.26	1.70	0.38	1.70
7	0.19	0.19	0.19	1.70
8	1.13	0.38	0.94	1.89
9	1.89	0.19	1.70	0.19
10	1.89	0.38	1.70	0.94
11	0.19	1.70	0.19	0.38
12	1.32	2.26	0.57	1.32
13	1.70	0.38	0.38	0.38
14	2.26	2.08	2.26	0.38
15	0.19	0.19	1.89	0.19
16	0.19	0.94	2.26	0.38
17	1.70	0.19	0.19	1.89
18	2.26	0.38	2.26	2.08
19	0.94	0.94	0.94	0.94
20	1.89	1.89	1.89	1.89
21	0.19	1.89	0.94	1.89
22	0.38	2.08	0.57	2.08
23	0.57	2.08	2.08	0.57
24	0.94	1.13	1.70	0.19

Table 2.2.3 Concentrations of gases (%) added to each sample

These consisted of 16 different combinations of high and low values, and 8 combinations that included medium values, where; 'high' amounts ranged from 1.51 to 2.26 %, 'low' values from 0.19 to 0.57 %, and 'medium' values between 0.94 and 1.32 % (corresponding to injection volumes between 1 and 12 ml).

A fixed amount of SF_6 (0.38 %) was added to each sample to act as an internal standard. Different data treatments were applied to the data including; dividing the data by the SF_6 reference peak (at wavenumber 941 cm^{-1}), and dividing the data by the SF_6 reference peak followed by normalising the data by sample. The data was also converted into absorption spectra by ratioing the spectra against scans made of the purged cell that were taken between samples.

More samples were prepared a couple weeks later (in the same manner) as an extra validation set for the modelling, this included repeats of samples 1 and 20. The concentrations are shown in table 2.2.4.

Concentration (%) of gases in each sample				
Sample	**CH_4**	**CO**	**CO_2**	**N_2O**
1	0.75	0.94	1.89	0.57
2	1.89	1.32	2.08	1.13
3	0.19	0.19	0.19	0.19
4	1.89	1.89	1.89	1.89
5	0.94	0.38	0.38	0.94
6	1.32	0.94	0.38	1.13

Table 2.2.4 Concentrations of gases (%) added to each additional sample

Effect of Resolution

To compare the effect of reducing resolution, spectra were collected at low resolution using both 16 cm^{-1} and 32 cm^{-1} resolutions. At 16 cm^{-1}, measurements were effectively the average response for a window of 15.4 cm^{-1}, with the centre of each window every 7.7 cm^{-1}. This resulted in 1024 measurements between 0 and 8000 cm^{-1}. At 32 cm^{-1} resolution, the windows are 30.9 cm^{-1}, with the centre every 15.4 cm^{-1} giving a total of 512 measurements. The spectra were divided by the SF_6 reference peak that occurred at 941 cm^{-1}; for the 32 cm^{-1} resolution spectra this corresponded to measurement number 62, for 16 cm^{-1} resolution data the same peak corresponded to measurement number 130.

Reproducibility of spectra

Samples consisting of 1.88 % (10 ml injections) each of CH_4, CO, CO_2, and N_2O and 0.38 % of SF_6 and of 0.19 % (1 ml injections) each of CH_4, CO, CO_2, and N_2O plus 0.38 % of SF_6 were prepared and scanned. Repeat samples, for both concentrations, were made two weeks after the first, and a further three repeat samples made four days later. The three repeat samples made on the same day were prepared and scanned within a two-hour interval.

The relative standard deviations between samples made on different weeks, and between samples made on the same day were compared for both raw data and scans divided by the SF_6 reference peak.

2.3 Multivariate Modelling with four component IR spectra

2.3.1 Optimisation of the VS-MLR modelling parameters

The VS-MLR algorithm is described in section 2.1.4. The effect of the different parameters (i.e. the convergence levels, number of iterations and maximum number of variables) was investigated using the four component mixtures at resolution 32 cm^{-1} using data that was divided by the reference peak and data that was divided by the reference peak and normalised. PCA was performed on the entire data set utilising the score plots to choose which samples to employ as the training data and to check for any outliers or abnormal spectra.

Peripheral samples plus a few central samples were selected from the score plots to use as training data (this ensured the calibration data was fully representative of all the samples) with the remaining samples employed as validation spectra.

From these and the other score plots samples 1, 2, 4, 5, 7-12, 14, 17, 19-22 and 24 were chosen as the training data and the remaining samples used as validation, VS-MLR was then applied to the data.

The effect of changing the cl's was investigated using different values of 0.1 increments for cl 1, this ranged from 0.5 to 2, for cl 2 this ranged from 0.6 to 3.

Originally each run was performed using 10 iterations; this was repeated with the number of iterations increased to 20, 50 and 100 to see if increasing the number of iterations had any significant effect. The maximum number of variables was also changed to determine how many variables were required in the process.

2.3.2 PLS Modelling

PLS modelling on entire spectra

The samples were again split into calibration and validation data sets for the VS-MLR modelling as described for the PLS modelling.

PLS models were built from the calibration set (using the entire spectra) employing different numbers of latent variables, the models were then used to predict concentrations for the calibration set, the validation set and the new validation sets. The SEPs obtained for the three different data sets with different numbers of latent variables were compared to determine the optimal number of latent variables required.

PLS models were also built after applying different data pre-treatments. This included: using the raw data, absorption data, data divided by a reference point corresponding to a

wavenumber with no absorption peaks (as opposed to an internal standard reference peak, in this case variable 130 corresponding to wavenumber 1990 cm^{-1} was employed), and data divided by a reference point followed by normalisation (by sample). In the case of the absorption data, using the entire spectrum gave high prediction errors due to the wide variation in response at cut off regions, therefore, these regions were removed to give better results.

PLS modelling on selected wavenumbers

Models using the selected variables at 139, 149, 200, 205 and 227 corresponding to 2129, 2284, 3071, 3148 and 3488 cm^{-1} respectively, were built for the raw data and for data with different pre-treatments applied (same pre-treatments employed as in the previous PLS modelling).

2.3.3 Conversion of wavenumbers to wavelength measurements

The selected variables relating to measurements at 2129, 2284, 3071, 3148, and 3488 cm^{-1} correspond to 4.6970, 4.3783, 3.2563, 3.1766, and 2.8670 μm (to 4 d.p.). To find out the effect on response when the equivalent wavelengths were rounded to 2 d.p. as well as to the nearest ± 0.05 μm, the units on the spectra were converted to wavelengths using the winIR software available on the FTIR and the corresponding response to the rounded wavelengths were determined.

The responses were determined for 10.60, 10.62, 4.7, 4.4, 4.35, 3.25, 3.15, 2.85, 2.9, 3.17, 7.15, 4.73, 2.7, 3.29, 3.3 and 3.9 μm.

2.4 Development of Dilution System

2.4.1 General Schematic

This was based on a simple exponential dilution system, as described in section 1.2.2, with samples being measured by GC and FTIR. The GC system was used to provide reference measurements.

The initial set-up is shown in figure 2.4.1.

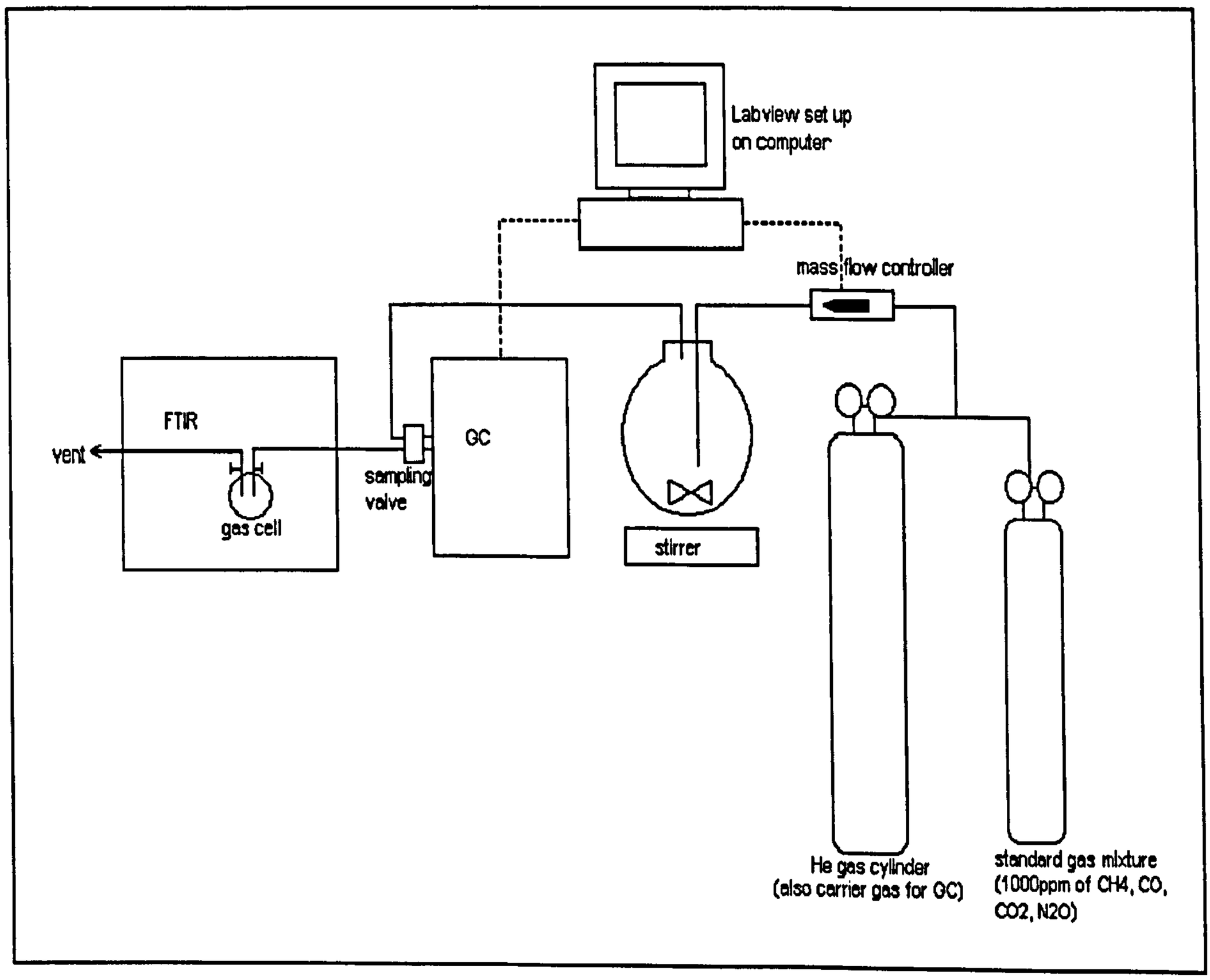


Figure 2.4.1 Initial dilution set-up

Equipment and reagents

- Gas Chromatography instrument (model 4300) from Carlo Erba Strumentazione, Italy, using TCD (thermal coupling detector) detection
- FTIR instrument from Bio-Rad (as described in section 2.1)
- Same gas cell as described in section 2.2
- Helium (used as diluent gas and for GC carrier gas) was supplied by Engineering and Welding Supplies Ltd.

- A standard gas mixture of 1000 ppm of each CH_4, CO, CO_2 and N_2O in nitrogen (certified as 1030, 1015, 989, and 1015 ppm respectively with ±2% certainty) supplied by Engineering and Welding Supplies Ltd. (from Linde Gas UK Ltd, Stoke-on-Trent, UK)
- Two more standard gas mixtures were later employed, also from Linde Gas UK Ltd., these were:
    - 1000 ppm of each CH_4, CO, CO_2 and N_2O in helium (certified as 1005, 1003, 993 and 999 ppm respectively, with ± 2 % certainty for CH_4, CO, and CO_2 and ± 5 % for N_2O)
    - 10,000 ppm each CH_4, CO, CO_2 and N_2O in helium (certified values 0.98 % CO, 0.95 % CO_2, 0.96 % N_2O, and 0.99 % CH_4 with ± 2 % certification accuracy for CO, CO_2 and N_2O and ± 5 % certified accuracy for CH_4)
- LabVIEW ™ version 5.0 with a PC-516/DAQ card ™
- Mass flow controllers: from Brooks Instrument B.V. model no. 5850 TR/DD1 B3DF) (calibrated against a bubble meter)
- A second flow meter incorporated in some of the later set-up was from MKS Instruments Inc. (1259B-001005V) (calibrated against a bubble meter)
- The gas cylinders were connected to a single mass flow controller *via* copper tubing and the remainder of tubing was plastic or teflon
- 2 l RB flask coated with Sigma Cote® supplied by Sigma Aldrich (to prevent the adsorption of gases onto the surface of the vessel), this included a magnetic stirrer to aid mixing

2.4.2 LabVIEW™ Programming

The original LabVIEW ™ (version 5.0 with a PC-516/DAQ card ™) program was written by Emery at University of Hull.[95] Figure 2.4.2 is the graphic diagram for the program, further details about the LabVIEW™ programming and wiring for the DAQ card can be found in the appendix (section 6).

On the DAQ card, channels 0 and 1 were wired to the mass flow controllers (referred to as controllers 1 and 2), channel 2 was for controlling the stirrer plate and channels 3 and 4 were connected to the GC. The program was designed to run through a sequence of

five steps. **Step 0**, shown in figure 2.3.2 in the box labelled '0[0…4]', checks the state of the controllers and stirrer, i.e. on or off.

The program could only be used to switch the mass flow controllers on and off, the actual flow rates had to be adjusted manually on the controllers.

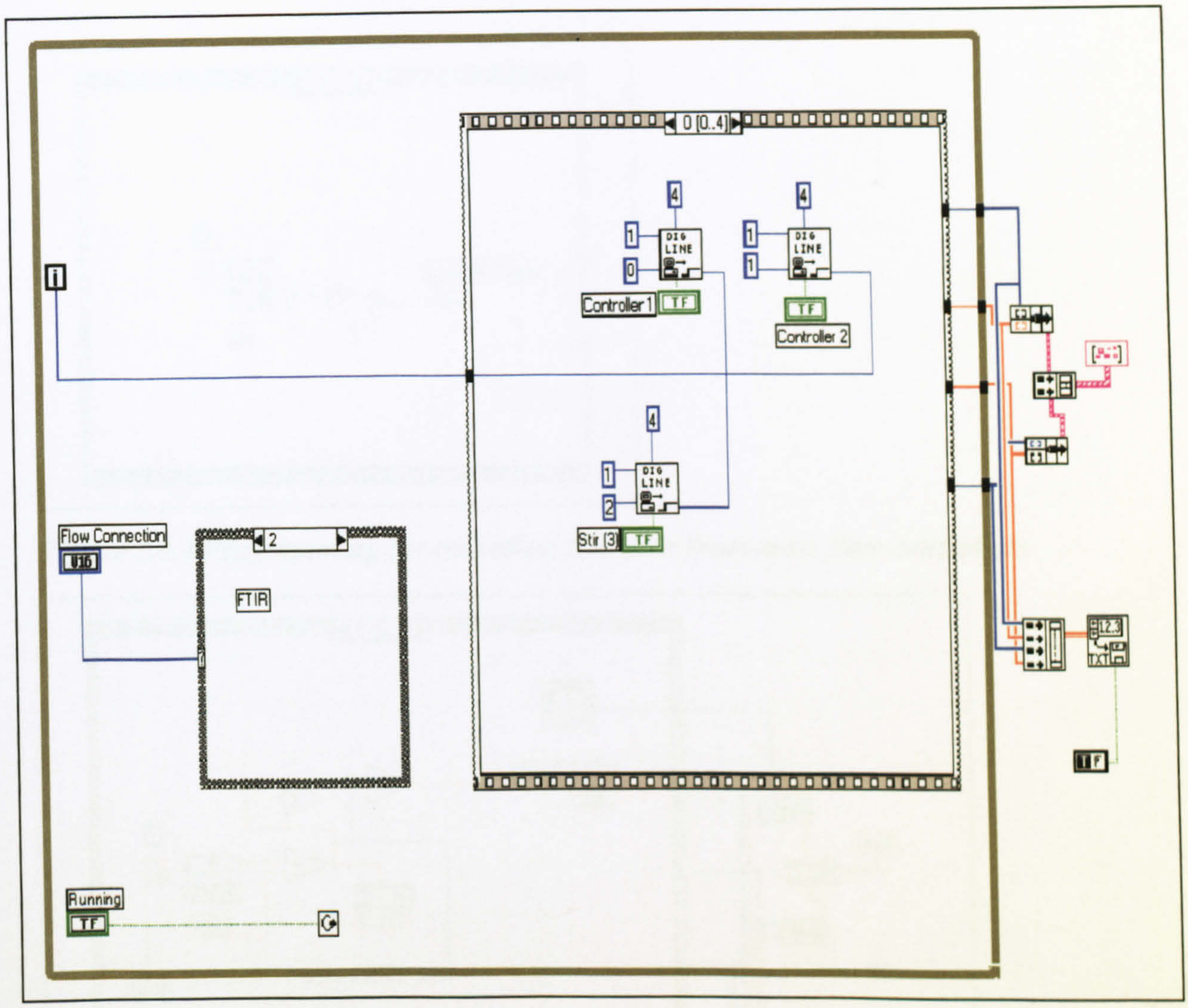


Figure 2.4.2 Diagram of LabVIEW™ programming

Step 1 acquires an immediate untimed single measurement from analog input channel 0, in this case the signal from the mass flow controller 1. In order to make the electrical signals relate to the actual flow rates the signal is multiplied by 200, rounded to the nearest whole number and divided by two. **Step 2** is the same process but acquires input from channel 1, which is wired to flow controller 2. Figure 2.4.3 shows LabVIEW™ diagram for this programming.

Steps 3 and 4, record output signals from the GC. The data is acquired as a waveform in which a specified number of samples are acquired at a specific sampling rate. The sampling rate was originally set at 1000 readings taken every second; from this 100

samples are selected. The samples that are collected as output are multiplied by 1000 to convert them to mV, the average of the samples is calculated and the data saved into a spreadsheet along with a time measurement. This programming is shown in figure 2.4.4.

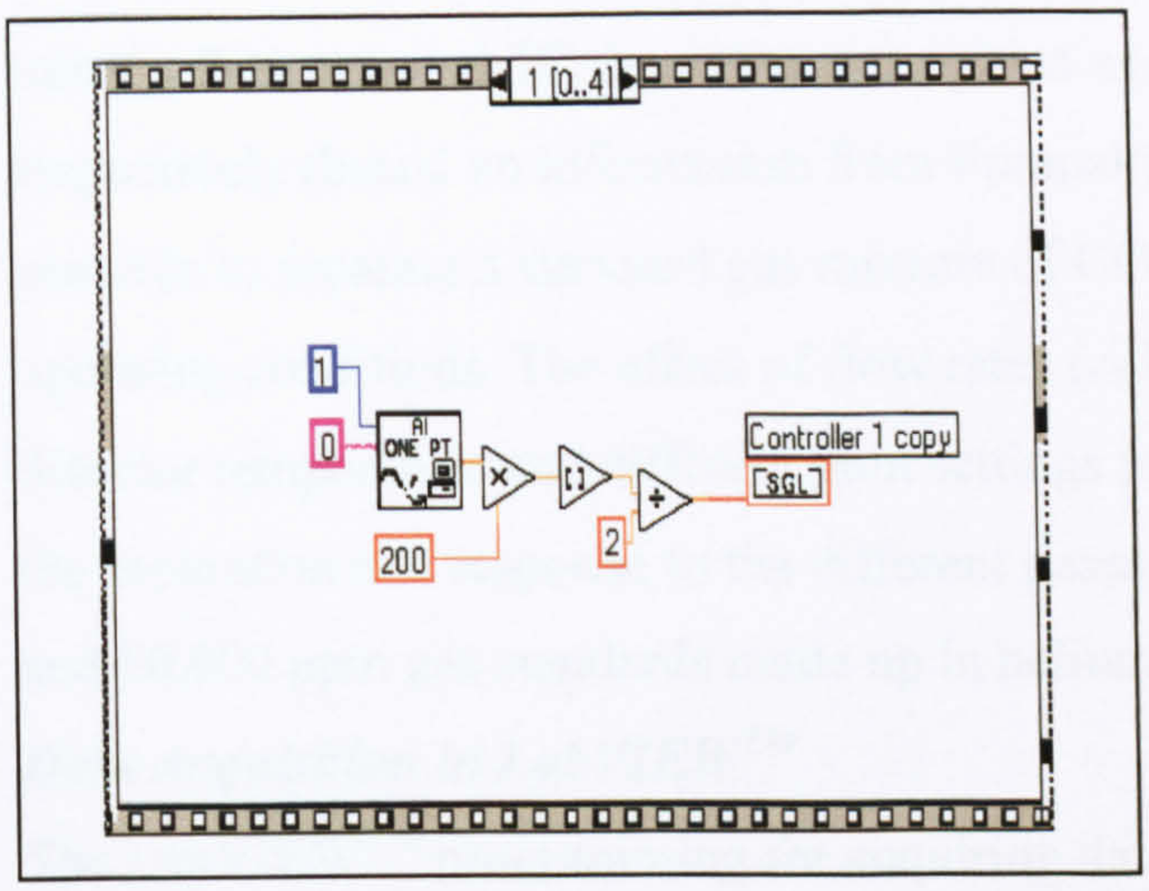


Figure 2.4.3 Programming for recording flow rate from mass flow controllers

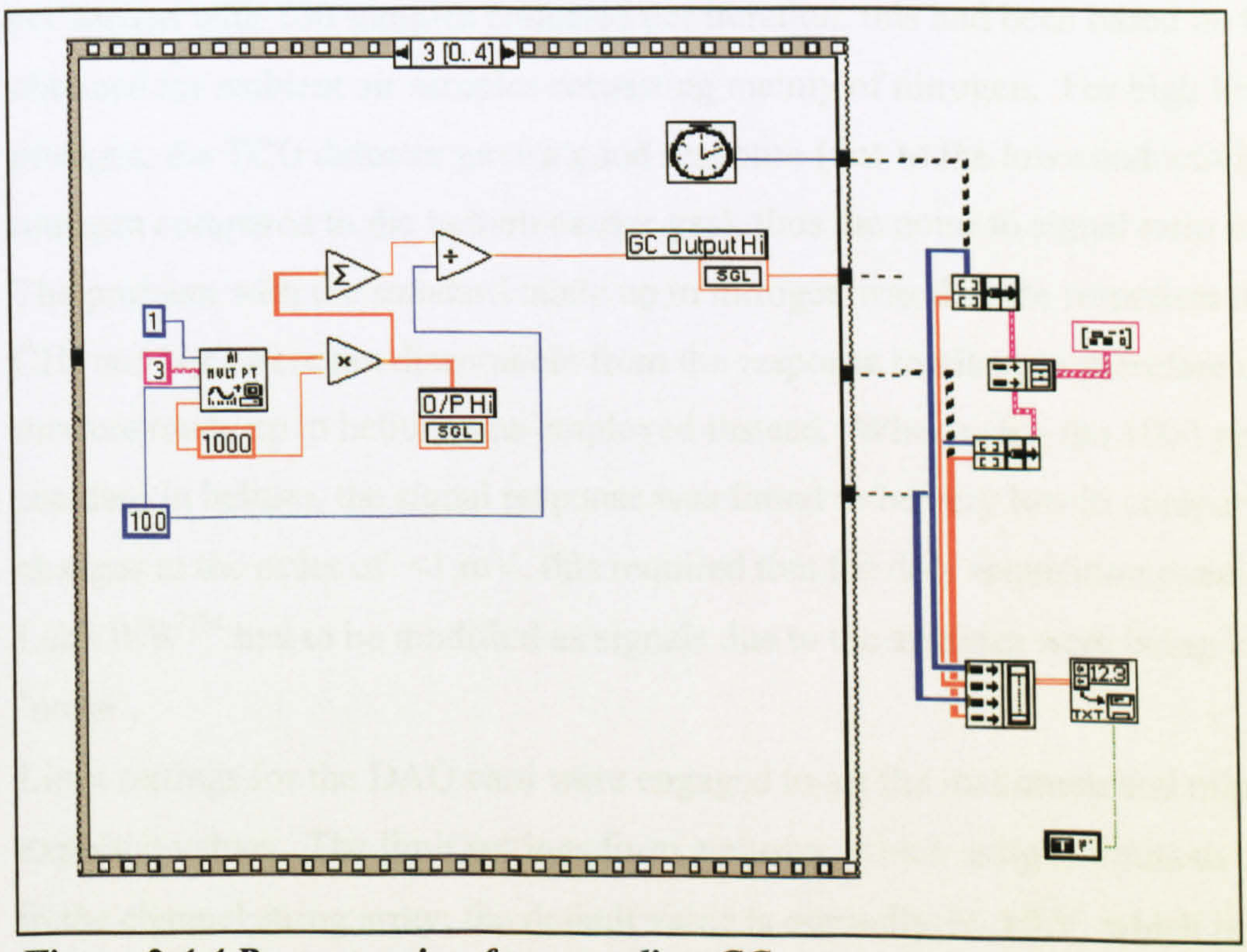


Figure 2.4.4 Programming for recording GC output

Subsequently the LabVIEW™ program was adapted such that the programming for recording the flow rate was removed and filtering and peak integration functions were added.

2.4.3 Optimising the Gas Chromatograph conditions

Detection and column

A 1 m x 2.3 mm diameter stainless steel column packed with Porapak Q^{96} was used, with helium as the carrier gas. Data for the retention times for CO, N_2, CO_2, N_2O and CH_4 on a 1 m x 2.3 mm Porapak Q column using a 30 °C operating temperature, 25 ml/min flow rate and TC detection were given as; 0.24, 0.24, 0.65, 0.76, 0.34 minutes respectively (based on information from Porapak). According to this data it should be possible to separate a standard gas mixture of CO, CO_2, CH_4 and N_2O using similar operating conditions. The effect of flow rates (column pressure), oven temperature, detector temperature and different gain settings (on both the GC and chart recorder), on the separation and response to the different gases was investigated using the 1,000 ppm and 10,000 ppm gas standards made up in helium.

Data Acquisition in LabVIEW™

The LabVIEW™ programming for acquiring data from the GC is described in section 2.4.2. Originally data was acquired as a waveform at a sampling rate of 1000 samples per second with 100 samples collected per iteration, this had been based on the response obtained for ambient air samples consisting mainly of nitrogen. For high levels of nitrogen, the TCD detector gave a good response (due to the low conductivity of nitrogen compared to the helium carrier gas), thus the noise to signal ratio was low. The problem with the standard made up in nitrogen was that the responses to CO, CO_2, CH_4 and N_2O were not discernable from the response to nitrogen therefore a standard mixture made up in helium was employed instead. When using the 1000 ppm gas standard in helium, the signal response was found to be very low in comparison, with changes in the order of <1 mV, this required that the data acquisition conditions in LabVIEW™ had to be modified as signals due to the analytes were being lost in 'noise'.

Limit settings for the DAQ card were engaged to set the maximum and minimum expected values. The limit settings form a cluster, which assigns limits to the channels in the channel string array; the default value is normally +/- 10 V, which in this case meant limited sensitivity. The gain (amplification or attenuation) is determined by the programming based on the limit settings.

The rate of sampling and the number of samples acquired by the LabVIEW™ program were both altered in order to find the optimum conditions. Due to the low response to the gas sample mixture, a small amount of ambient air was let into the system so that

the nitrogen peak could be used as an indication of whether the signal to noise ratio was improved. A median filtering function was also included into the LabVIEW™ program.

Peak Integration

To calculate the peak areas, peak detection and integration functions (from the LabVIEW™ software) were incorporated into the LabVIEW™ program.

The peak detection function identifies peaks above a particular threshold with a specified peak width (or below the threshold if looking at negative peaks). The outputs are the number of peaks detected, the peak indices and the amplitude.

Reproducibilty of response

Reproducibility of response was tested using an undiluted 0.1 % standard mixture in helium; the peak areas were calculated in Excel. This was performed for different column pressures (thus changing the He flow rates) of 140, 160, 200 and 260 kPa. The standard deviation between peak height and peak areas were then calculated and compared.

2.4.4 FTIR

IR spectra were collected as single beam spectra, using the range 400-4200 cm^{-1}, with a resolution of 4 cm^{-1} resolution with 8 scans co-added. The data was saved as .csv files allowing them to be analysed in Excel and MATLAB®.

2.4.5 Different dilution set-ups

Trial Run Dilution System (using standard made up in nitrogen)

The original set-up (figure 2.4.1) was tested using the 1000 ppm mixture in nitrogen, the sampling chamber was filled with the standard mix at a flow rate of 30 cm^3min^{-1} for 10 minutes, this was diluted with helium using a flow rate of 10 cm^3min^{-1} for 90 minutes, this was later increased to a flow rate of 20 cm^3min^{-1} for 120 minutes. Samples were analysed by GC using the gas-sampling valve (with a 5 cm^3 sample loop), with responses recorded both in LabVIEW™ and with a chart recorder, and FTIR spectra were collected at timed intervals (using single beam energy scan range 0-8000 cm^{-1} with a resolution of 2 cm^{-1}).

Dilution with 1% standard mixture in helium

Set-up 2 shown in figure 2.4.5 was employed. This 2 L mixing chamber was removed (thus reducing time required to purge system) and instead the gas cell was utilised as the

mixing chamber. The gas cell was filled with the 1 % standard gas mixture and then diluted with helium using a 10 cm^3min^{-1} flow rate with samples taken approximately every 6 minutes (5 cm^3 sample loop). This was repeated three times and the GC data was analysed (employing a GC column pressure 200 kpa). The response was recorded in LabVIEW™ using a sampling rate of 600 s^{-1} with 60 samples taken as average and median filtering with a window size of 40 was also applied

The effect of stopping dilution while collecting IR data

As previously, set-up 2 was employed and the gas cell was filled with 1 % standard gas mixture. In this case the helium was added to the cell for a specific amount of time then the helium flow stopped while GC and IR data was collected.

The dilution times between each sample were at: 1, 1.5, 2.5, 5, 5, 10, 15, and 20 minutes.

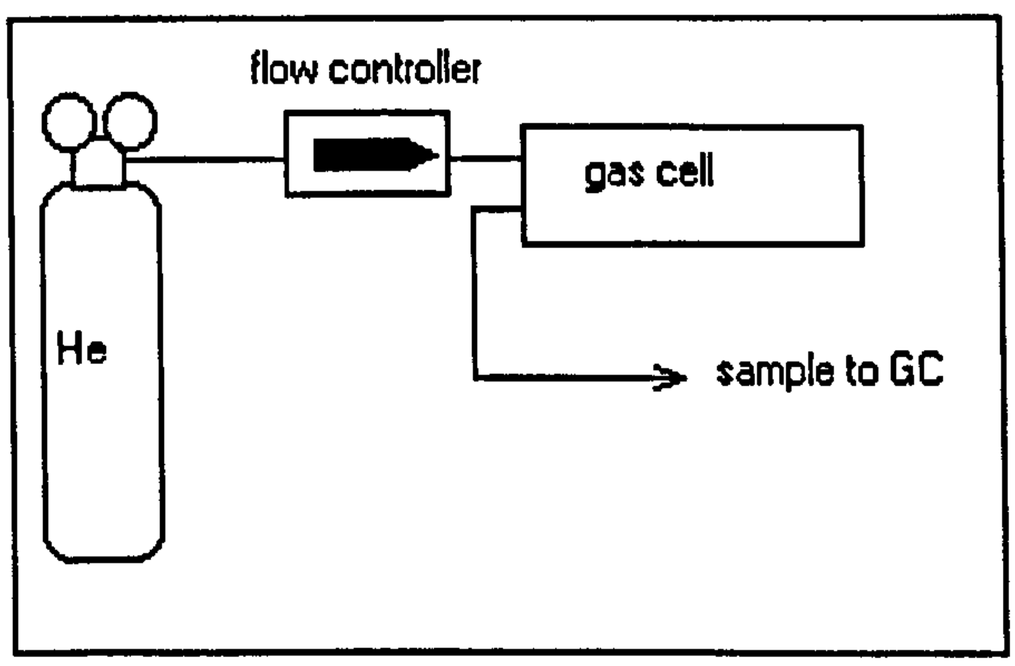


Figure 2.4.5 Set-up 2, dilution of samples made up in gas cell

The effect of bleeding SF_6 into the system

This work investigated the effect of adding SF_6 to this system. The SF_6 was bled in at a fixed rate using a second flow controller. The set-up is shown in figure 2.4.6. IR spectra were collected at different intervals to see if a constant concentration of SF_6 could be maintained.

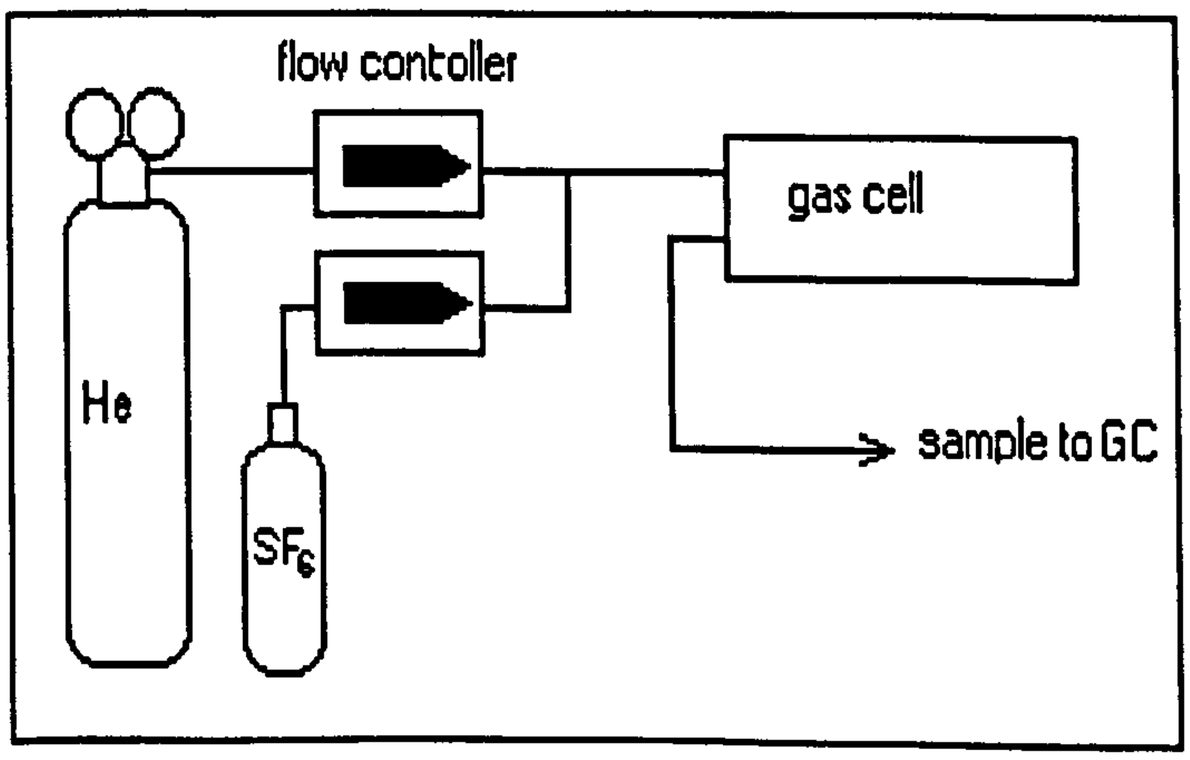


Figure 2.4.6 Set-up for bleeding SF_6 into the system

2.4.6 Correlation between GC and IR measurements

Set-up 2 was employed for this work (figure 2.4.5), samples were prepared in the long path gas cell and spectra were collected by FTIR. The concentrations of gases prepared in each sample are shown in table 2.4.1.

After scanning by FTIR each sample was then diluted with helium using a 10 cm^2min^{-1} flow rate, this was repeated twice for each sample. From sample L onwards, spectra were also collected using the 4.35 μm and 3.25 μm filters. In an effort to reduce the amount of air getting into the gas cell during sample preparation, the gas cell outlets were plugged with rubber bungs and injections of each gas was made through them. Plots relating the GC response (peak heights) to IR response (this was the transmission at wavenumber relating to the gases of interest, or peak area in the case of spectra collected using the filters) were generated and the correlation coefficients determined.

Concentration ppm of gases in each sample				
Sample	**N_2O**	**CO**	**CH_4**	**CO_2**
A	1886.79	1886.79	1886.79	1886.79
B	188.68	188.68	188.68	188.68
C	1132.08	188.68	188.68	754.72
D	188.68	943.40	188.68	943.40
E	1886.79	188.68	1886.79	188.68
F	188.68	1886.79	1886.79	1886.79
G	943.40	943.40	188.68	943.40
H	943.40	943.40	905.66	188.68
I	1886.79	188.68	1886.79	1886.79
J	566.04	1320.75	377.36	1509.43
K	377.36	377.36	1320.75	377.36
L	1886.79	1509.43	377.36	566.04
M	1509.43	1886.79	188.68	1509.43
N	377.36	1509.43	1509.43	377.36
O	188.68	566.04	1509.43	1509.43
P	377.36	566.04	377.36	1320.75
Q	566.04	1698.11	1698.11	566.04
R	1698.11	1698.11	1698.11	188.68
S	943.40	1886.79	1886.79	943.40
T	1698.11	377.36	377.36	1886.79
U	1886.79	188.68	188.68	188.68
V	188.68	1509.43	377.36	377.36
W	1886.79	377.36	943.40	1509.43
X	1698.11	1698.11	943.40	377.36
Y	188.68	943.40	1509.43	1509.43

Table 2.4.1 Concentrations of gases (ppm) added to each sample (for correlating GC and IR responses)

2.5 Characterisation of narrowband filters by FTIR

2.5.1 Characterisation of 3.25 μm and 4.35 μm filters

Originally just two narrowband filters from Northern Optical Coatings (NOC), Northumberland, UK, were tested. One with band centre at 3.25 μm (3077 cm^{-1}) and one centred at 4.35 μm (3000 cm^{-1}) with a bandwidth of 1 %.

The filters were placed into lens holders and placed one at a time next to the gas cell in the sample chamber of the FTIR.

Calibrations with samples prepared statically

Gas samples were prepared by injecting different amounts (between 1 and 30 ml) of the 1% standard mixture to give sample concentrations between 18 and 566 ppm. FTIR single beam scans (using the same conditions as previously in section 2.3) were taken. The total response of light transmitted when the filters were in place was calculated by WinIR software using peak templates (this function calculated the area between two specified wavenumbers using a baseline determined from two other wavenumbers). Initially each sample prepared was measured with both filters (first one and then switching to the other) but the continuously swapping round of the filters meant variations arose due to changes in alignment, subsequently separate samples were prepared to test each filter.

Calibrations using the dilutions

The gas cell was filled with 1 % standard mix and diluted with helium using a 10 cm^3min^{-1} flow rate and again repeated with a 20 cm^3min^{-1} flow rates. This was done for both filters.

Scans were taken at timed intervals using the same conditions as described in section 2.3.

2.5.2 Using filters at 2.9 μm, 3.25 μm, 4.35 μm, 4.73 μm and 10.62 μm

Three more filters at 2.9 μm, 3.73 μm and 10.62 μm were employed (also from NOC). The response obtained with different filters was measured using samples prepared by injecting different amounts of the 1% standard (the amounts injected were 20, 50, 60, 80, 100, 120, 150 and 200 ml, corresponding to final concentrations of 377, 943, 1132, 1509, 1887, 2264, 2830 and 3774 ppm)

The total response, i.e. amount of light detected with filter in place, was calculated in the Bio-Rad WinIR software using peak templates.

Different data treatment was applied to the responses; this included taking logs and subtracting the response obtained for the 10.62 μm filter (reference peak).

Variation in response due to filter alignment

The variation in response due to realigning filters was investigated using the 2.9 μm filter by measuring the same gas sample. Initially the sample was scanned, then measured again after a 1-minute and a 10-minute interval (without moving the filter). The filter was moved and realigned and the sample measured again, this was then was repeated a second time. The filter was subsequently placed in a different filter holder and the sample measured, this was also repeated switching the sample back to the original holder.

Automated Filter Wheel

The filters were placed in a 6 position automated filter wheel (figure 2.5.1), the 4.35 μm filter was placed in position 1, 4.73 μm in position 2, 10.62 in position 3, position 4 was left blank (allowing normal scans to be taken when the wheel was in this position), 2.9μm in position 5 and 3.25 μm in position 6. The wheel was placed in the sample chamber of the FTIR next to the gas cell as shown in figure 2.5.2. The wheel could be controlled manually (with a push button) or via computer.

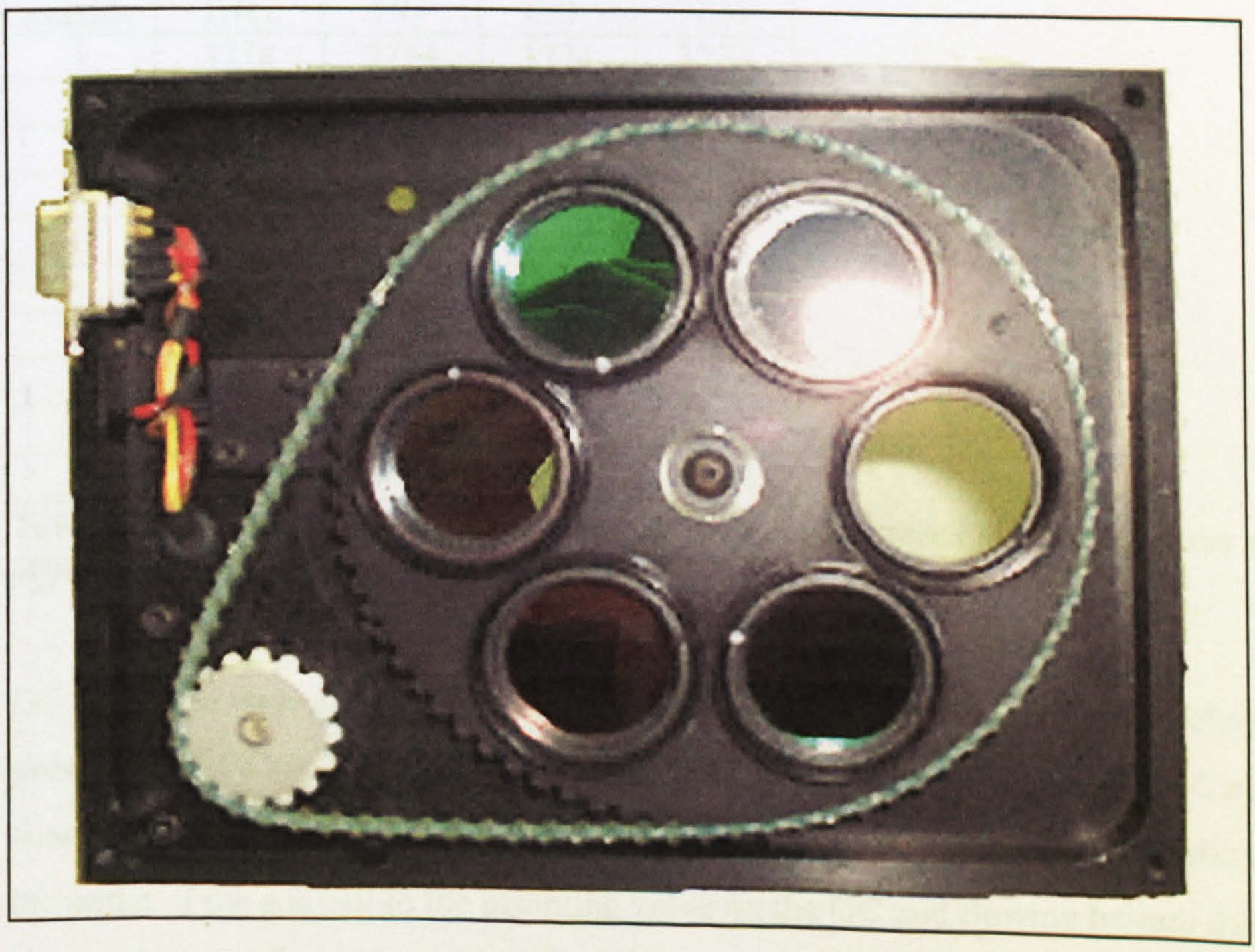


Figure 2.5.1 Inside view of filter wheel with narrow band filters in place

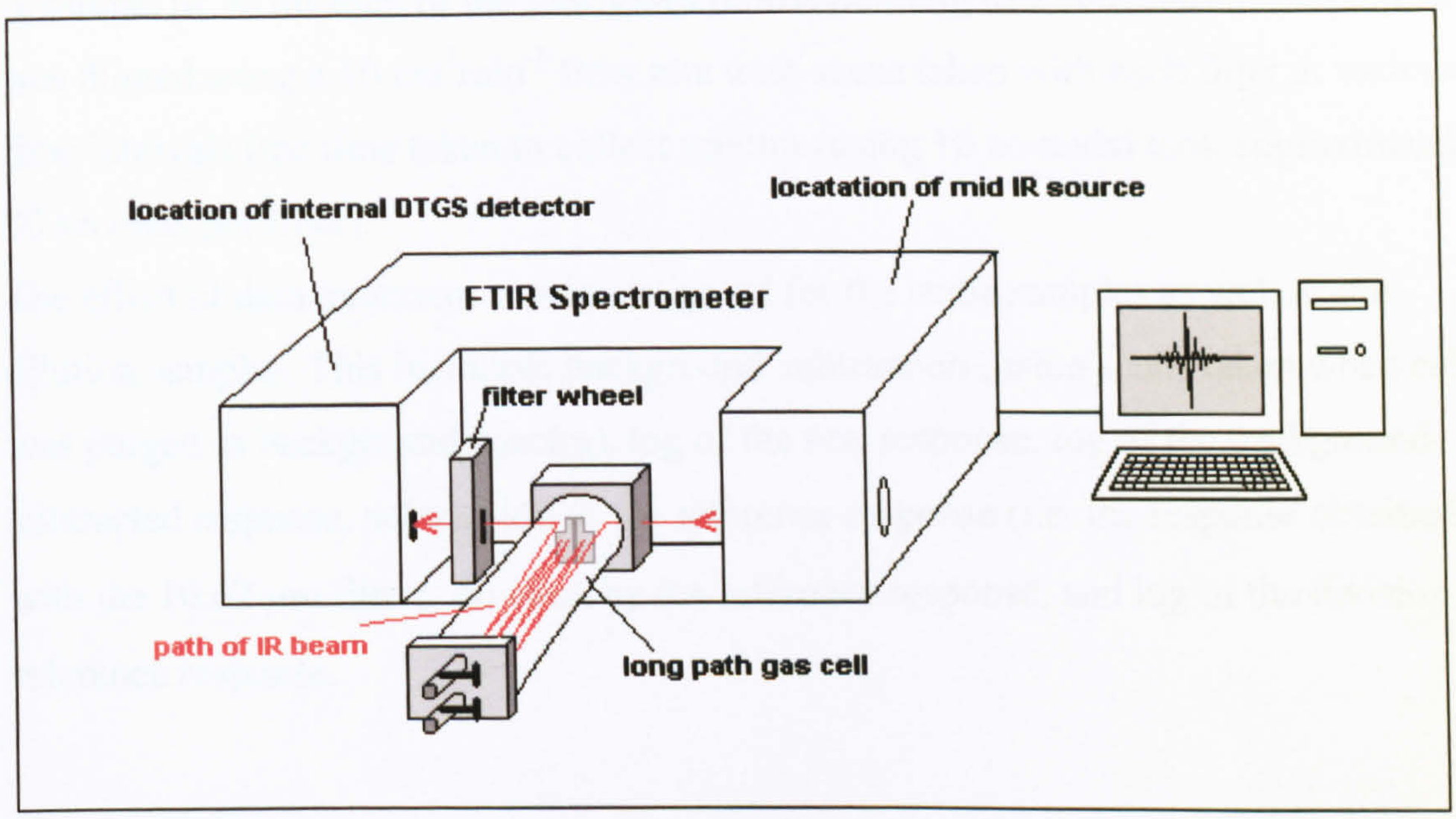


Figure 2.5.2 Diagram showing placement of Filter wheel in FTIR

Sample mixtures of the four gases (CO, CO_2, CH_4 and N_2O) were prepared in the same manner as described in section 2.1. Table 2.5.1 lists the concentrations of each gas in the different samples.

Concentrations in ppm of gases in each samples				
sample	**CH_4**	**CO**	**CO_2**	**N_2O**
1	3774	3774	3774	3774
2	18868	18868	18868	18868
3	9434	3774	13208	5660
4	3774	1887	18868	16981
5	1887	15094	1887	1887
6	18868	18868	3774	18868
7	3774	3774	1887	15094
8	16981	3774	3774	3774
9	15094	1887	15094	1887
10	-	9434	9434	9434

Table 2.5.1 Concentrations of each gas for samples prepared to test responses with different filters

GC responses were also collected for each sample (using GC conditions of column pressures at 200 kPa, oven temperature at 25°C, detector temperature at 200°C and filament temperature at 250°C); these were sampled by connecting (with plastic tubing) the outlet of the gas cell to the sampling valve on the GC and flowing helium through the cell (using a flow rate of 10 cm^3min^{-1}) to ensure the sampling valve was filled with sample before injecting into the GC.

A sample of 50 ml each of the four gases (corresponding to a concentration of 9.4 %) was diluted using a 10 cm^3min^{-1} flow rate with scans taken with each filter at various time intervals (the time taken to collect spectra (using 10 co-adds) took approximately 30 seconds per filter).

The effect of data treatment was investigated for the static samples as well as the dilution samples. This included; background subtraction (using scans taken when cell was purged as background spectra), log of the raw response, log of the background-subtracted response, subtraction of the reference response (i.e. the response obtained with the 10.62 μm filter), division by the reference response, and log of the division by reference response.

2.6 Investigations into different optical layouts and signal processing

The aim of this section was to investigate different optical layouts using the available equipment to determine an optimal set-up.

Equipment

- Mid-IR ceramic air-cooled source (PERMAGLOW^{TM}) from Bio-Rad
- InSb detector from Hamamatsu Photonics K.K., Japan
- MCT (mercury cadmium telluride) detector from Bio-Rad (range 10,000 -700 cm^{-1})
- Peltier-cooled DTGS (deuterated trigylcine sulfate) detector from Bio-Rad
- Pre-amplifier from Hamamatsu Photonics K.K.
- Lock-in-amplifier from Thors, Model LIA100, programmable AC input amplifier, a monolithic phase-sensitive detector (PDS), and two-stage programmable output low-pass filter (with adjustable settings)
- Chopper from Bentham Instruments Ltd., Berkshire, UK
- Farnell adjustable Power Supply, 10 V, 0.6 amp)
- CaF (bi-concave lens) lens

Wiring and connection box for detectors

The MCT and DTGS detectors had 9 pin connectors for use with the BioRad FTIR. To enable the detector responses to be sent to the lock-in amplifier, the cable connecting the detectors to the FTIR was adapted so that the signal was also sent to a coaxial fitting.

The InSb detector was wired to a coaxial cable so that it could be connected directly to the lock-in amplifier. To enable the InSb signal to be processed by the FTIR a connection box was built to connect the coaxial cable to the 9-pin cable. More details about the wiring can be found in the appendix (section 6).

Optimisation of lock-in amplifier settings

The effect of different settings on the lock-in amplifier was investigated to determine which settings gave the best results. This included; investigating different chopper speeds (no chopper, and speeds of 5 Hz, 124 Hz and 380 Hz), changing the frequency range (5-10 Hz, 150-300 Hz, 300-600 Hz), the sensitivity (the lock-in had a 6-position rotary switch to set the gain of the input, (0, 20, 40, 60, 80, 100 dB) as well as a x10 gain) and time output constant. The main criteria for a good result being; stable signals

(i.e. not much drift or fluctuation) and different responses with different filters (NB. the emission was from a black body source, therefore there was a higher amount of emission around 4 μm of light while much lower response for 10 μm (in the case of the InSb detector, 10 μm is beyond the cut-off point so no response would be expected with the 10.62 μm filter).

Different Optical Set-ups

Set-up (1) Employing the InSb detector with a multimeter to measure detector response

A schematic for set-up 1 is shown in figure 2.6.1.

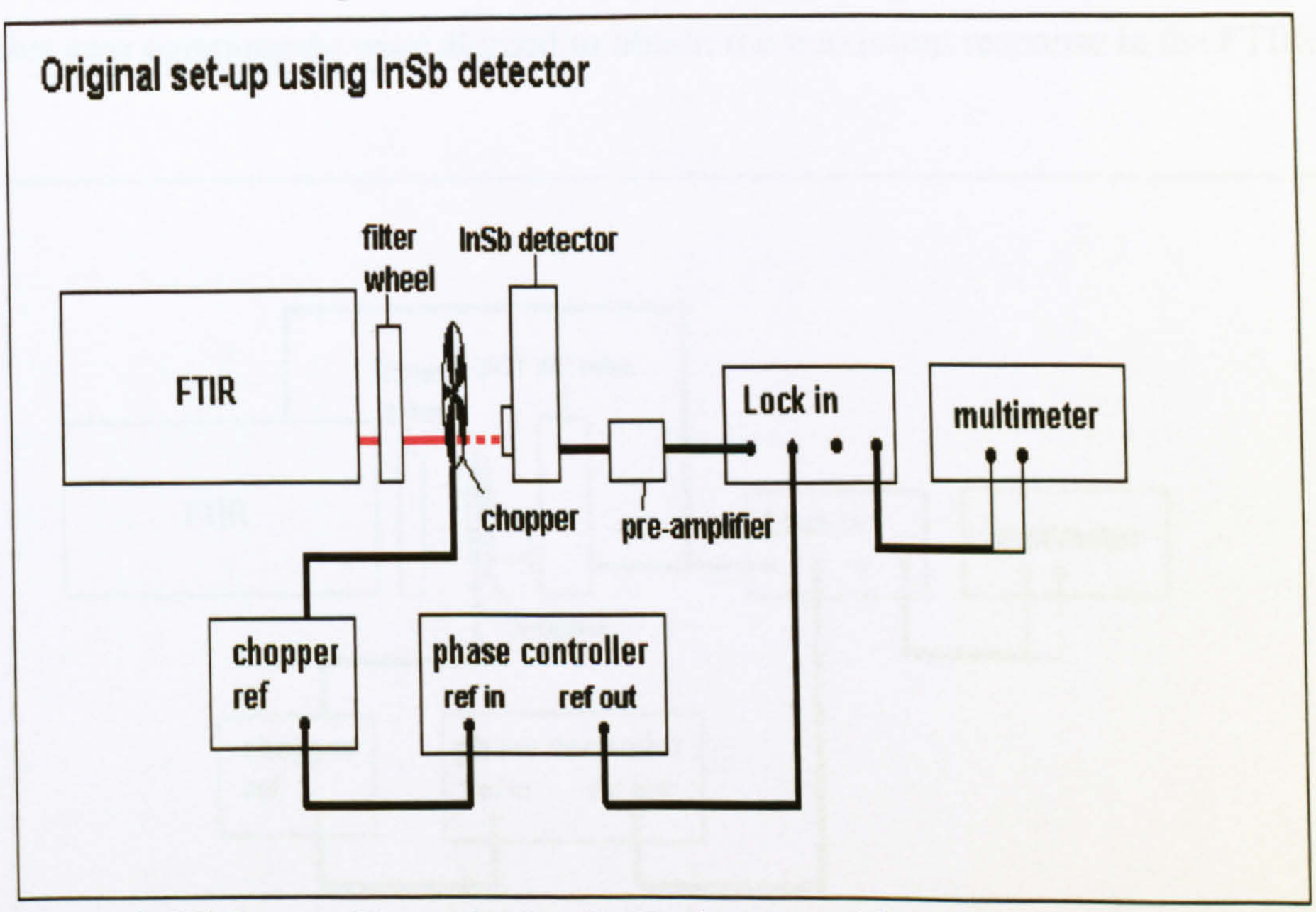


Figure 2.6.1 Set-up (1) employing InSb detector with multimeter to measure response

The components were aligned by centring the He laser beam (from the FTIR provided for alignment purposes) on the detector window. The infrared beam passes from the external right window of the FTIR through the selected filter (held by the filter wheel) to the detector. The signal is increased with a pre-amplifier (which had three different settings of low, medium and high) before reaching the lock-in amplifier. A chopper (placed between the filter wheel and the detector) is used to act as a signal reference. The output (V) from the lock-in amplifier is shown on the digital display on the multimeter. The different settings on the pre-amplifier (low, medium and high), changing the chopper speed, and altering settings on the lock-in amplifier (i.e. frequency range, sensitivity and output time constant) were applied to try and optimise the response.

Set-up (2) using MCT detector

The cable for connecting external detectors the FTIR instrument was rewired so that the signal from the detector was also sent to the lock-in amplifier, this meant the MCT detector (which gave a higher signal response) could replace the InSb detector and simultaneous measurements could be made with the FTIR. The corresponding values from the FTIR were from the total response of the interferrogram in the set-up mode of the software. When obtaining spectra with the FTIR the chopper had to be switched off as it interfered with the interferrogram signal. The set-up is shown in figure 2.6.2; in this case components were aligned to obtain the maximum response in the FTIR.

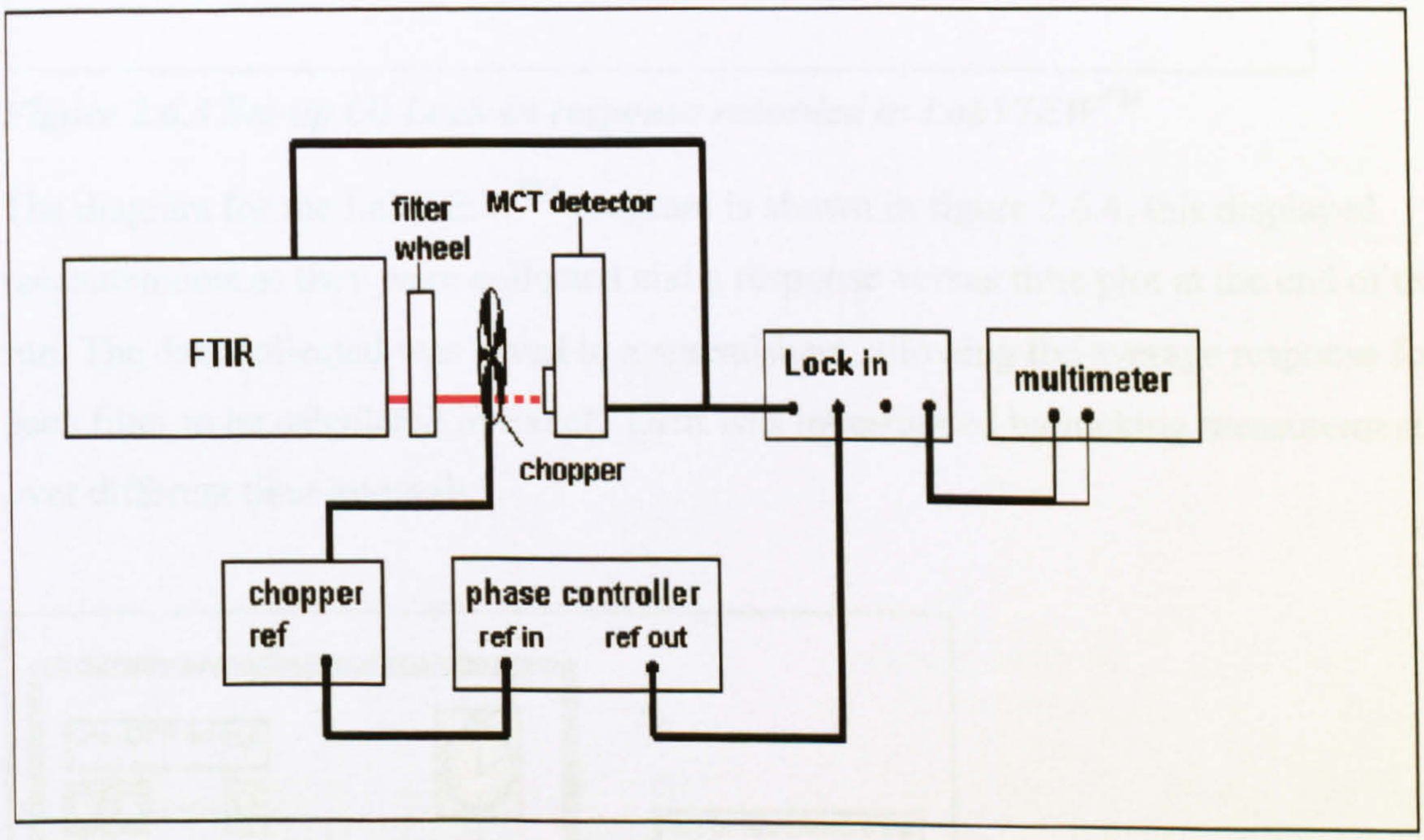


Figure 2.6.2 Set-up (2) using MCT detector with multimeter to record the output

Set-up (3) using MCT detector and LabVIEW™

Set-up (3) is shown in figure 2.6.3. The output from the lock-in amplifier was wired to a DAQ card (same card that was used for recording GC measurements as in section 2.3) and a simple LabVIEW™ program was written to record the output as an analogue signal.

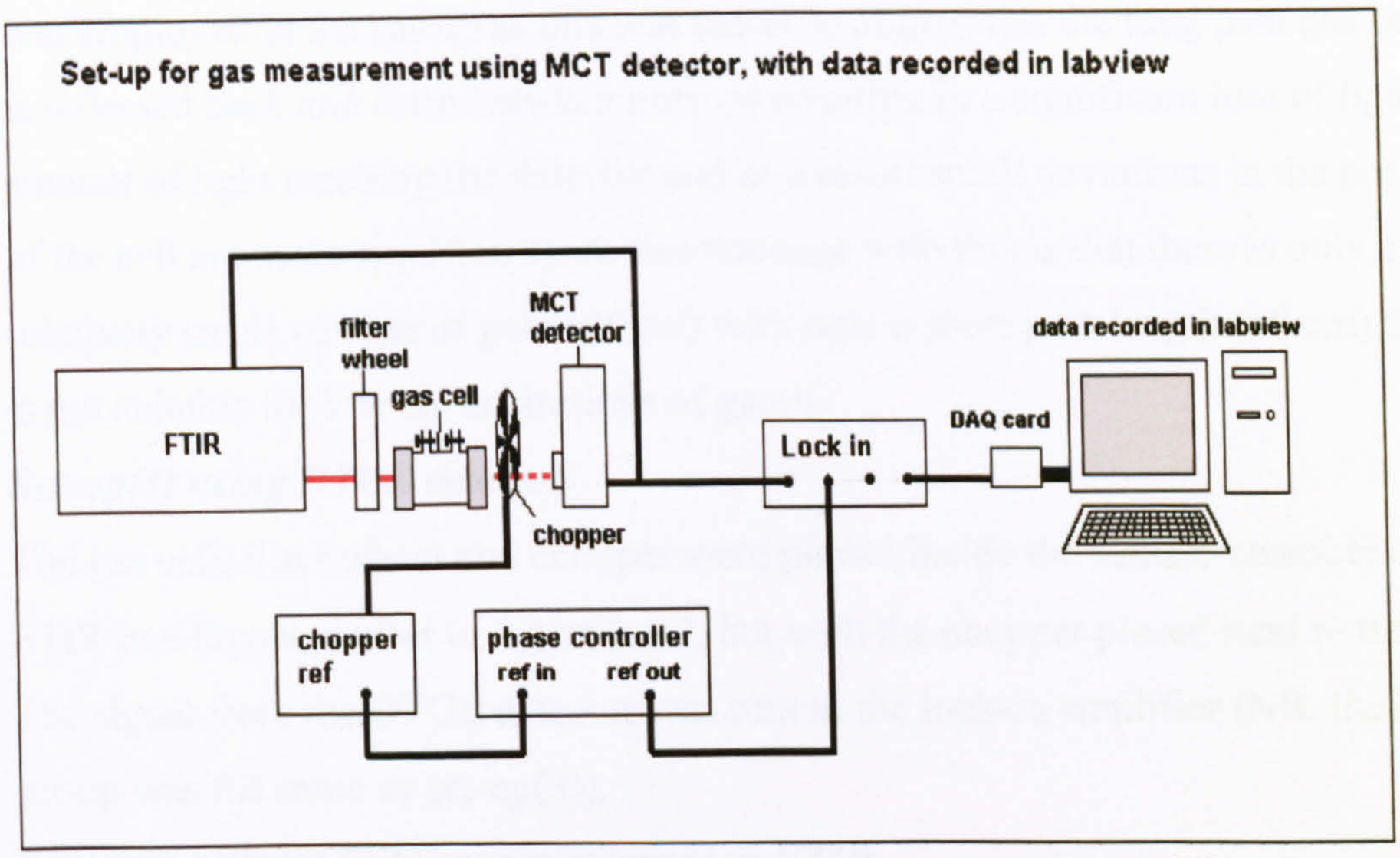


Figure 2.6.3 Set-up (3) Lock-in response recorded in LabVIEW™

The diagram for the LabVIEW™ program is shown in figure 2.6.4, this displayed measurements as they were collected and a response versus time plot at the end of the run. The data collected was saved to a spreadsheet, allowing the average response for each filter to be calculated in Excel. Drift was investigated by making measurements over different time intervals.

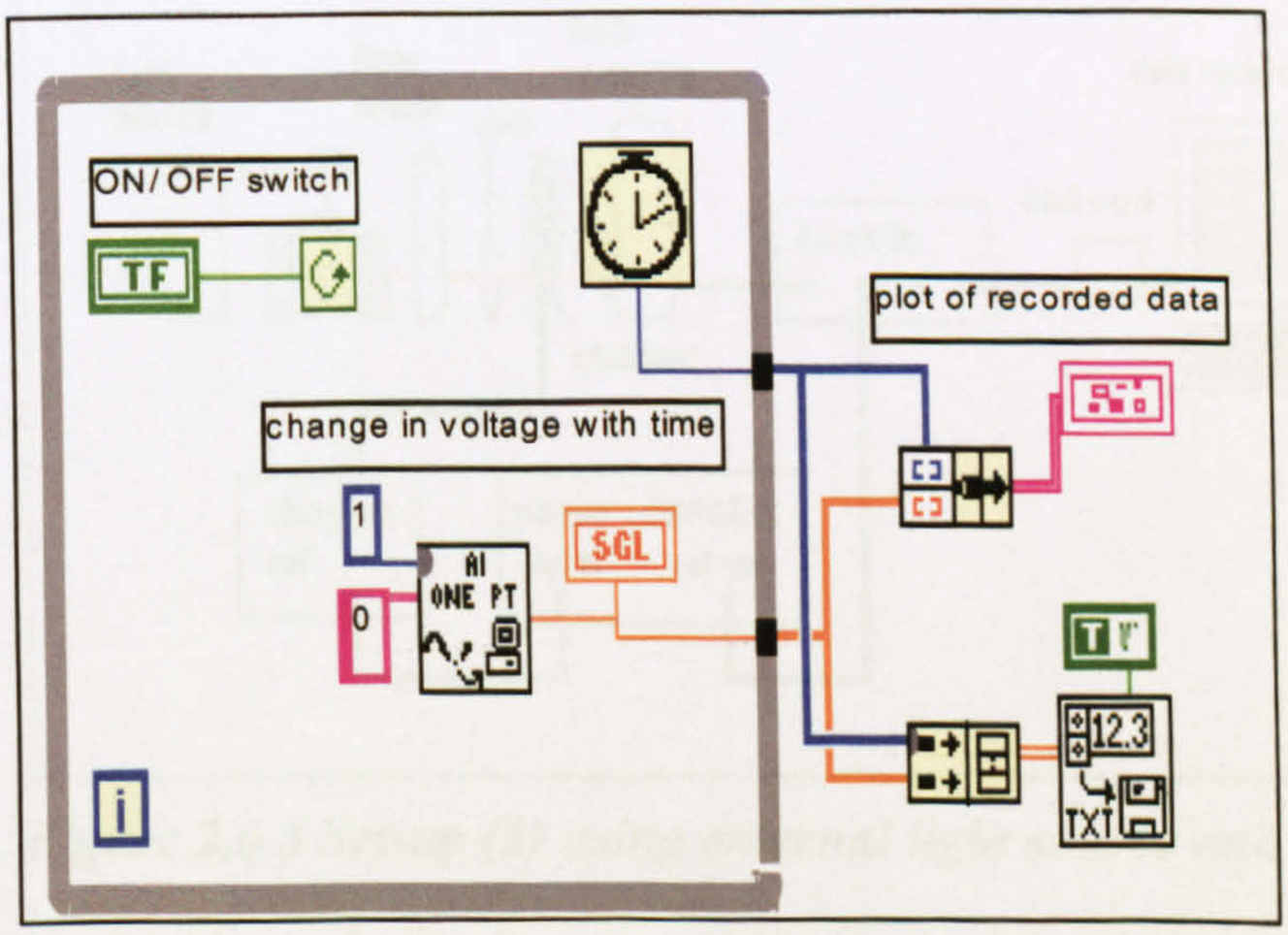


Figure 2.6.4 LabVIEW™ program for recording response from lock-in amplifier

A small 100 ml gas cell (with 10 cm path length) was placed in the set-up and different amounts of gases where injected into the cell to check if the detector response obtained with different filters varied in relation to the gases injected. The small (100 ml) gas cell

was employed in the set-up as this was easier to align (with the long path gas cell light is reflected back and forth between mirrors resulting in a significant loss of light in the amount of light reaching the detector and as a result small deviations in the positioning of the cell are more significant). A disadvantage with this is that there is only a relatively small volume of gas (100 ml) with only a short path length (10 cm) therefore is not suitable for low concentrations of gases.

Set-up(4) using DTGS detector

The gas cell, filter wheel and chopper were placed inside the sample chamber of the FTIR in a layout similar to figure 2.6.2, but with the chopper placed next to the gas cell. The signal from the DTGS detector was sent to the lock-in amplifier (NB. the lock-in set-up was the same as set-up(3)).

Set-up(5) using a light source external to FTIR

In this case the IR source was removed from FTIR and used externally, otherwise the set-up was similar to set-up (3). Since the IR beam from the source no longer passed through an interferometer the detector response was no longer recorded by FTIR. A lens was placed between the filter wheel and chopper to focus the light on to the detector; this set-up is shown in figure 2.6.5.

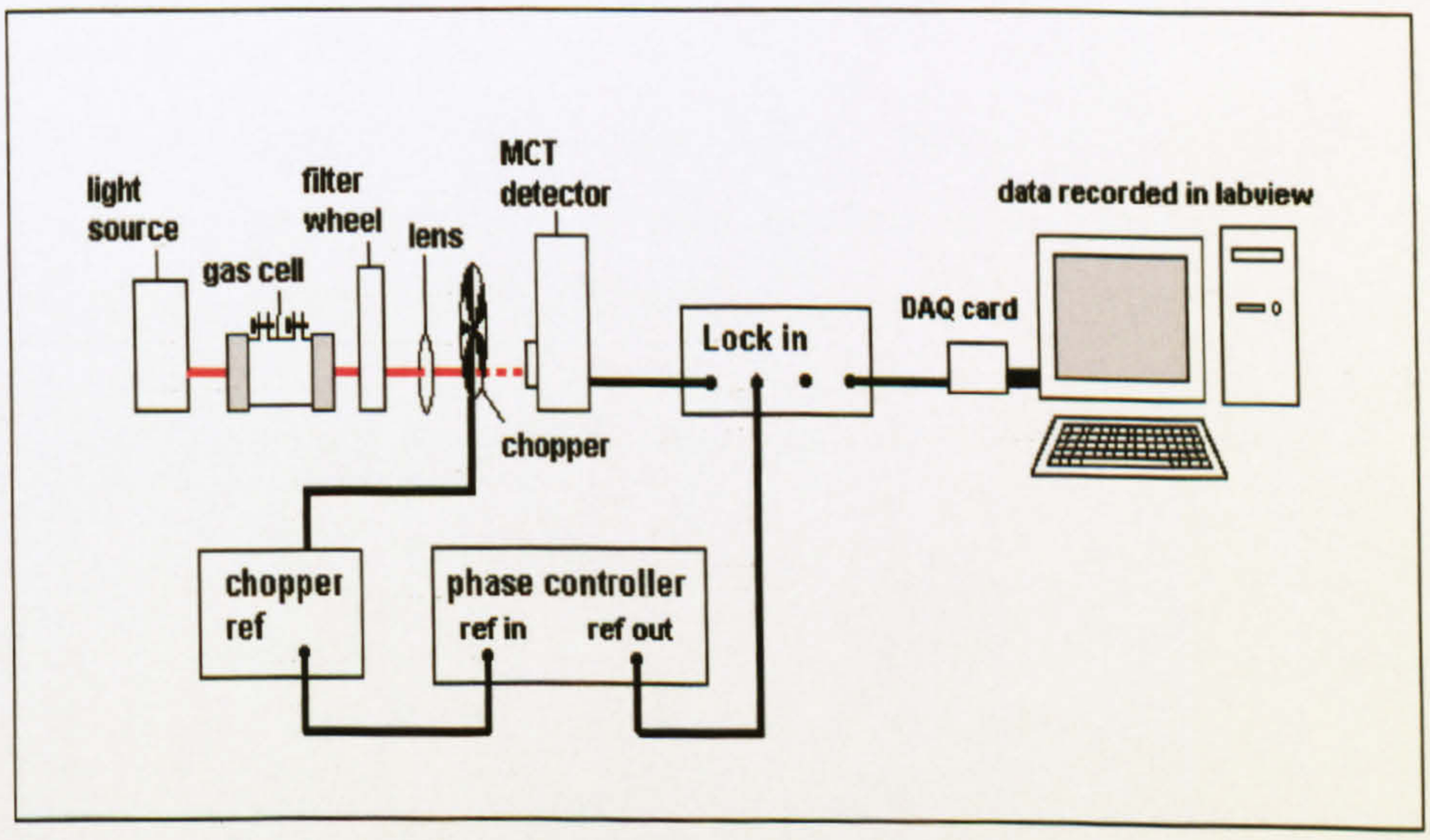


Figure 2.6.5 Set-up (5) using external light source and including a lens

Sample Dilutions with set-up (5)

Set-up (5) gave the best results therefore further work was carried out to investigate responses with this set-up. The 100 ml gas cell was replaced with the long path gas cell employed in previous work. A sample consisting of 9.4 % each of CH_4, CO, CO_2 and

N_2O was prepared and diluted with helium using a 85 cm^3min^{-1} flow rate for 1 ½ hours with measurements made every 2-5 minutes.

This was repeated using an initial sample mixture of 9.4 % CH_4, 9.4 % CO, and 7.5 % CO_2, this time diluted at a 60 cm^3min^{-1} He flow rate. This was also performed twice using initial sample concentrations of 7.5 % each of CH_4, CO and CO_2 with a flow rate of 10 cm^3min^{-1}.

The data collected was analysed to look at the sensitivity of the measurements and reproducibility. Data treatment applied included subtraction of purge responses, subtraction of the reference response (10.62 μm, for samples with no N_2O present the 2.9 um filter was also tried as a reference response), division of reference response and logs.

2.7 The purgable box

2.7.1 Design and development of a purgable box

To remove the effects of ambient air, in particular CO_2, it was decided to build a black sealed box that could be purged.

The main considerations were:

- The optical alignment of the components
- Inlet and outlets to allow the box to be purged
- Inlet and outlets for the gas cell to allow sample preparation and purging of gas cell without requiring the box to be opened
- Cables and connections for the source, filter wheel, chopper and detector (all requiring connections to power sources)

Based on the results of the previous section (2.6), set-up (5) gave the optimal results, hence this alignment was used as the basis for the design of the box. Figure 2.7.1 shows the basic schematic for the box.

A rectangular box (although this resulted in a lot of empty space (increasing the amount of air that required purging) it was the simplest design to construct) was constructed of a black vinyl material; this had the advantage of minimising the amount of stray light entering the system.

The optical components were aligned and mounted on a removable metal plate. A gas cell holder (as used in the FTIR instrument) held the gas cell in place, with rubber tubing connecting the cell outlets to the external outlets on the box.

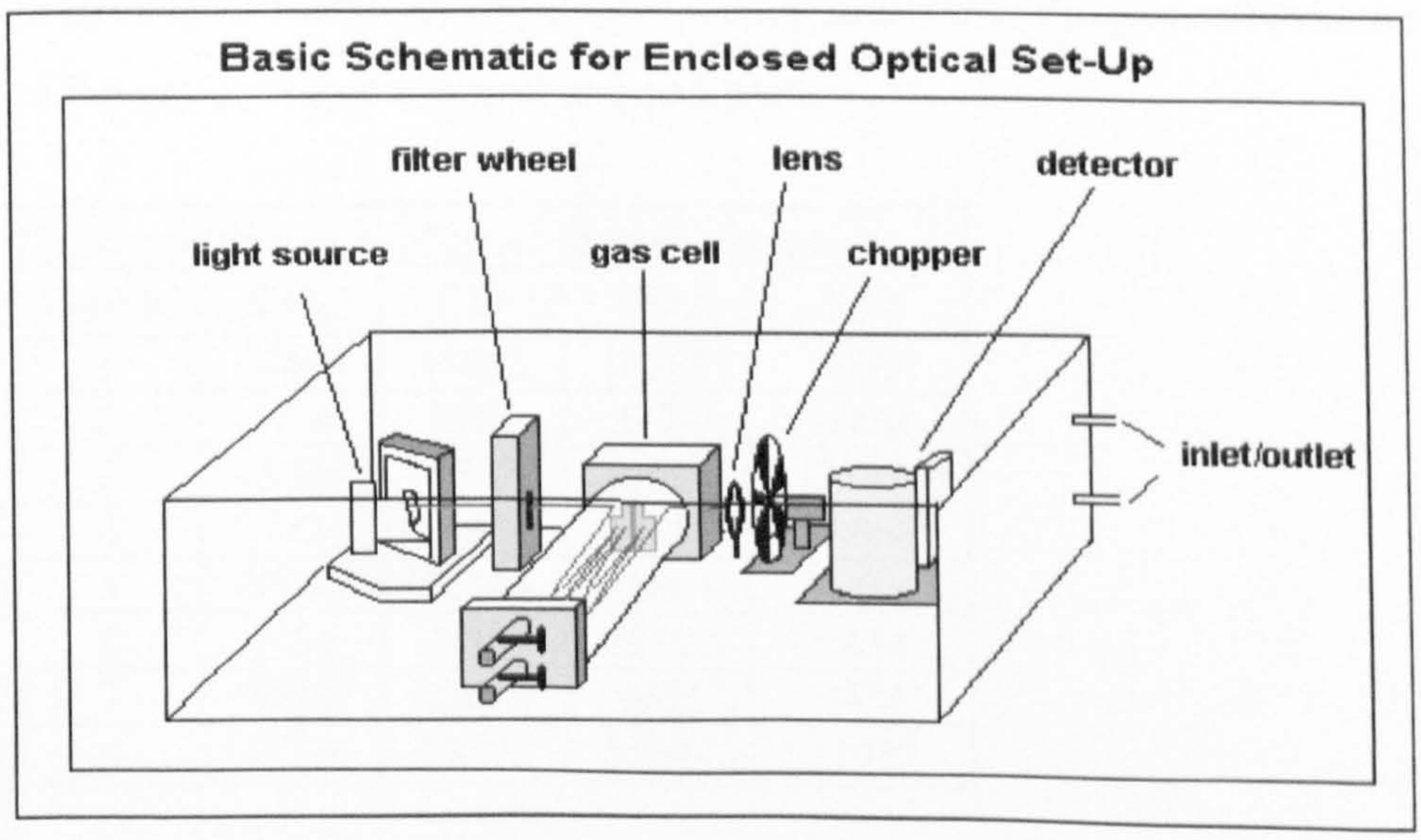


Figure 2.7.1 Schematic for the design of the box

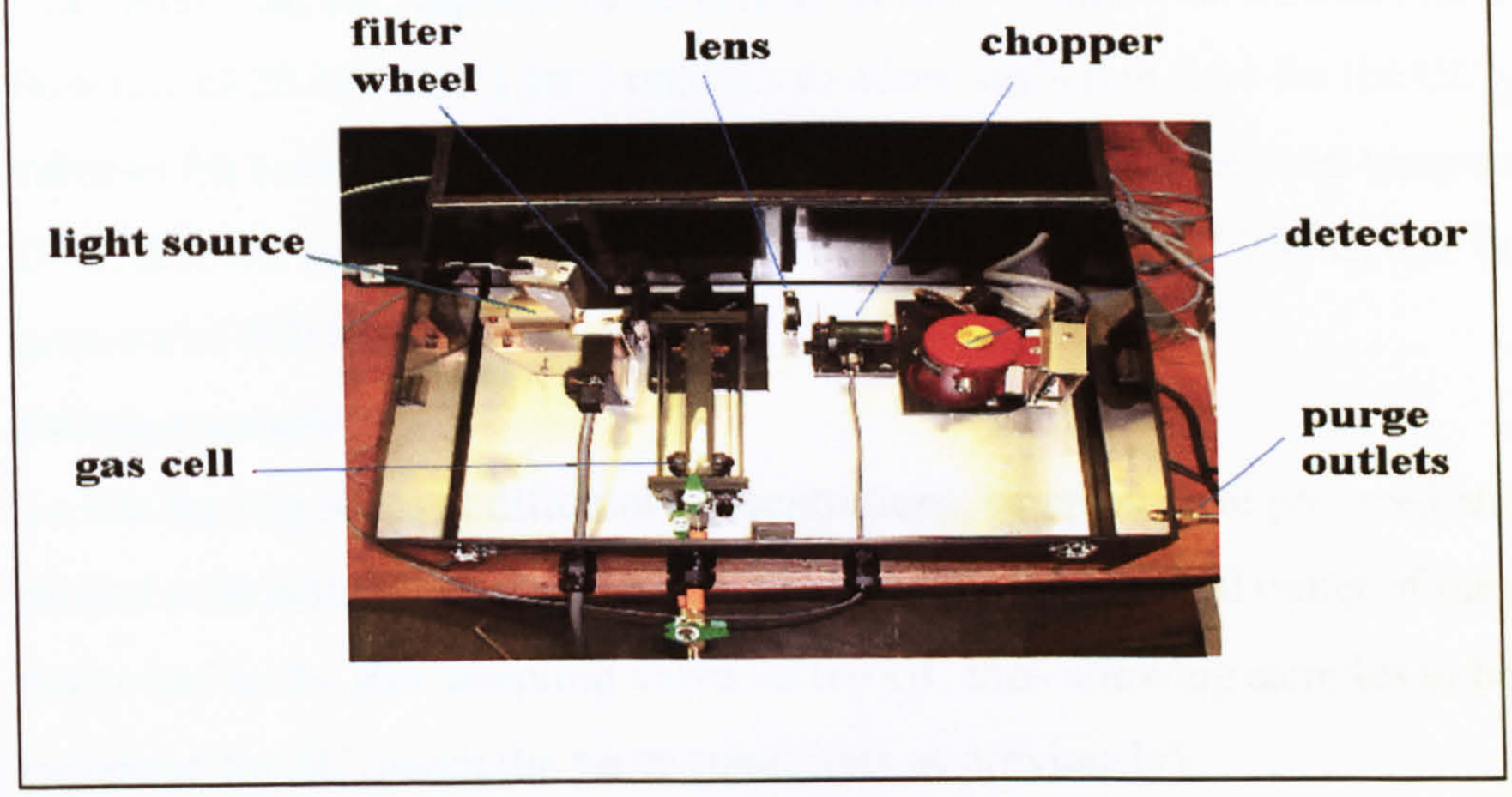


Figure 2.7.2 Photo of the inside of the box

2.7.2 Effect of purging

The box was initially tested with the MCT detector comparing the responses between no purging and with purging the box at a high flow rate (>100 cm^3min^{-1}) and a low flow rate (10-20 cm^3min^{-1}). Due to problems with the MCT detector, the DTGS detector was utilised instead (originally the DTGS detector was not used as this was the internal detector inside the FTIR and required an additional cable to control the peltier cooling).

2.7.3 Testing the box

Static samples

To test the box, samples of the four gases (the concentrations in each sample are shown in table 2.7.1) were prepared by injecting gases into the gas cell (via the external inlets to the gas cell) and allowed to equilibrate.

Concentration ppm of gases in each sample				
sample	**CH_4**	**CO**	**CO_2**	**N_2O**
1	18868	18868	18868	18868
2	3774	3774	3774	18868
3	15094	7547	15094	7547
4	37736	18868	18868	18868
5	49057	18868	56604	56604
6	9434	28302	28302	9434
7	9434	22642	9434	9434
8	37736	28302	15094	33962

Table 2.7.1 Concentration of gases in ppm added to each sample to test purgable box

After recording the response of each filter in turn, helium was flowed into the cell (at a flow rate of 20 cm^3min^{-1}) for 5 minutes to allow sufficient time for the GC gas sampling valve to fill before sampling by GC (The GC conditions were; oven temperature of 25°C, detector temperature of 200°C, filament temperature of 250°C, and flow carrier pressure of 200 kPa).

Dilution samples

To test the responses at different concentrations, samples were prepared and then diluted with helium at a flow rate of 20 cm^3min^{-1}, the external outlet of the gas cell was connected to the gas-sampling valve on the GC thus allowing samples to be also measured by GC (using the same conditions as previously).

Sample	Concentration
1	9.4 % CH_4, 7.5 % CO, 6.6 % CO_2, 8.5 % N_2O
2	4.7 % CH_4, 5.7 % N_2O
3	3.8 % CO_2
4	3.8 % CO
5	4.7 % N_2O
6	4.7 % CO_2

Table 2.7.2 Concentration of samples used for dilutions

2.7.4 Reproducibility

To test for reproducibility, repeat samples consisting of 1887 ppm each of the four gases and of 18880 ppm each of the four gases were prepared. These were measured a number of times at different time intervals investigate signal drift.

2.7.5 Five Level Experimental Design Calibrations

Preparation of samples and collection of data

To build a calibration model, 17 samples were preparared based on a five level experimental design. The level combinations for each sample (shown in table 2.7.3) were determined by D-optimal design, Design Expert 6.0.1 StatEase, USA. The concentrations (ppm) for each sample are shown in table 2.7.4. These were prepared in the same way as the previous samples. After 20 minutes, to allow the sample to equilibrate, the responses were measured twice with the different filters before sampling with GC (employing a 5 minute dilution with helium was used to fill the sampling value). Repeat samples were prepared and measured a week later.

Level of component for given sample				
Sample	**C1**	**C2**	**C3**	**C4**
1	3	2	2	1
2	1	5	1	1
3	4	3	2	4
4	3	4	1	4
5	1	4	5	5
6	2	5	5	2
7	1	1	4	4
8	2	3	1	3
9	5	3	4	5
10	3	5	4	3
11	4	2	1	2
12	2	1	3	5
13	5	2	3	4
14	5	1	5	1
15	2	4	4	1
16	5	4	2	2
17	3	3	3	2

Table 2.7.3 Five level experimental design for four components (C1-C4)

Concentration ppm of gases in each sample				
Sample	**CH_4**	**CO**	**CO_2**	**N_2O**
1	11321	7547	7547	3774
2	3774	18868	3774	3774
3	15094	11321	7547	15094
4	11321	15094	3774	15094
5	3774	15094	18868	18868
6	7547	18868	18868	7547
7	3774	3774	15094	15094
8	7547	11321	3774	11321
9	18868	11321	15094	18868
10	11321	18868	15094	11321
11	15094	7547	3774	7547
12	7547	3774	11321	18868
13	18868	7547	11321	15094
14	18868	3774	18868	3774
15	7547	15094	15094	3774
16	18868	15094	7547	7547
17	11321	11321	11321	7547

Table 2.7.4 Concentration of gases in ppm, based on a five level experimental design, added to each sample

2.7.6 Investigating the reproducibility of CH_4 responses

The reproducibility of CH_4 response was investigated by preparing repeat samples of CH_4 at concentrations of 3774 and 18868 ppm. After measuring the responses with just CH_4, varying amounts of the other 3 gases were added (to check for interferences), then GC measurements where made. Table 2.7.5 lists the concentrations of gases in each sample.

Concentration (ppm) of gases in each sample				
Sample	**CH_4**	**CO**	**CO_2**	**N_2O**
1	18868	3774	18868	3774
2	3774	3774	3774	18868
3	18868	18868	3774	3774
4	3774	18868	18868	18868
5	18868	-	-	-
6	3774	18868	15094	56604

Table 2.7.5 Concentrations of CH_4 and other gases used in samples to investigate reproducibility of CH_4 response

2.7.7 Repeat of Five Level Calibrations with higher CH_4 concentrations

The five level calibration samples were prepared as previously; however, the concentration of CH_4 was doubled. The same data collection and modelling was applied.

2.7.8 Sample preparation using a sample loop

To improve the reproducibility of samples it was decided to use a sample loop, similar to the gas-sampling valve on the GC.

The valve used was a 6-way valve, this allowed a sample loop to be filled with gas from the gas lecture bottle and then by switching the valve the loop was emptied into the cell using a controlled flow of helium (figure 2.7.3).

More gas mixtures were prepared employing the same levels from the five level experimental design in previous work.

Three different lengths of sample loop were employed; a small size at 20 cm, a medium size at 40 cm and a longer size measuring 58 cm.

The different loops could then be placed into the valve as required and sealed with PTFE seal tape.

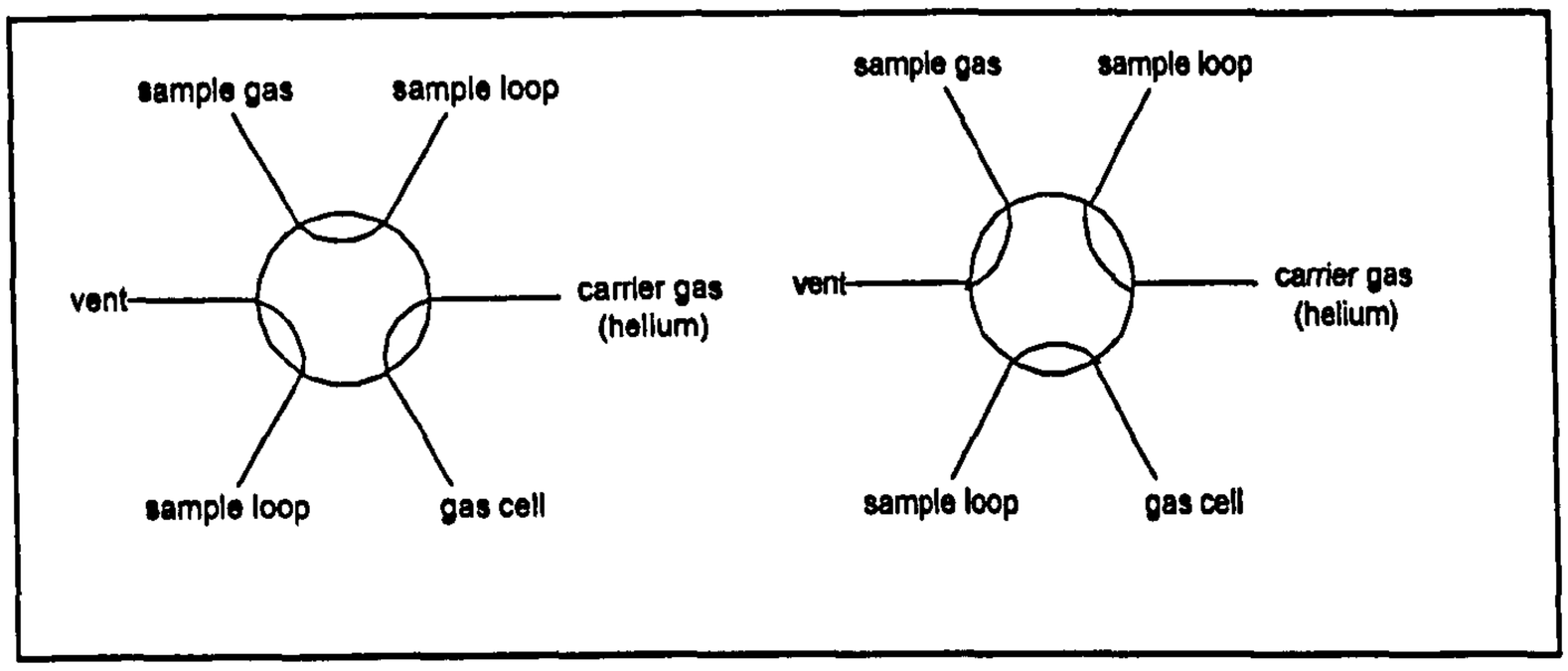


Figure 2.7.3 Connections on the sampling valve, left: valve positioned to fill sample loop, right: valve positioned to inject sample

Level one was achieved with the small loop, level two the medium, level three the long, level four using the medium loop twice and level five the long loop twice.

These equated to the concentrations shown in table 2.7.6.

level	ppm
1	1185
2	2371
3	3438
4	4742
5	6876

Table 2.7.6 Concentrations ppm prepared for different levels in the experimental design by employing sample loops

2.8 Volatile Organic Compound Analysis

2.8.1 Samples collected by vapour sampling

Collection of spectra

For this work spectra were collected for a-pinene, limonene (S), benzene, toluene, p-xylene, o-xylene, m-xylene, methanol, ethanol, butan-1-ol, cumene, dodecane, heptane, cyclohexane, cyclopentane, t-butanol, and 4-isopropyltoluene.

The infrared spectra were collected on a Perkin Elmer Paragon 1000 FTIR Instrument using the range 4000- 500 cm^{-1} with a resolution of 1 cm^{-1}.

Samples were obtained by using a 10 ml gas syringe to sample vapour directly from the bottles containing the compound of interest, this was then injected into a 100 ml gas cell. The cell was scanned and ratioed against a background spectra obtained from scanning the gas cell when filled with nitrogen. For less volatile compounds, the sampling was repeated till significant absorption peaks were obtained. Between samples the cell was flushed with nitrogen until the IR scans of the cell showed no sign of any residual sample peaks.

For some of the more volatile compounds, more than one spectrum was collected as samples were diluted to give similar sized absorption peaks to some of the compounds with weaker absorption spectra.

The spectra collected were compared to reference spectra from Aldrich[97] and NIST[64] to check for impurities.

The scans showed large amounts of water vapour present in the spectra, therefore the solvents were dried using molecular sieves. The scans still exhibited a number of peaks due to water vapour and carbon dioxide present in the atmosphere which were unavoidably collected in the sampling process.

PCA was performed on the data to investigate whether the water vapour peaks could be modelled out of the spectra; however, as most of the principal components included various water peaks this was unsuccessful.

Table 2.8.1 gives a list of spectra analysed, including duplicates at different concentrations. 'Noise' refers to samples of ambient air (for comparison purposes).

Spectra	Compound
1	toluene
2	α-pinene
3	α-pinene
4	limonene
5	noise
6	benzene
7	benzene
8	benzene
9	noise
10	p-xylene
11	p-xylene
12	p-xylene
13	toluene
14	α-pinene
15	α-pinene
16	cyclohexane
17	cyclohexane
18	cyclopentane
19	cyclopentane
20	dodecane
21	dodecane
22	heptane
23	heptane
24	methanol
25	m-xylene
26	o-xylene
27	o-xylene
28	noise
29	butan-1-ol
30	butan-1-ol
31	ethanol
32	ethanol
33	toluene
34	4-isopropyltoluene
35	t-butanol
36	t-butanol

Table 2.8.1 List of spectra and corresponding compounds

Analysis of Spectra

Principal Component Analysis (PCA)

PCA was employed to examine patterns and trends within the data. This was first performed on the entire spectra, and subsequently on different regions of the spectra. Region 1 covered the wavenumber range 4000-3400 cm^{-1}; region 2 the range 3200-2600 cm^{-1}; region 3: 2100-500 cm^{-1} and region 4: 1300-500 cm^{-1}. These regions were based on where spectral peaks occurred. Sample scores were plotted for different principal components to investigate whether clustering of similar compounds had occurred.

Different data pre-treatments were applied to the spectra including mean centring and standard normal variate (SNV).

Cluster Analysis

As well as score plots, dendrograms were also employed to investigate grouping between different compounds. These were built for both untreated and SNV treated data. The different types built included k distancing by neighbours and by groups, and by using 10 principal components with and without Mahalanobis distancing. For more details refer to section 2.1.

2.8.2 Improvement of Spectra by removal of water vapour

In order to decrease the amount of the CO_2 and water entering the cell, a sampling rig as shown in figure 2.8.1 was set up. This was used to evacuate the cell, then the tap leading to the cell was closed. The liquid sample of the compound to be analysed was cooled in liquid nitrogen and the surrounding air pumped out. The sample was warmed so that vapour could saturate the rig, then the tap to the cell was opened allowing vapour to fill into the evacuated gas cell.

The sample spectra were ratioed against spectra obtained for the cell without any sample present. The concentration of sample was dependent its volatility. For the more volatile compounds (such as benzene, toluene and methanol) the noise to sample ratio was low. For the less volatile compounds (e.g. limonene and pinene) the low concentrations of sample meant there was a high noise to sample ratio.

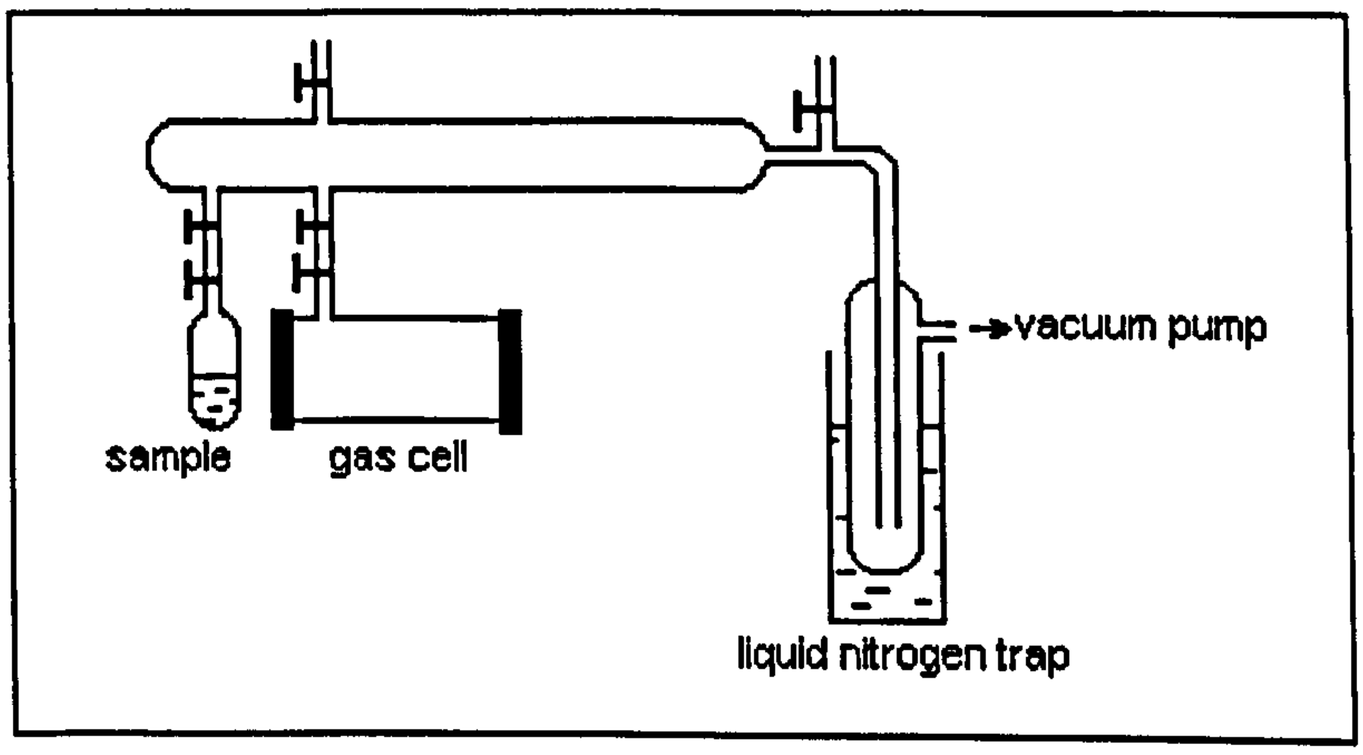


Figure 2.8.1 Schematic diagram of gas sampling rig

Improved spectra were collected for all of the following; a-pinene, butan-1-ol, cumene, cyclohexane, octane, p-isopropyltoluene, propanoic acid, butan-2-ol, limonene, benzene, p-xylene, cyclopentane, heptane, methanol, m-xylene, o-xylene, ethanol, toluene, t-butanol, t-butanol (high concentration) and noise.

PCA and dendrogram analysis was performed on the improved spectra in the same manner as with the previous spectra.

2.8.3 Vapour samples prepared by evaporation

Sample preparation

The aims of the previous sections (2.8.1 and 2.8.2) were to obtain qualitative spectra of different compounds to compare different spectra and look at clustering and grouping of

similar compounds. In order to build calibration models and prepare mixtures of different compounds samples of known concentrations were required.

To do this, μl litre amounts of liquid compound being investigated at were injected directly into the gas cell and allowed to evaporate.

Between samples the gas cell was purged with nitrogen and scans were ratioed against the corresponding spectra of the purged gas cell. The gas cell was scanned with a Perkin Elmer Paragon 100 spectrometer for the range 4000-400 cm^{-1}, using 4 cm^{-1} resolution, with measurements every 2 cm^{-1} .

The concentrations of samples prepared are shown in table 2.8.2.

Compound	Concentration (ppm)									
heptane	3.34	5.02	6.69	10.03	1.67					
methanol	18.15	18.15	30.26	54.46	66.57	96.83	24.21			
hexane	1.88	5.63	7.51	11.26	1.88	5.63	7.51	11.26	1.88	5.63
hexane (cont.)	11.26	15.02	1.88	5.63	11.26	1.88				
limonene	3.03	4.54	7.56	13.61						
butan-1ol	5.35	6.05	10.59	10.59	10.59	5.35	6.05	6.05	10.59	
ethanol	8.39	12.59	20.99	37.78	41.97	8.39	20.99	41.97	8.39	20.99
acetone	6.66	10.00	23.32	6.66	13.32	23.32				
hexene	1.96	3.92	7.84	15.68	1.96	7.84	13.72			
toluene	2.31	6.92	11.53	2.31	11.53	23.05	2.31	11.53	23.05	
cyclohexane	2.27	11.33	2.27	4.53	2.27	4.53	6.80	2.27	11.33	
o-xylene	2.02	10.12	12.14	2.02	10.12					
m-xylene	2.00	10.00	16.00	2.00	6.00	16.00				
p-xylene	2.00	10.00	10.00	14.00	4.00					
propanoic acid	3.28	6.57	16.42	9.85	9.85	3.28				
tert-butyl alcohol	2.68	5.35	10.71	5.35	2.68	8.03	13.38			

Table 2.8.2 Concentrations of samples prepared for different compounds (assuming entire volume injected into cell is vapourised)

For each compound a minimum of at least 3 sample spectra of different concentrations were collected. This allowed individual calibration plots for each compound to be produced as well as allowing molar absorptivities to be calculated.

Spectra Analysis

PCA was utilised to compare spectra for different compounds. MLR (with leave one out cross validation) was also performed using all the entire spectra, as well as plots at specific wavelengths corresponding to main absorption peaks.

Dendrograms with different clustering techniques were also employed.

PCA was performed on all the data as well as selected groups. The groups looked at were alcohols, xylenes and alkanes. Also investigated was acetone, propanoic acid and tertbutyl alcohol; cyclohexane, limonene, xylenes, toluene and hexane; ethanol,o-xylene, heptane and cyclohexane.

2.8.4 Preparation of mixtures of heptane, cyclohexane, ethanol and o-xylene

Mixtures were prepared using mixtures of low, medium and high values based on experimental design. The concentrations prepared (assuming complete vaporisation) are shown in table 2.8.3.

Sample	Cyclo-hexane	Ethanol	Heptane	Xylene
1	2.27	4.20	1.67	2.00
2	6.80	12.59	5.02	6.00
3	11.33	20.99	8.36	9.99
4	11.33	4.20	1.67	2.00
5	2.27	20.99	1.67	2.00
6	6.80	12.59	1.67	9.99
7	6.80	4.20	5.02	2.00
8	2.27	4.20	8.36	2.00
9	11.33	20.99	8.36	2.00
10	11.33	12.59	5.02	2.00
11	2.27	12.59	8.36	9.99
12	11.33	4.20	1.67	6.00
13	11.33	20.99	1.67	9.99
14	2.27	12.59	1.67	6.00
15	2.27	20.99	8.36	2.00
16	6.80	12.59	1.67	6.00
17	2.27	12.59	5.02	9.99
18	11.33	20.99	1.67	2.00
19	2.27	4.20	1.67	9.99
20	11.33	4.20	1.67	9.99
21	11.33	4.20	8.36	9.99
22	11.33	4.20	8.36	2.00
23	6.80	20.99	1.67	2.00
24	6.80	12.59	8.36	9.99
25	2.27	20.99	8.36	9.99
26	2.27	4.20	8.36	9.99
27	6.80	20.99	8.36	6.00
28	2.27	20.99	1.67	9.99
29	2.27	12.59	1.67	2.00
30	11.33	4.20	5.02	6.00
31	11.33	20.99	8.36	9.99

Table 2.8.3 Concentrations of components in each sample assuming complete vaporisation of liquid injected into the gas cell

Samples were prepared by injecting µl amounts (using a 25 µl syringe) of corresponding liquid directly into the gas cell then allowing it to evaporate. High-level

amounts corresponded to 5 µl, medium levels to 3 µl and low levels to 1µl. Initially 22 samples were prepared, a further 9 samples were prepared the next day.
PCA was performed on the spectra and score plots used to check for any outliers.
MLR and VS-MLR were performed on the first 22 samples using different splits for training and validation sets and the additional 9 spectra were used for further validation.

2.8.5 Investigations into different sample preparation techniques

Mixtures prepared as liquid before injecting into cell

As the previous method involved injecting µl amounts a high percentage of error would be expected therefore other methods examined.
Mixtures were prepared using a 1 ml syringe, by injecting 0.3 ml for high volumes, 0.2 for medium volumes and 0.1 ml for low volumes, plus 0.1 ml of acetone as an internal standard; µl amounts of the sample mixtures (the amount injected was based on the % acetone in the mixture so that the sample concentration of acetone (corresponding to 3.33 ppm) was present in each sample) were then injected into the gas cell and allowed to evaporate. The final concentrations of components in each sample are shown in table 2.8.4.

Sample	Cyclo-hexane	Ethanol	Heptane	Xylene
1	2.27	4.20	1.67	2.00
2	4.53	8.39	3.34	4.00
3	6.80	12.59	5.02	6.00
4	6.80	4.20	1.67	2.00
5	2.27	12.59	1.67	2.00
6	4.53	8.39	1.67	6.00
7	4.53	4.20	3.34	2.00
8	2.27	4.20	5.02	2.00
9	6.80	12.59	5.02	2.00
10	6.80	8.39	3.34	2.00
11	2.27	8.39	5.02	6.00
12	6.80	4.20	1.67	4.00
13	6.80	12.59	1.67	6.00
14	2.27	8.39	1.67	4.00
15	2.27	12.59	5.02	2.00
16	4.53	8.39	1.67	4.00
17	2.27	8.39	3.34	6.00
18	6.80	12.59	1.67	2.00
19	2.27	4.20	1.67	6.00
20	6.80	4.20	1.67	6.00
21	6.80	4.20	5.02	6.00
22	4.53	4.20	5.02	2.00

Table 2.8.4 Concentrations of components in each sample assuming complete vaporisation of liquid injected into the gas cell

Models were then built comparing the absorption spectra with no further treatment, absorption data was divided by the acetone peak at 1742 cm^{-1} and absorption data divided by acetone peak at 532 cm^{-1}. The effect of applying other data treatments such as normalisation, SNV and baseline subtraction methods was also investigated.

Mixtures using heptane as internal standard

Mixtures of ethanol, o-xylene and acetone were prepared using heptane as an internal standard, with injections made using a Neelson pippette with high values of 200 µl volumes, medium values of 125 µl and low volumes of 60 µl. Models were then built as previously (section 2.8.4).

Sample	Ethanol	Xylene	Acetone
1	839.50	399.68	666.41
2	251.85	119.90	199.92
3	251.85	119.90	416.51
4	524.69	119.90	416.51
5	251.85	119.90	199.92
6	251.85	119.90	416.51
7	839.50	399.68	666.41
8	251.85	119.90	666.41
9	839.50	399.68	199.92
10	524.69	119.90	199.92
11	839.50	119.90	199.92
12	251.85	249.80	199.92
13	251.85	399.68	666.41
14	524.69	399.68	199.92
15	251.85	399.68	199.92
16	251.85	399.68	666.41
17	839.50	119.90	666.41
18	524.69	119.90	666.41
19	839.50	249.80	199.92
20	524.69	399.68	666.41
21	524.69	399.68	666.41
22	524.69	249.80	199.92

Table 2.8.5 Concentrations (ppm) of components in each sample assuming complete vaporisation of liquid injected into the gas cell

2.8.6 PLS Investigations

In order to investigate the data and look for possible outliers and sources of error, PLS was employed. This included building models with different numbers of latent variables and comparing the effects on prediction error both for the calibration and validation data sets. Also investigated was the removal of samples that contained high amounts of error or had high leverage effects.

3 Results and Discussion

3.1 Gas Sample Spectra and the effect of different pre-treatments on data

The aim of this stage of the work was to collect IR spectra suitable for multivariate models. Initially just single gas samples were prepared allowing simple models to be built to compare the effect of different data pre-treatments. Later mixtures of four gases (CH_4, CO, CO_2 and N_2O) were prepared for further in depth analysis.

3.1.1 Comparison of different data treatments on spectra for single gas samples

3.1.1.1 Spectra collected

Normally with FTIR analysis, spectra are ratioed against background scans to obtain absorption spectra. Ratioing against a background scan has the advantage of removing artefact peaks and variations from source emission and detector drifts, however in some applications obtaining a background scan is not always feasible or possible.

In order to see if it was possible to proceed without the need to run separate blank samples, the data was left as single beam transmission scans and various data treatment techniques applied.

Figure 3.1.1 displays the raw single beam spectra obtained for single component spectra. The change in background emissions between samples can be clearly seen thus demonstrating the importance of finding an appropriate method of background correction. The results obtained with different data treatments were compared to results obtained with the raw data.

It is clear from figure 3.1.1 that all the spectra contain peaks due to CO_2 absorption. This is caused by ambient CO_2 present in the sample chamber and the interior of the instrument. Even with constant purging, it was difficult to fully remove all traces of CO_2 due to the design of the FTIR; hence for this stage of the work no attempts were made to remove the ambient CO_2 and it was generally assumed that ambient levels of CO_2 were more or less constant. Also visible in the spectra were water vapour peaks in the 1400-2000 cm^{-1} and 3550-4000 cm^{-1} regions.

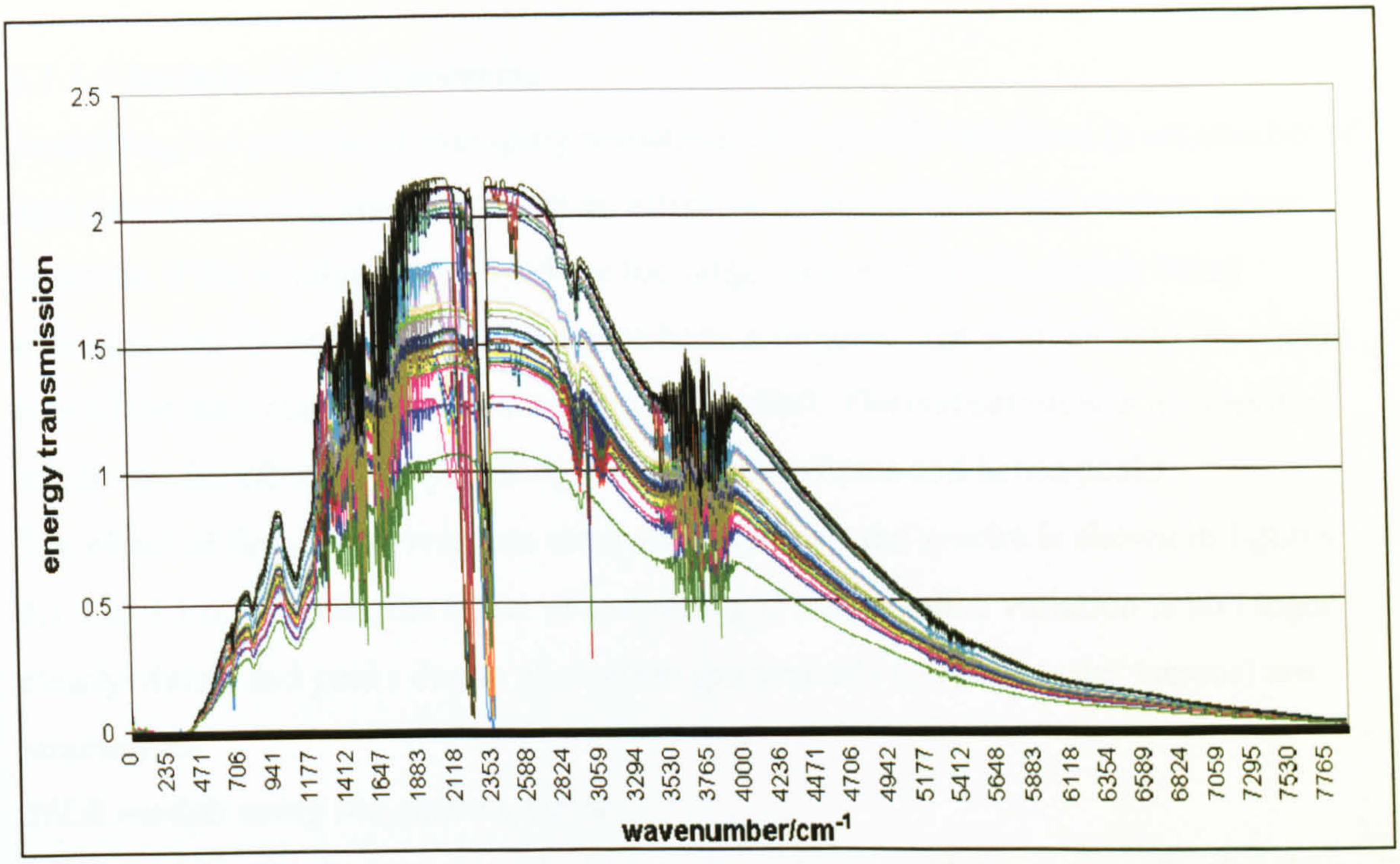


Figure 3.1.1 Single beam transmission spectra obtained for single component gas samples

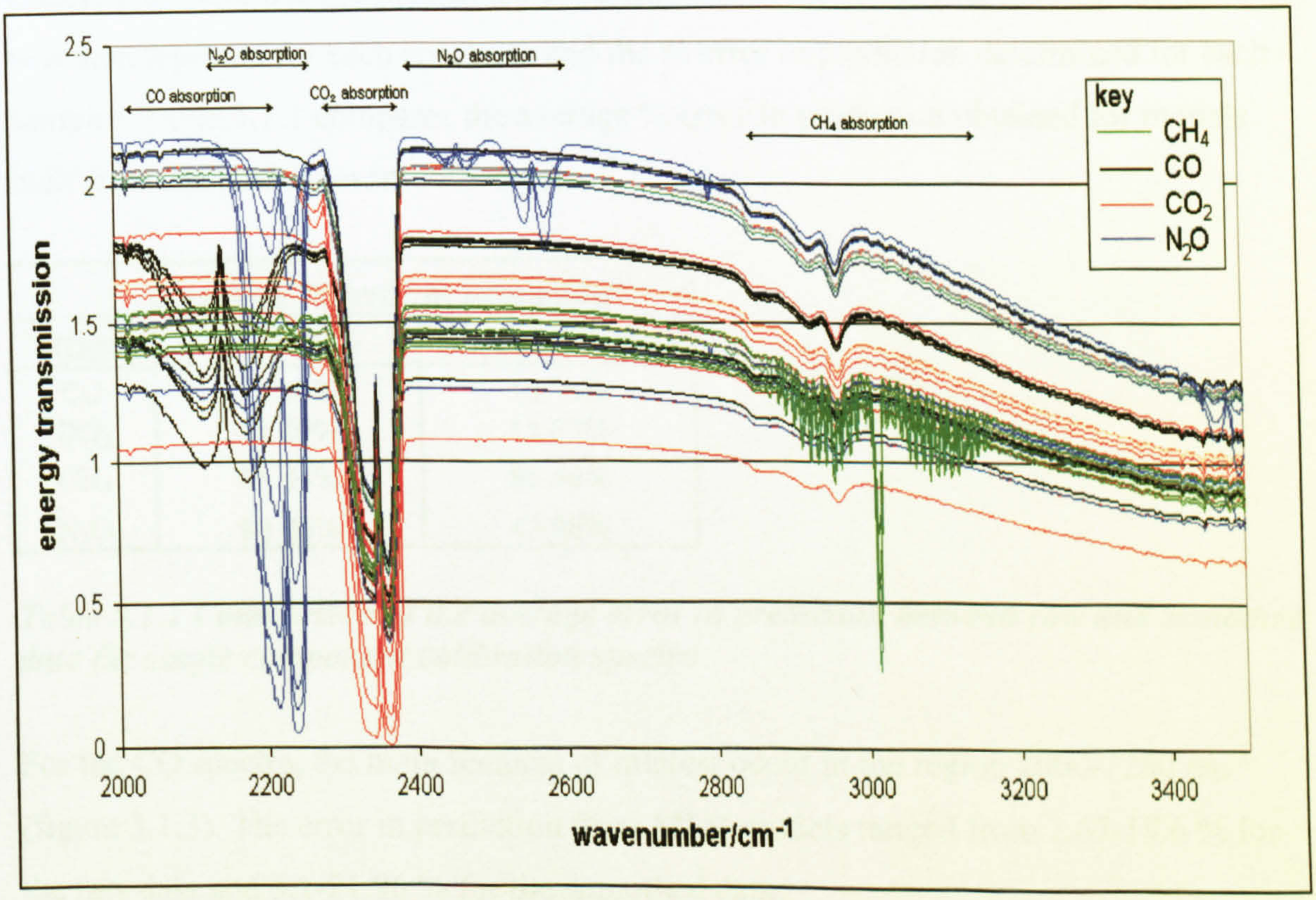


Figure 3.1.2 Close-up of region 2000-3500 cm^{-1} (high lighting the main absorption peaks) from single beam transmission spectra obtained for single component gas samples

Figure 3.1.2 is a close up of the region 2000-3500 cm^{-1} highlighting the regions of absorption for the gases of interest.

3.1.1.2 Savitsky-Golay Smoothing

Smoothing is a process of averaging responses over specific windows (a set number of data measurements), hence it can be an effective means of removing random noise; however, if the windows employed are too large this can result in signals being averaged out. Generally, as was the case here, a window size equivalent to the square root of the total number of measurements is applied. Derivatisation was also applied, which has the effect of emphasising changes in gradients and hence peaks.

The effect of first derivative, first order smoothing on the spectra is shown in figures 3.1.3 to 3.1.6. Overall, the effect of smoothing is that baseline variation is no longer clearly visible and peaks due to absorption (particularly CO_2 and water vapour) are emphasised.

MLR models using the entire spectra

MLR models were built as described in section 2.1, using "leave one out" validation (where all but one spectrum were used to calculate the MLR coefficients, these coefficients were used to predict the concentration of the remaining spectra; this process was then repeated for each spectrum and the % error in prediction determined for each sample). Table 3.1.1 compares the average % error in prediction obtained for models built from raw and from smoothed data.

	Average % error in prediction	
Gas	**Raw data**	**Smoothed data**
CO	11.55%	10.15%
CO_2	14.80%	11.63%
CH_4	31.44%	56.46%
N_2O	85.56%	42.88%

Table 3.1.1 Comparison of the average error in prediction between raw and smoothed data for single component calibration spectra

For the CO spectra, the main features of interest occur in the region 2000-2200 cm^{-1} (figure 3.1.3). The error in prediction from MLR models ranged from 2.67-19.6 % for the raw data and 3.1-21.96 % for the smoothed data.

The main features for CO_2 spectra (figure 3.1.4) are the absorption peaks at 2200-2400 cm^{-1}, 3600-3800 cm^{-1} and 600-800 cm^{-1}. The prediction errors for raw data varied from 6.34-20.5 % and for smoothed data from 3.28-20.93 %.

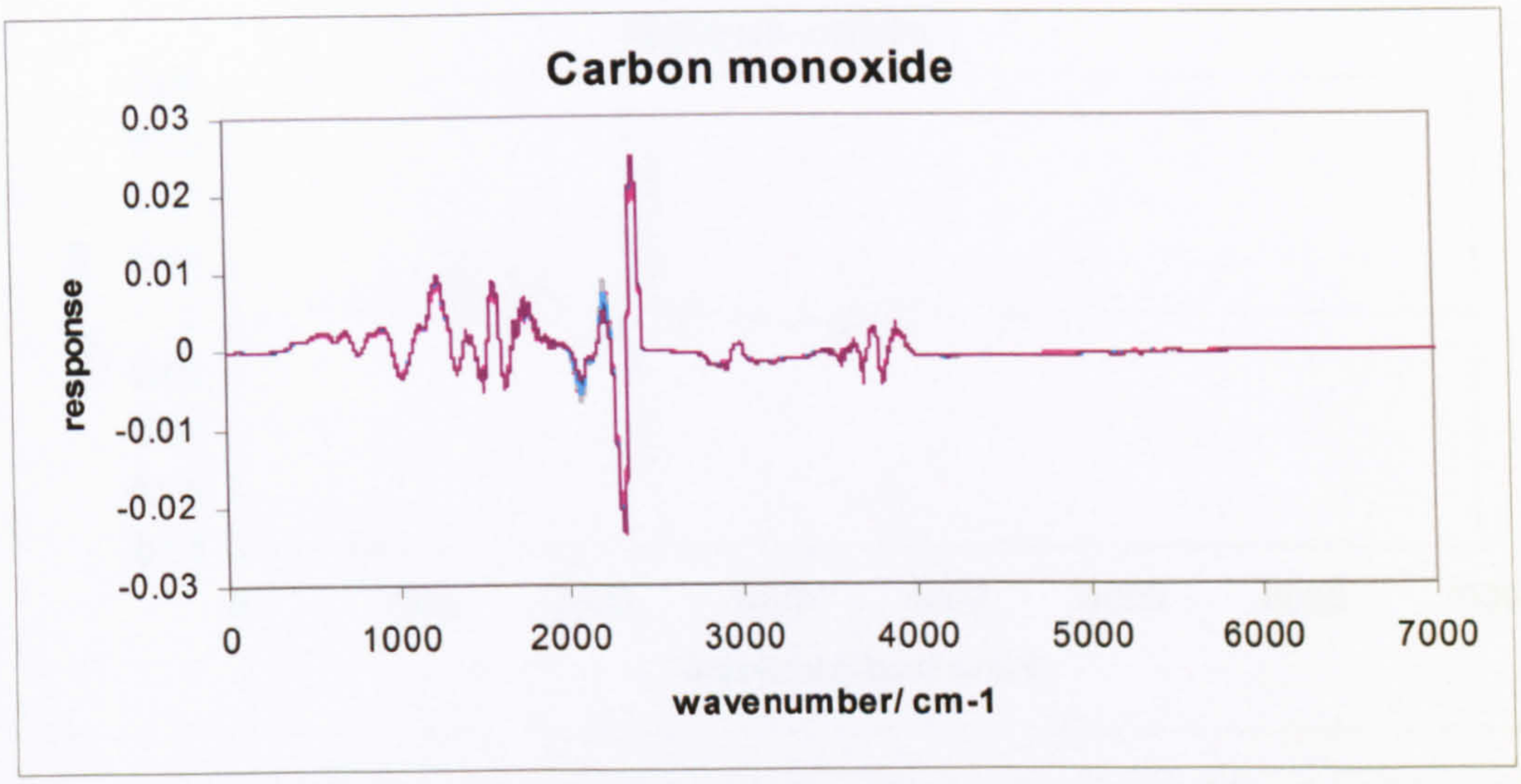


Figure 3.1.3 CO data after first derivative, first order smoothing has been applied

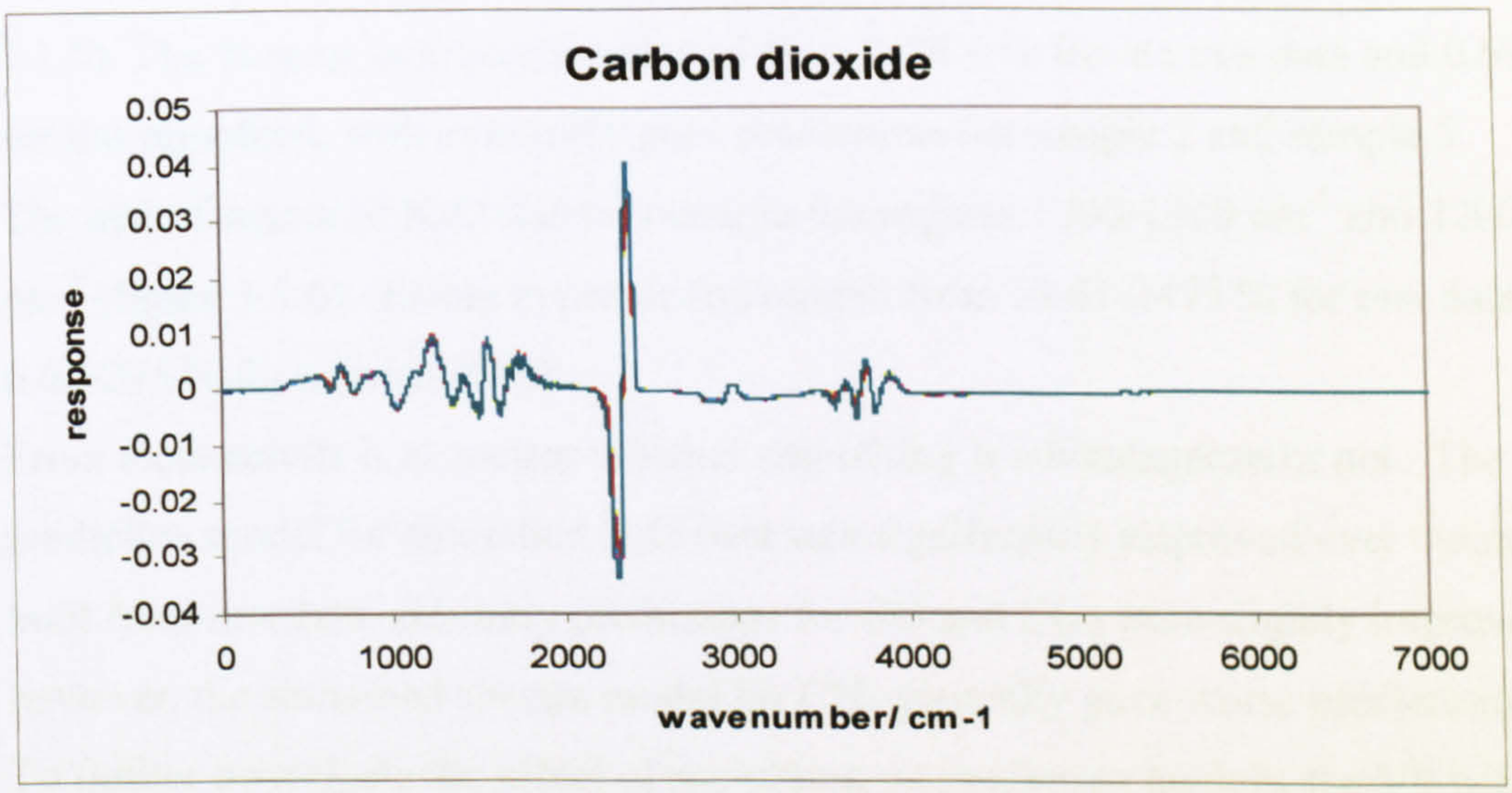


Figure 3.1.4 CO_2 data after first derivative, first order smoothing has been applied

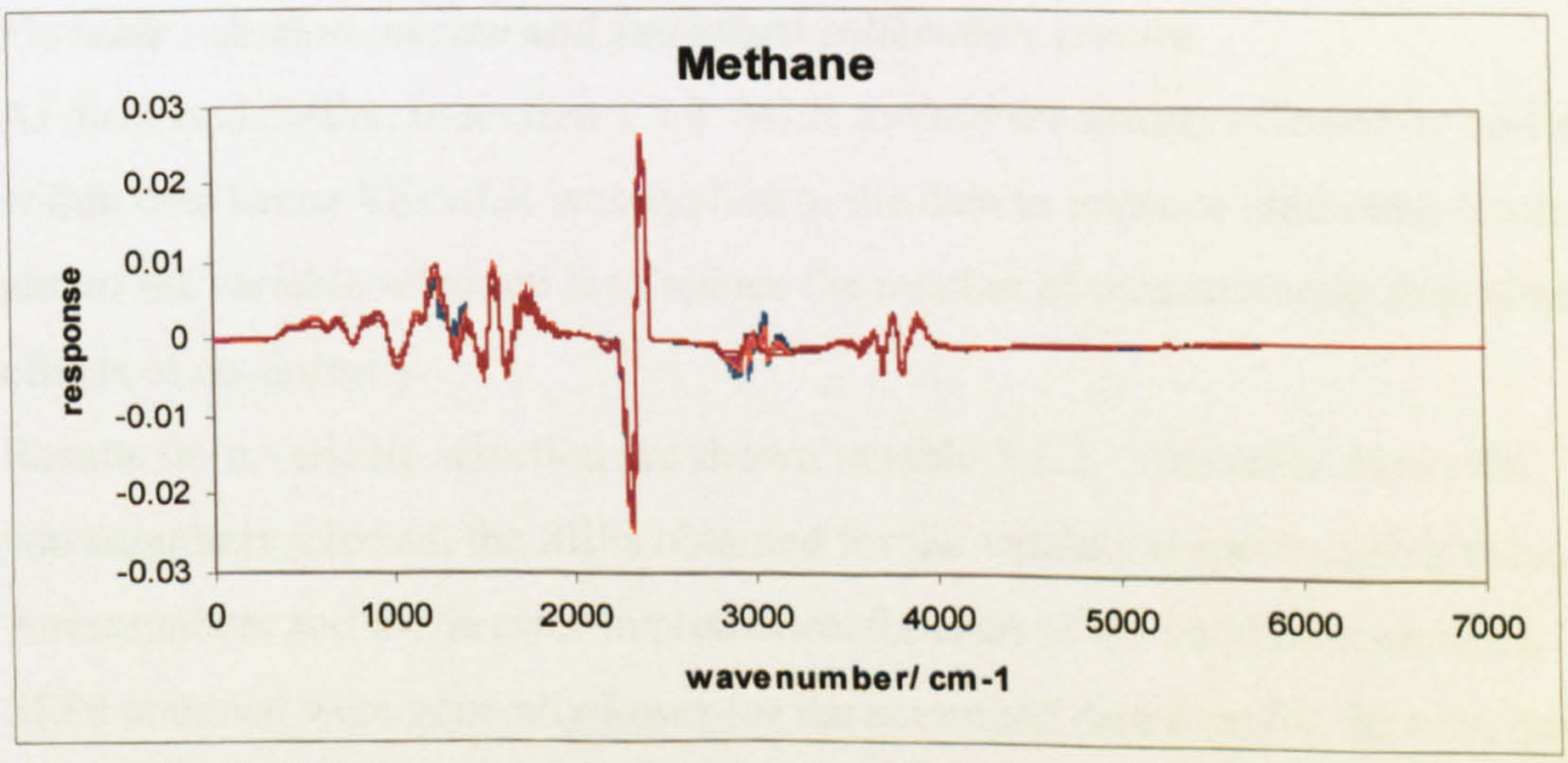


Figure 3.1.5 CH_4 data after first derivative, first order smoothing has been applied

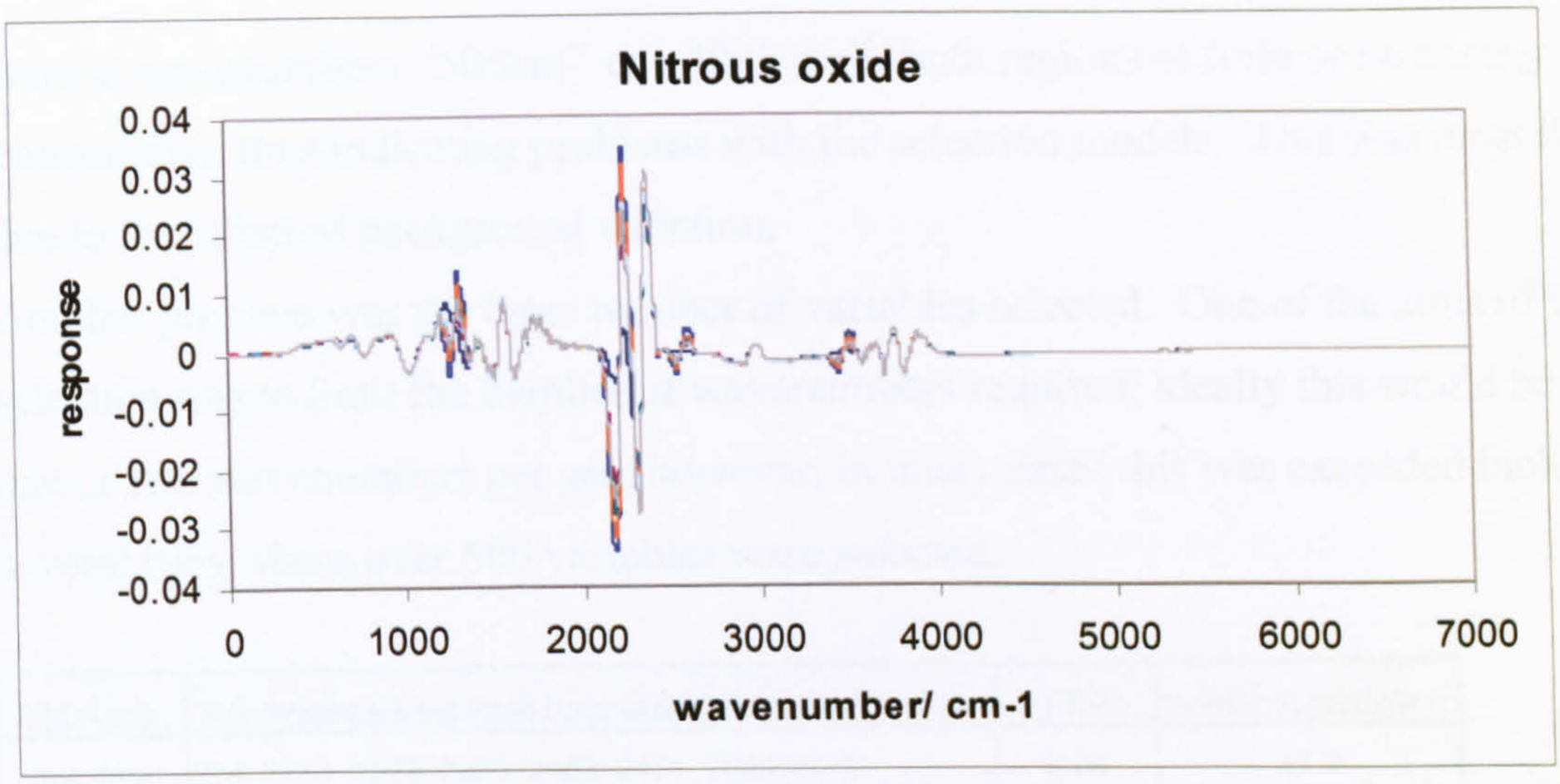


Figure 3.1.6 N_2O data after first derivative, first order smoothing has been applied

For CH_4 the main absorption peaks occur at 2800-3200 cm^{-1} and 1200-1400 cm^{-1} (figure 3.1.5). The % error in prediction ranged from 0-98.6 % for the raw data and 0.69-204 % for the smoothed, with extremely poor predictions for sample 2 and sample 5.

The main features of N_2O spectra occur in the regions 1200-1300 cm^{-1} and 2200-2300 cm^{-1} (figure 3.1.6). Errors in prediction ranged from 33.61-2475 % for raw data and 6.63-295 % for smoothed data.

From these results it is unclear whether smoothing is advantageous or not. The prediction model for smoothed N_2O data was significantly improved over the model built from raw data, similarly predictions for CO and CO_2 were slightly improved; however, the smoothed spectra model for CH_4 generally gave worse predictions.

To further investigate the effect of smoothing on prediction models, the VS-MLR algorithm was applied to the data.

Variable selection on raw and smoothed calibration spectra

As discussed earlier, in section 1.3.3, MLR models are greatly affected by co-linearity within data hence VS-MLR was applied to the data to improve prediction errors. The aim of the variable selection is to reduce the number of measurements thus eliminating effects of co-linearity.

Results from variable selection are shown in table 3.1.2. This table shows the wavenumbers selected, the SEPs obtained for the validation spectra using the selected wavenumbers and the % error in prediction for each of the validation samples. The SEPs obtained were generally lower for the smoothed data than for the raw; however, comparing the selected wavenumbers to where they occur on the spectra it is clear that none of these related to the main regions of absorption. Many of the selected variables

were at wavenumbers <500cm^{-1} or >4000 cm^{-1}, both regions of little or no energy transmission thus indicating problems with the selection models. This was most likely due to the effect of background variation.

Another problem was the large number of variables selected. One of the aims of the selection was to limit the number of wavenumbers required; ideally this would be only one or two wavenumbers per gas; however, in many cases this was exceeded including several runs where over 500 variables were selected.

N_2O data	Wavenumbers selected from variable selection models	SEP	% error in prediction
raw data	274 7753 3503 2380 3483 2171 7883 cm-1	2.06	41.9
	458 wavenumbers selected	1.52	18.61
	157 62 351 359 1281 cm-1	2.35	92.9
	201 48 46 164 220 cm-1	1.97	49.55
Smoothed	5709 1348 3825 1308 3530 1836 2255 cm-1	1.48	13.76
	2519 2517 2133 2534 2513 cm-1	1.9	10.07
	7870 2149 6502 2475 3462 6408 7146 cm-1	1.53	14.09
	2627 2477 3537 cm-1	1.69	15.04
CH_4 data	Wavenumbers selected from variable selection models	SEP	% error in prediction
raw data	262 174 13 cm-1	3.79	29.71
	243 368 370 257 137 cm-1	2.32	20.06
	368 262 60 156 cm-1	1.54	8.87
	147 137 71 cm-1	2.18	18.44
Smoothed	75 variables selected	2.76	21.04
	65 variables selected	1.77	17.3
	7794 1381 3055 cm-1	0.76	6.38
	26 variables selected	1.71	12.69
	220 222 218 cm-1	1.64	22.15
CO_2 data	Wavenumbers selected from variable selection models	SEP	% error in prediction
raw data	12 7760 7854 cm-1	1.68	7.41
	177 197 339 cm-1	1.57	10.1
	123 270 129 cm-1	0.69	5.59
Smoothed	301 3634 cm-1	0.46	1.82
	over 620 variables selected	0.32	3.27
	3180 variables selected	0.52	3.32
	1668 2438 3732 1086 cm-1	0.93	5.43
	526 variables selected	0.32	3.27
CO data	Wavenumbers selected from variable selection models	SEP	% error in prediction
raw data	104 243 cm-1	0.14	0.82
	214 187 cm-1	1.38	5.13
Smoothed	797 cm-1	0.3	105.24
	324 301 95 2750 cm-1	0.01	0.04

Table 3.1.2 Errors in prediction obtained for different sets of wavenumbers selected from models built using smoothed and raw data from single component samples

Correlation at different wavenumbers for raw and smoothed data.

Correlation coefficients were calculated at each wavenumber between the response and the amount of gas injected. This was performed on the raw and the smoothed data to see if this had any bearing on why the selected variables were not in the regions of interest. Correlations for the raw data, shown in figure 3.1.7, generally resemble absorption spectra; however, the correlations for the cut-off and low energy transmission regions (i.e. <500 and >5000 cm^{-1}) were high (both positive and negative) and this possibly explains why most of the selected variables were chosen from these regions.

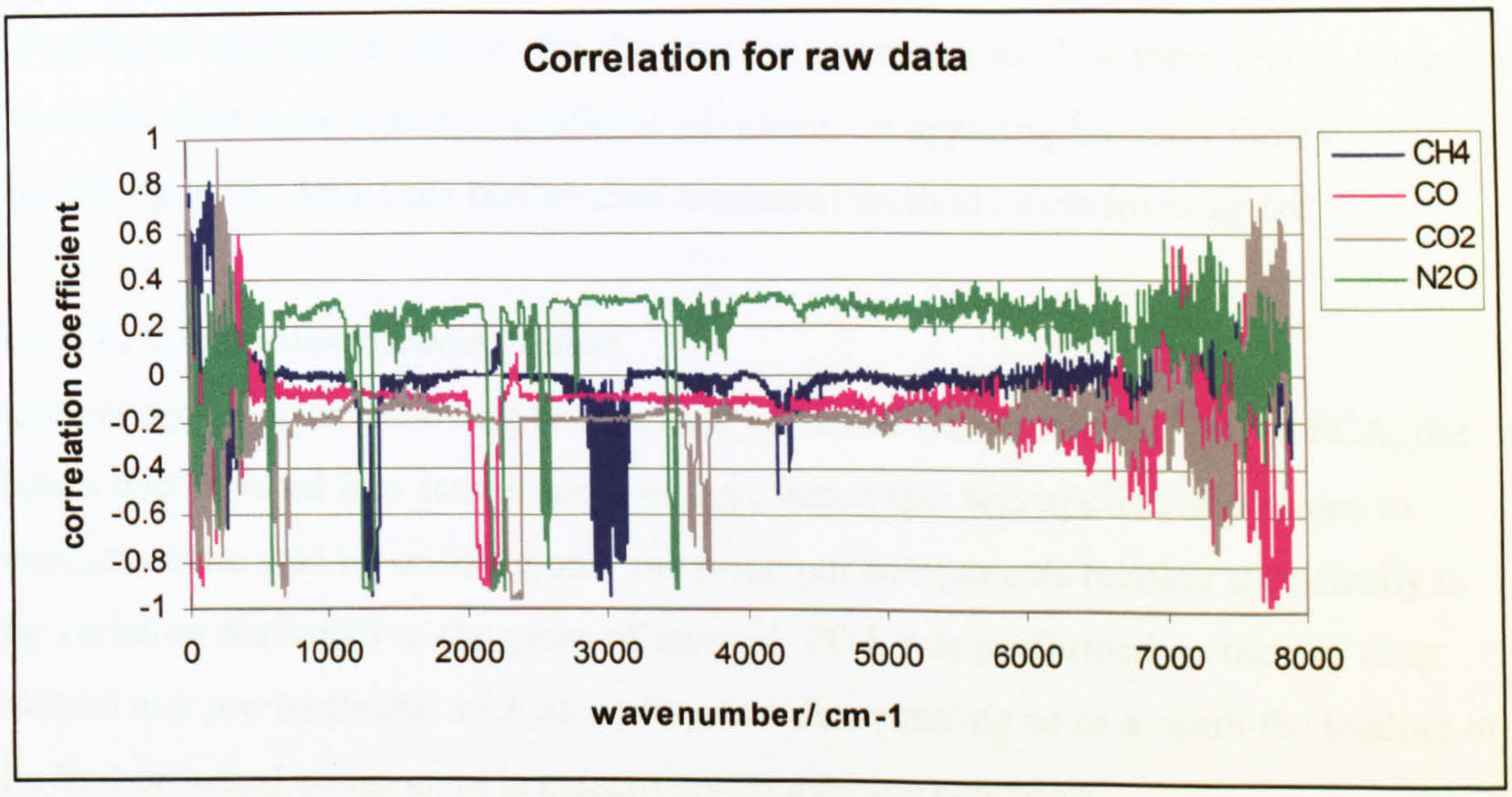


Figure 3.1.7 Correlation coefficients between response and concentration at different wavenumbers for single component gas samples

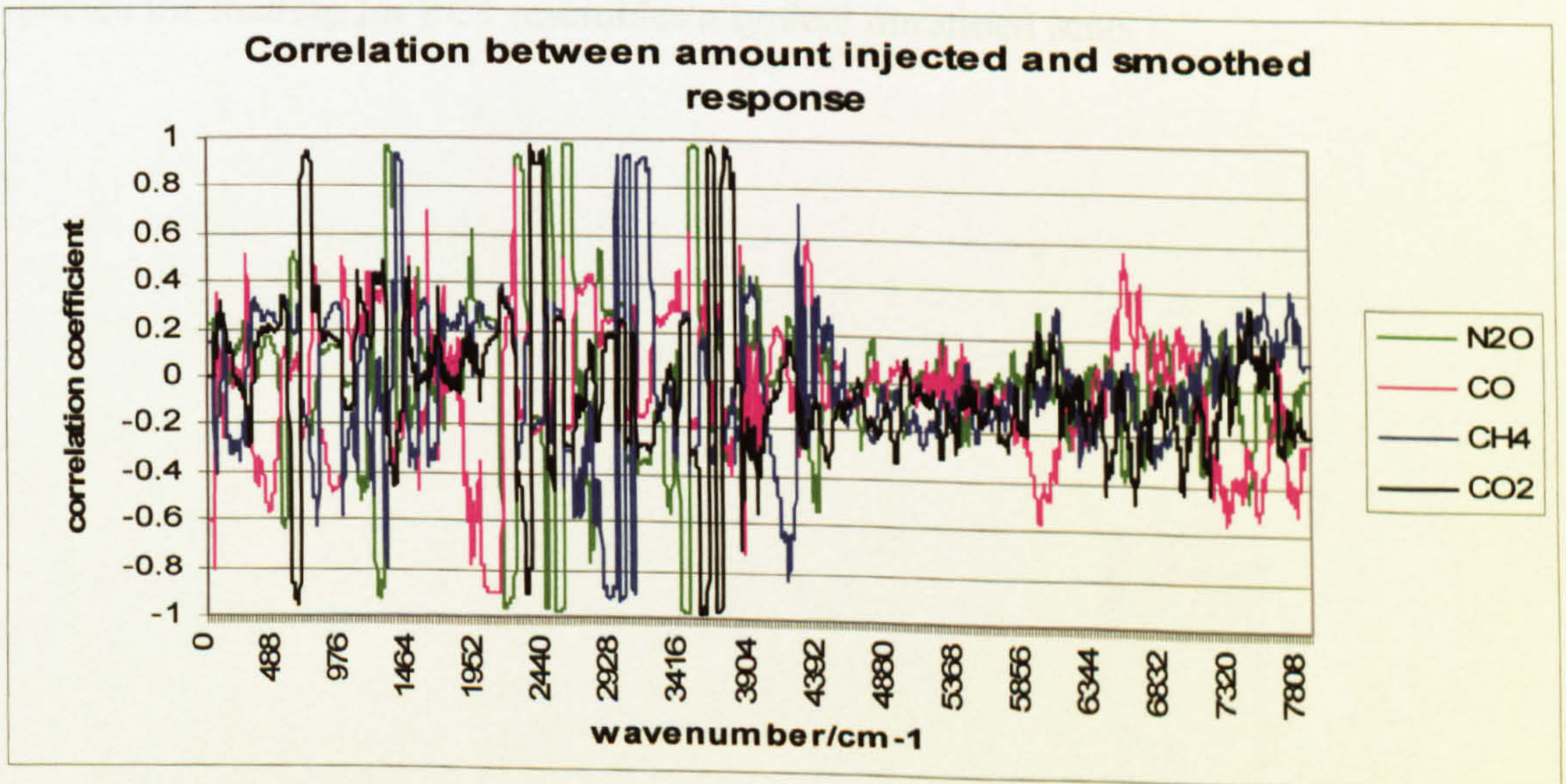


Figure 3.1.8 Correlation coefficients between response and concentration at different wavenumbers for smoothed data

The results for the smoothed response, shown in figure 3.1.8, gave high correlations (both positive and negative due to derivatisation) for variables relating to regions of gas absorption; however, unlike the analysis performed on the raw data, the correlation did not increase at the end portions of the spectra.

Conclusion on Savitsky Golay Smoothing

MLR models using the entire spectra gave no clear indication that smoothing improved prediction ability. Applying the VS-MLR algorithm to the data generally gave lower SEPs than for the raw untreated data; however, the selected wavenumber sets were unsatisfactory as they either contained too many wavenumbers or wavenumbers that did not relate to absorption regions for the gases of interest. Based on these results it can be concluded that there was no significant advantage in applying Savitsky Golay smoothing to the data, thus further data treatment methods were investigated.

3.1.1.3 PCA Reconstruction of data

Another approach to removing background variation was to utilize PCA. In PCA, the data is decomposed into scores and loadings, hence this was applied as a means to reconstruct the data by utilising only the principal components relating specifically to the variation attributed to the gases of interest. PCA was performed on the raw data without any pre-treatment such as scaling or mean centring so as a result the loading of the first principal component is effectively an average spectrum.

For PCA performed on the spectra collected for CH_4, 99.98 % of variation was described by PC1 (principal component one) and a further 0.02 % described by PC2. As expected the loading for PC1 resembles a typical unratioed scan.

The loading for PC2 resembles a typical absorption spectrum for CH_4. These loadings, the scores for PC2 and the reconstructed data are shown in figure 3.1.9. Reference spectra for CH_4, CO, CO_2 and N_2O (taken from the NIST database[64]) can be found in the appendix (section 6).

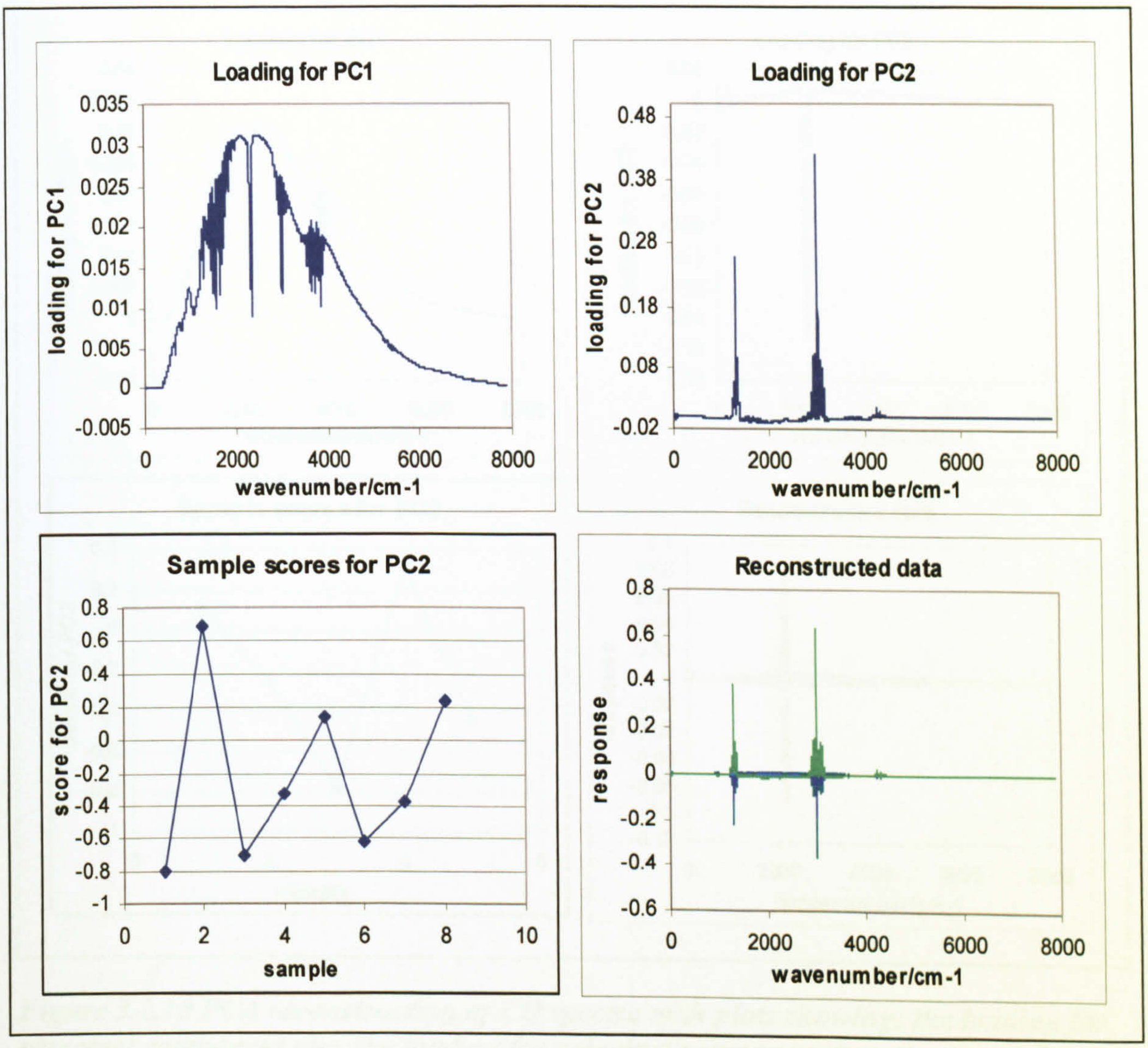


Figure 3.1.9 PCA reconstruction of CH_4 spectra with plots showing; the loading for principal component one, the loading for principal component two, the scores for principal component two, and the reconstructed data obtained by multiplying the scores and loadings for principal component two

Similar results were obtained for PCA on the CO spectra with PC1 describing 99.98 % of variation and component two accounting for 0.02 %. Examining plots of the results (figure 3.1.10) it can be seen that the loading for PC2 gave a similar pattern in response to reference spectra for CO.

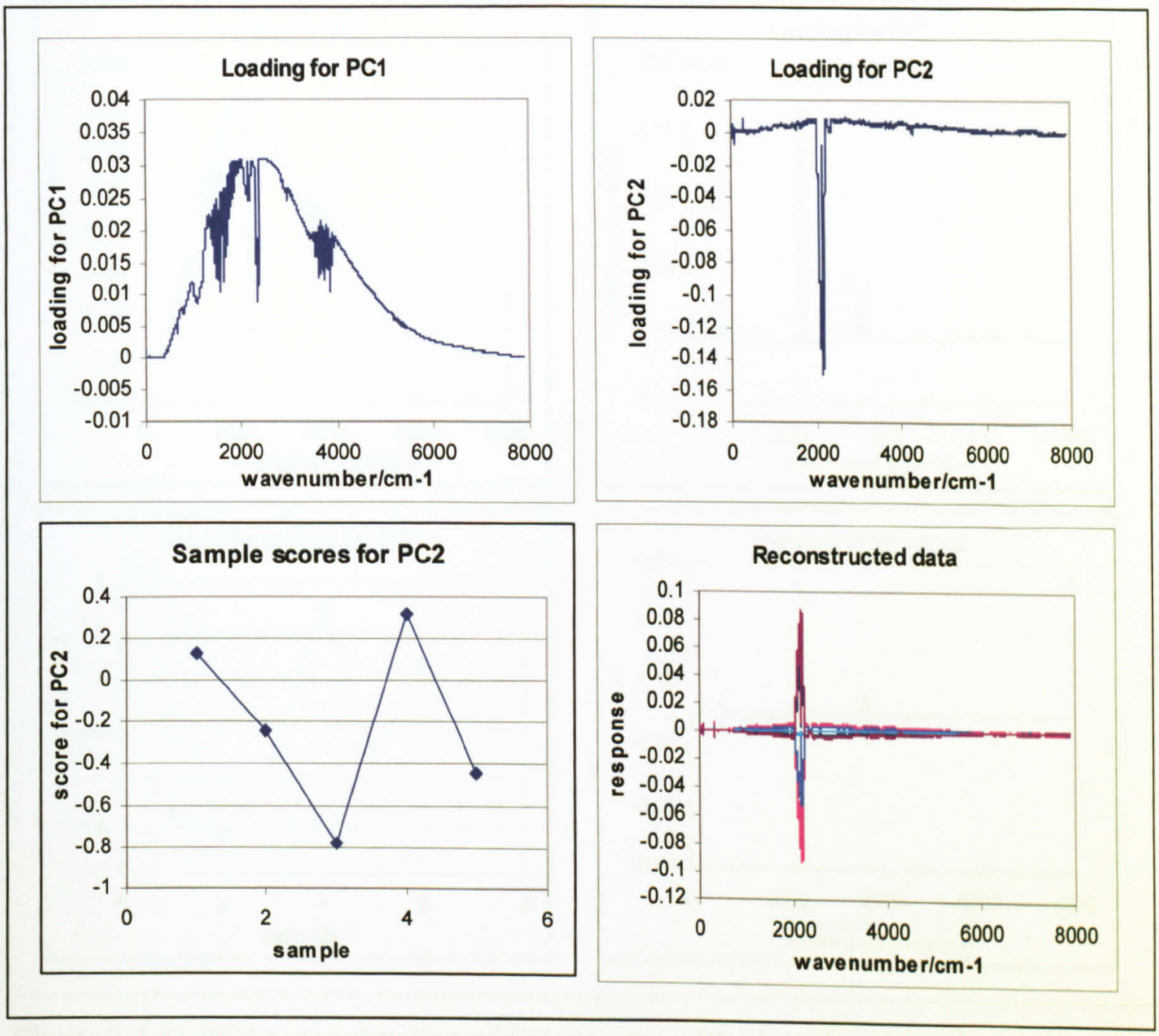


Figure 3.1.10 PCA reconstruction of CO spectra with plots showing; the loading for principal component one, the loading for principal component two, the scores for principal component two, and the reconstructed data obtained by multiplying the scores and loadings for principal component two

PCA on the CO_2 spectra gave 99.95 % of the variation described by PC1, with a further 0.05 % described by PC2. Again the reference spectra resembled most closely the loading for PC2 (figure 3.1.11).

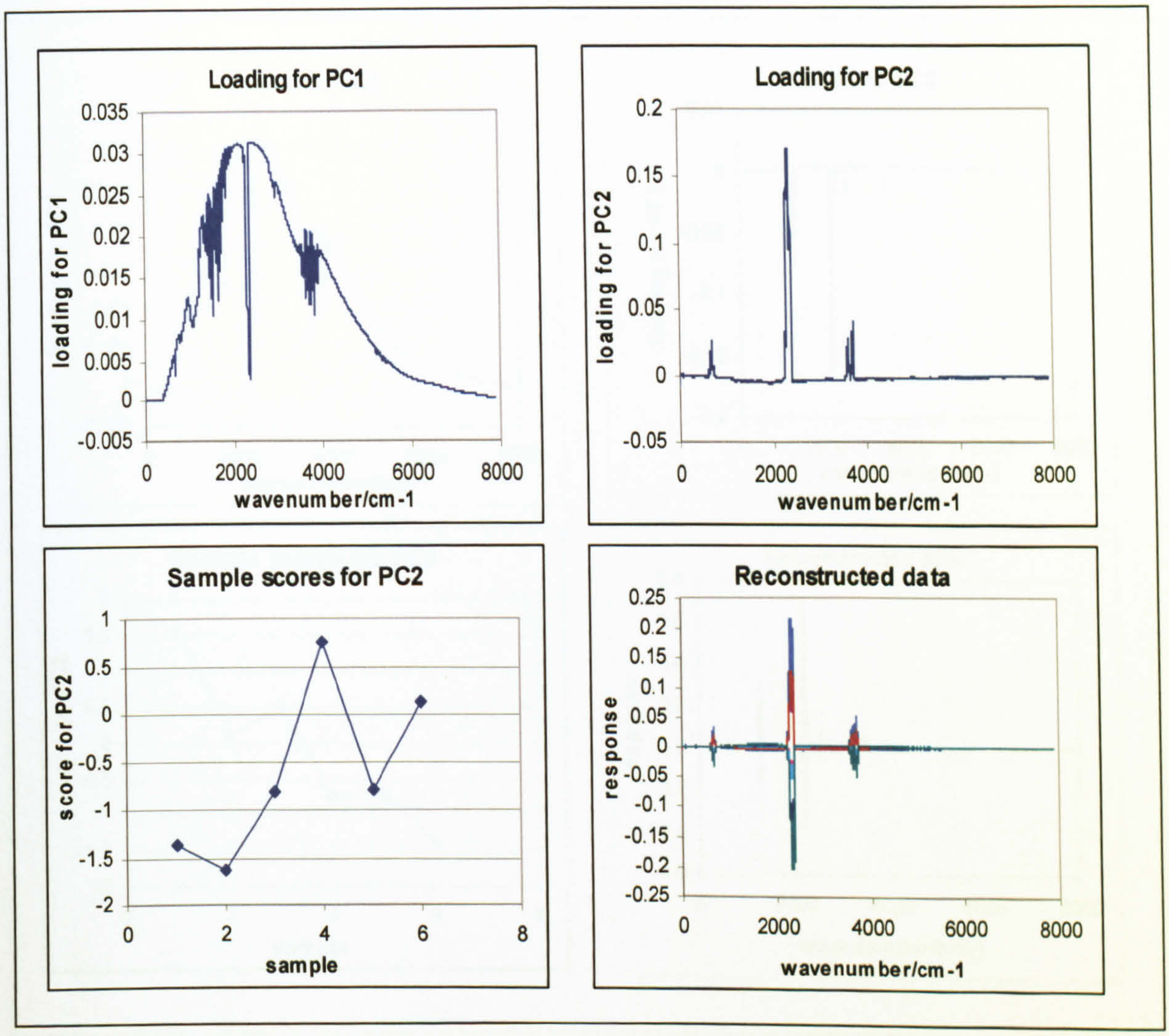


Figure 3.1.11 PCA reconstruction of CO_2 spectra with plots showing; the loading for principal component one, the loading for principal component two, the scores for principal component two, and the reconstructed data obtained by multiplying the scores and loadings for principal component two

PCA on N_2O resulted in 99.7 % of variation that was described by PC1, and 0.3 % described by PC2. As with the analysis on the other gases, the loading for PC2 resembled the reference spectra (figure 3.1.12).

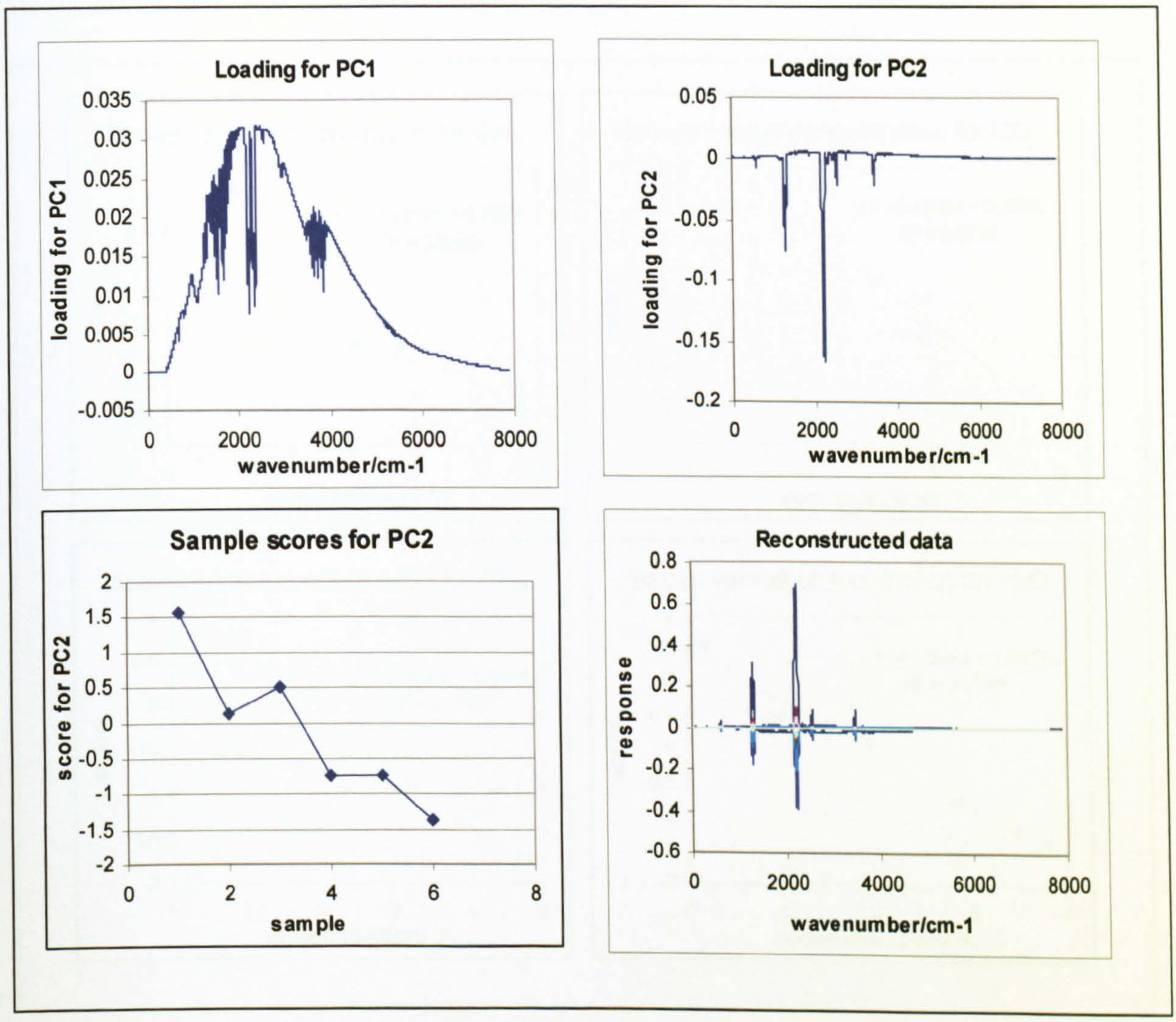


Figure 3.1.12 PCA reconstruction of N_2O spectra with plots showing; the loading for principal component one, the loading for principal component two, the scores for principal component two, and the reconstructed data obtained by multiplying the scores and loadings for principal component two

Figure 3.1.13 displays plots of the amount of gas injected against the sample scores for PC2 for each of the gases. The plots give correlations between the scores and concentrations ranging from 0.9 to 0.98, thus demonstrating that the variation described by PC2 does indeed relate to the gases of interest.

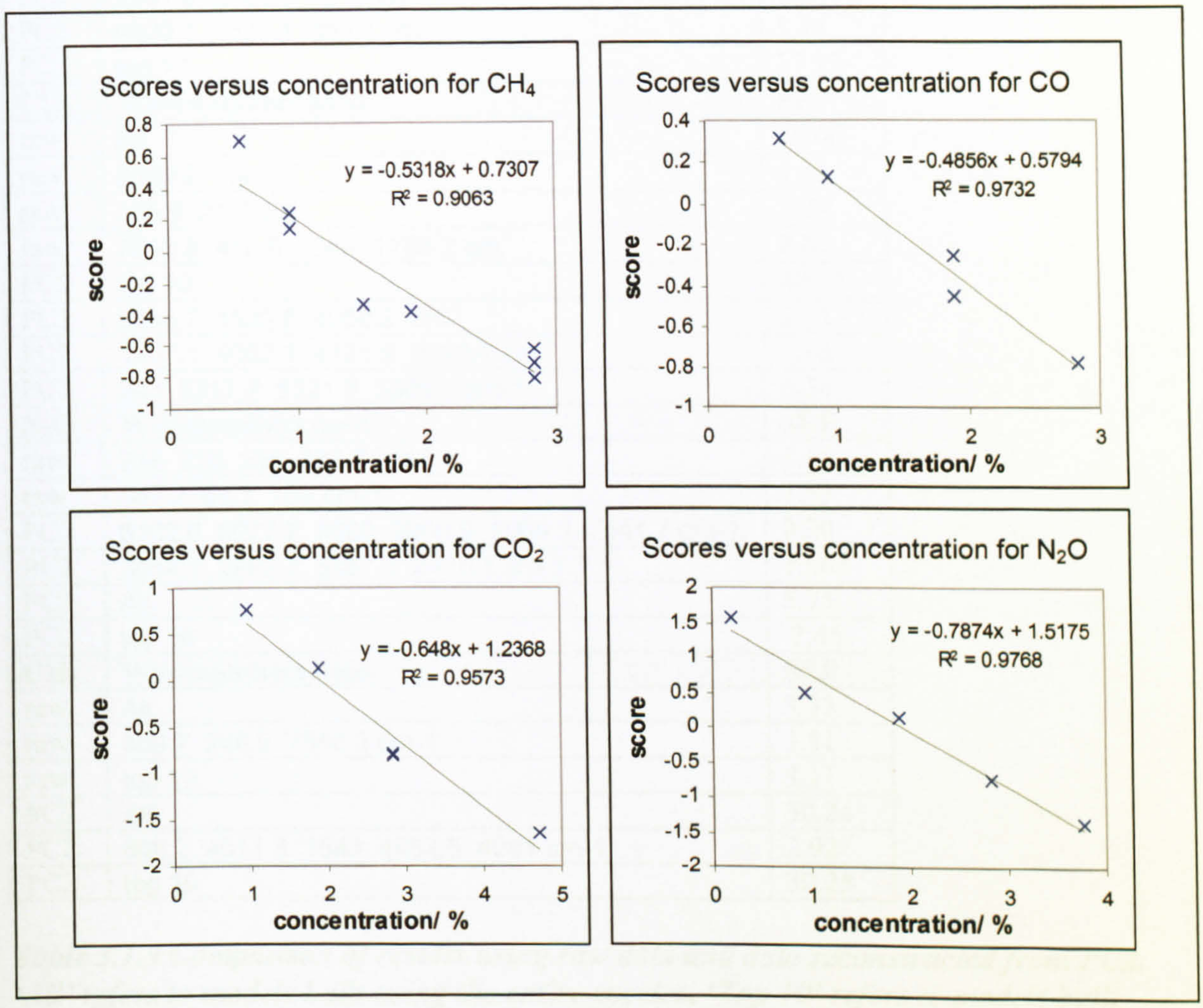


Figure 3.1.13 Sample scores for principal component two versus the corresponding sample concentration (%)

Table 3.1.9 gives comparisons for some of the results obtained from variable selection on raw data and variable selection on the data reconstructed from PC2. Included are results from models employing all the variables and from models using the top 10 variables that gave the best correlations for the gas of interest.

Direct comparisons are difficult to make as the results obtained depend on which wavenumbers were selected, therefore the average SEP was used as a general indication of performance. In almost all cases, the SEP for the reconstructed PC2 data was worse than with raw data; however, the wavenumbers selected from the PC2 data were closer to absorption regions.

CO	Wavenumbers used	SEP
raw	All	34.61
raw	163.9 302.8 38.6 cm^{-1}	0.57
raw	167.8 177.5 cm^{-1}	0.17
raw	top 10	4.60
PC2	All	107.74
PC2	2931.8 2208.5 1219 cm^{-1}	1.14
PC2	4820.1 4629.1 883.4 cm^{-1}	3.44
PC2	top 10	17.23
CO_2	**Wavenumbers used**	**SEP**
raw	All	19.06
raw	2360.9 3.9 cm^{-1}	1.51
raw	165.9 21.2 cm^{-1}	2.29
raw	7618.8 416.6 7046 1186.2 cm^{-1}	2.62
PC2	top 10	40.33
PC2	5346.7 4530.8 4050.5 cm^{-1}	5.43
PC2	1107.1 4087.1 4121.9 3990.7 cm^{-1}	7.16
PC2	243 5317.7 5321.6 5308.1 cm-1	6.34
N_2O	**Wavenumbers used**	**SEP**
raw	218 216 254 333.7 cm-1	1.16
raw	347.2 61.7 108 cm-1	2.53
PC2	6162.6 6077.7 5690 2900.9 1969.3 7541.7 cm-1	0.20
PC2	3954.1 2962.7 3487.3 3410.1 cm-1	10.63
PC2	All	5.75
PC2	top 10	7.45
CH_4	**Wavenumbers used**	**SEP**
raw	All	3.33
raw	360.7 246.9 7848.3 cm-1	3.51
raw	top 10	3.11
PC2	All	30.24
PC2	806.2 4611.8 1543 4862.5 4091 cm-1	2.92
PC2	top 10	30.25

Table 3.1.9 Comparison of results using raw data and data reconstructed from PC2. 'All' refers to models built using the entire spectra, 'Top 10' refers to models built using measurements from the top ten wavenumbers of highest absorption, otherwise the wavenumbers used were from variable selection models

3.1.1.4 Work using reference peaks

CH_4 employed as an internal reference for CO samples

To compare the deviation of response without the influence of background variation, the spectra were converted to absorption spectra by ratioing against scans obtained when the cell was purged between samples. The plot of the % RSD (figure 3.1.14) over different wavenumbers shows the RSD for CH_4 absorption to be less than 2 %. As expected, the highest deviations are obtained at wavelengths relating to the CO absorption with the magnitude directly related to the molar absorptivity for the given wavenumber.

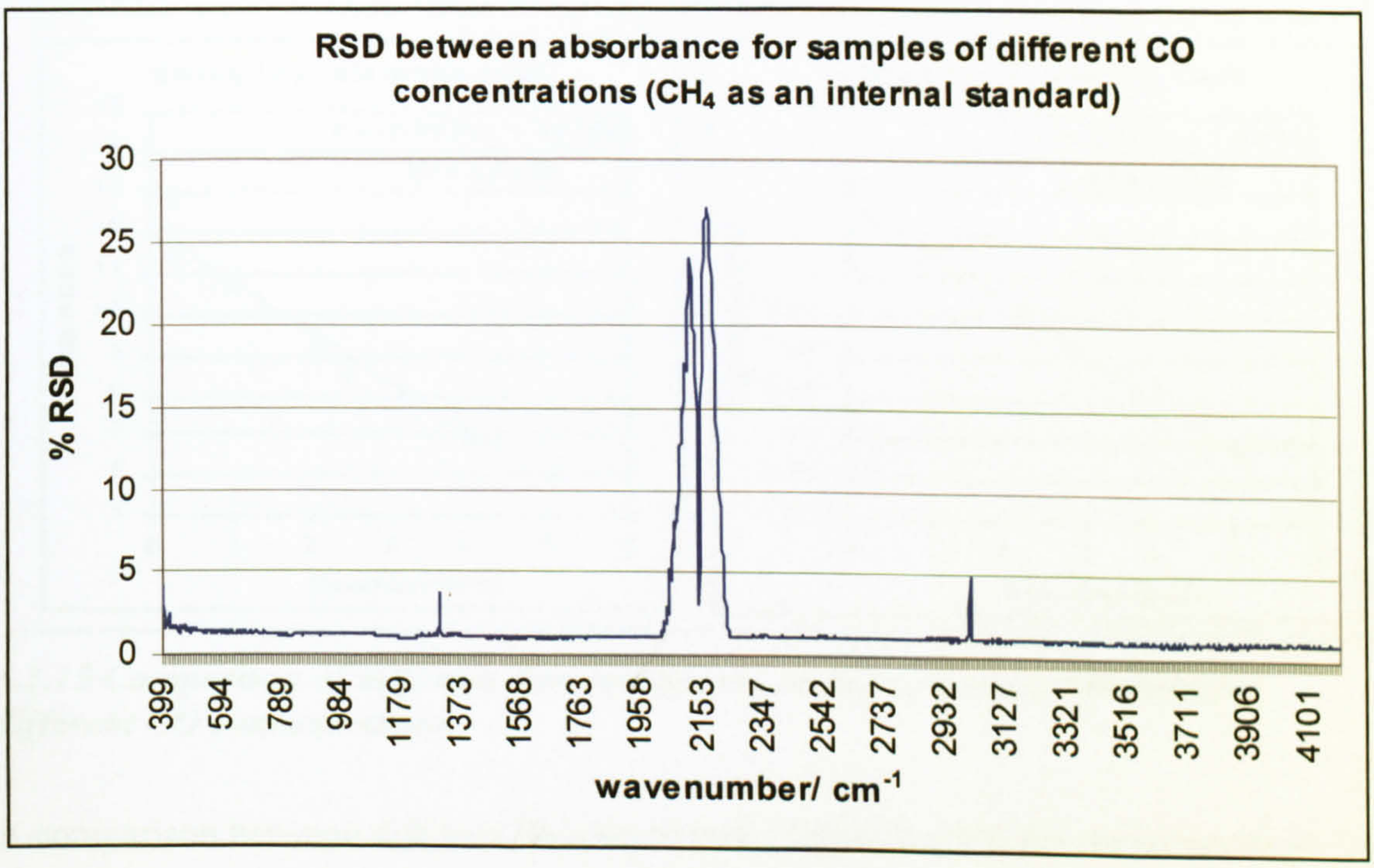
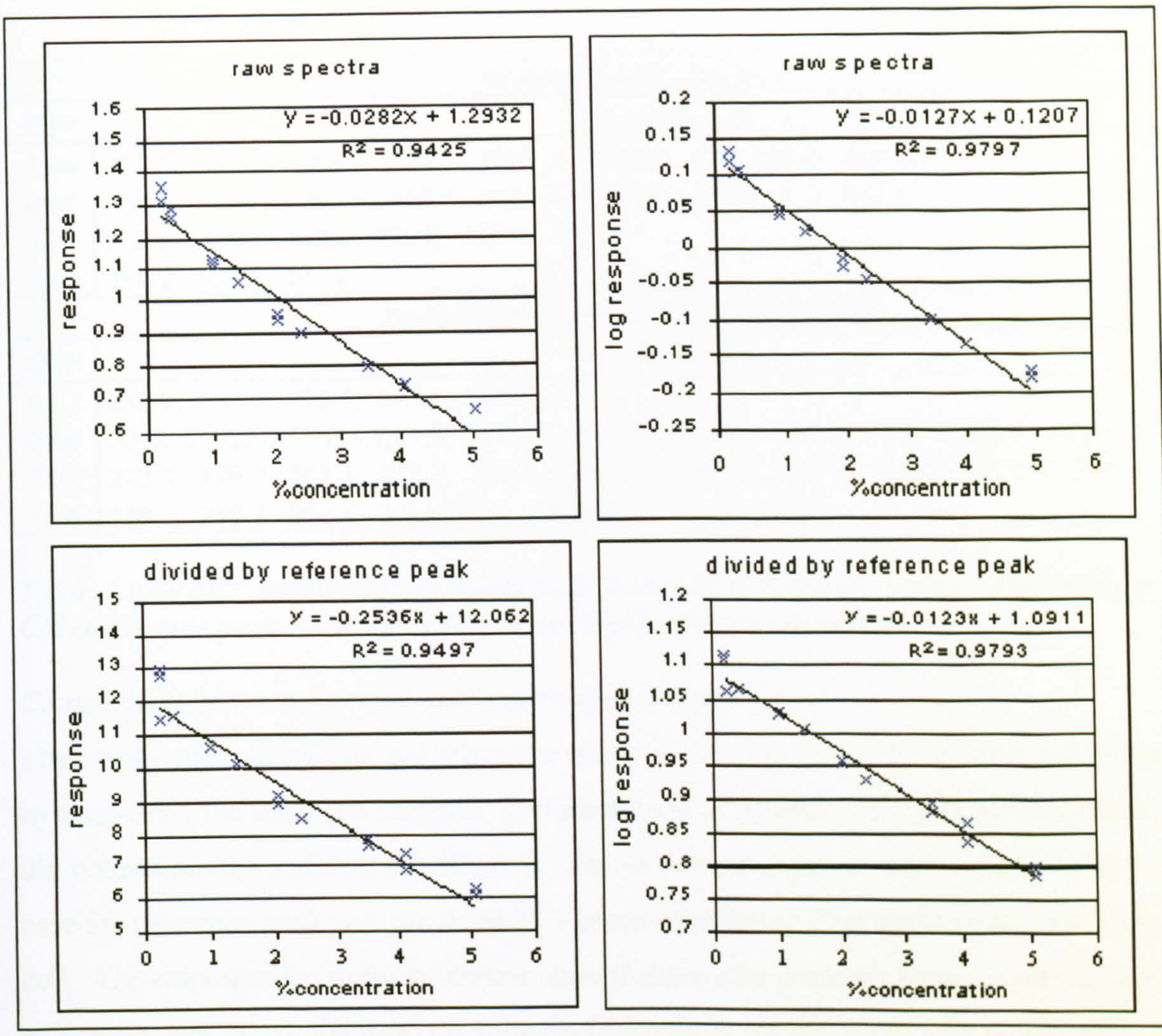


Figure 3.1.14 Standard deviation for absorption spectra for samples employing CH_4 as an internal reference



3.1.15 Comparison of different data treatments on the responses obtained for different CO concentrations

A comparison between different data treatments (figure 3.1.15) for the response at wavenumber 2200 cm^{-1} shows that similar correlations were obtained between response and concentration when the data was divided by the reference peak, and for that between concentration and untreated data. The best correlations were obtained when logs were taken of the responses.

Variable selection was then performed on the data to further assess the effect of dividing by a reference peak.

Results from the selection models gave variable results; some of the results are shown in table 3.1.10. In many of the cases, wavenumbers <500 and >5000 cm^{-1} were being selected. The lowest SEP obtained was 0.28, this occurred when the data was divided by the CH_4 reference peak and then normalised, this model included three wavenumbers directly related to CO absorption i.e. in the range 2000-2300 cm^{-1}. This was regarded as a promising result so further investigations were carried out employing reference peaks.

Divided by CH_4 peak	
SEP	**Wavenumbers**
1.84	7697.9, 364.5, 358.8, 229.5, 7784.7, 7809.8, 3.9, 366.5, 381.9, 1.9
0.97	356.8, 7736.5, 366.5, 358.8, 6654.4, 6669.8, 229.5, 1.9, 6473.1, 7271.6, 7734.5, 3.9
0.4	312.5, 326.0, 192.9, 229.5, 300.9, 358.8
3.43	354.9, 358.8, 163.9
Divided by CH_4 peak and normalised	
SEP	**wavenumbers**
0.67	358.8, 356.8, 229.5, 312.5, 326.0
3.06	3012.8, 358.8, 7611.1, 1.9, 7719.1, 368.4, 7829.0, 7645.8, 3.9, 7699.8, 229.5, 7790.5
0.28	273.9, 299.0, 241.1, 277.7, 358.8, 2081.2, 2085.0, 2139.1, 345.3, 2210.4, 229.5
1.08	356.8, 212.2, 358.4, 3.9, 229.5, 7676.7

Table 3.1.10 SEP obtained for models built from CO calibration spectra divided by a CH_4 reference peak, with and without normalisation pre-treatment

Comparison between internal and external references

The employment of an external reference was compared to that of an internal reference by measuring the same gas samples. The advantage of an external reference was that it did not require the addition of a fixed amount of reference gas to each sample. In this case the reference peak was provided by a piece of polymer film placed next to the gas cell. The response due to the reference should always be constant, hence it can be used as a 'reference' to compensate for changes in response arising from such factors as changes in source emission. The disadvantage with an external reference is that it can not compensate for variations due to the sample itself (for example errors in the amounts of gas injected) and changes within the gas cell e.g. a slight change in mirror alignment. The advantages of an internal reference is that it can be used to compensate for changes within the cell itself, as well as for external changes in room temperature and pressure. The gas samples were prepared by injecting fixed volumes of gases into the gas cell at room temperature and pressure, and to simplify the calculations it was generally assumed that the temperature and pressure were effectively constant when converting these amounts into the corresponding concentrations. Variations in room pressure and temperature would also affect the concentration of the internal standard added, thus allowing these effects to be compensated for; however, the amount of internal standard injected is still subject to the same experimental errors as the other gases.

Table 3.1.11 lists results obtained for variable selection runs using raw data and for data divided by the N_2O reference.

Spectra without Polymer Film		
Data	**Wavenumbers**	**SEP**
Raw	All	0.51
Raw	150 selected	0.21
Raw	44 selected	0.16
Divided by N_2O peak at 1304 cm^{-1}	All	0.69
Divided by N_2O peak at 1304 cm^{-1}	2735, 3011, 2399, 3019 cm^{-1}	0.36
Divided by N_2O peak at 2579 cm^{-1}	All	0.39
Divided by N_2O peak at 2579 cm^{-1}	3028, 3015, 2351, 3007, 2401 cm^{-1}	0.13
Divided by N_2O peak at 2579 cm^{-1}	1196, 1263, 417, 1076, 472 cm^{-1}	0.09
Spectra with Polymer Film (same training and validation spectra):		
Data	**Wavenumbers**	**SEP**
Divided by peak at 758 cm^{-1}	All	0.59
Divided by peak at 758 cm^{-1}	3221, 2825, 2574, 3017, 2097 cm^{-1}	0.22
Divided by peak at 758 cm^{-1}	2814, 2097, 2353, 2824, 2575 cm^{-1}	0.19
Divided by peak at 1030 cm^{-1}	All	0.59
Divided by peak at 1030 cm^{-1}	1653, 2515, 2758, 2262, 2031 cm^{-1}	0.21
Divided by peak at 1030 cm^{-1}	3628, 2758, 2031, 2951, 2509 cm^{-1}	0.22

Table 3.1.11 SEP obtained for models built from CH_4 calibration spectra (plus N_2O as an internal standard) with and without polymer film

Variable selection on the raw spectra resulted in a large number of wavenumbers selected. Dividing the spectra by the N_2O transmission response at 1304 cm^{-1} gave worse predictions than the raw data; however, dividing the data by the N_2O transmission response at 2579 cm^{-1} improved the errors in prediction and smaller numbers of variables were selected.

Results with an external reference gave better predictions than the models built from spectra divided by the N_2O response at 1304 cm^{-1}, but not as low as those for data divided by the response at 2579 cm^{-1}.

CO_2 and N_2O samples with SF_6 as an internal standard

Results from the models built from CO_2 spectra are shown in table 3.1.12, and the results for N_2O spectra are shown in table 3.1.13.

CO_2 calibration using SF_6 as an internal standard		
Data used	**Wavenumbers**	**SEP**
Raw *	102 selected	0.79
Raw *	5512.5, 4258.8, 4476.8, 4586.7, 4820.1, 4773.8 cm^{-1}	0.66
Raw *	565 selected	0.99
Raw #	4104.5, 4750.7, 4771.9, 4287.7, 4675.4, 4604.1, 4476.8 cm^{-1}	0.53
Raw #	3730.3, 3639.7, 3255.8, 3809.4, 3736.1, 3512.4 cm^{-1}	0.54
Divided by SF_6 peak *	4384.2, 5146.1, 4594.4, 4820.1, 4634.9, 4474.8 cm^{-1}	0.49
Divided by SF_6 peak *	144 selected	0.70
Divided by SF_6 peak *	406 selected	0.93
Divided by SF_6 peak #	3938.6, 4038.9, 3608.8, 3639.7, 3689.8, 3512.4 cm^{-1}	0.59
Divided by SF_6 peak #	4135.4, 6226.2, 6679.5, 6696.8, 6532.9, 6585.0 cm^{-1}	0.55

* using training spectra 1,3,4,5,7 and validation spectra 2,6,8,9

using training spectra 2,4,7,8,9 and validation spectra 1,3,5,6

Table 3.1.12 SEP obtained for models built from CO_2 calibration spectra plus SF_6 as an internal standard

N_2O calibration using SF_6 as an internal standard		
Data used	**Wavenumbers**	**SEP**
Raw	212 selected	3.21
Raw	2125.6, 3487.3, 3358.1, 3844.1, 1967.4, 3388.9 cm^{-1}	3.3
Raw	1423.5, 3728.4, 3728.4, 3475.7, 3410.1, 3385.1, 3851.8 cm^{-1}	2.69
Divided by SF_6 peak	4181.7, 1975.1, 4521.1, 4810.5, 4374.5, 3441.0 cm^{-1}	1.67
Divided by SF_6 peak	4403.5, 4887.6, 4515.3, 3660.9, 4806.6, 3441.0 cm^{-1}	3.07
Divided by SF_6 peak	2522.9, 2607.8, 4351.4, 4777.7, 5367.9, 4820.1, 4353.3 cm^{-1}	2.84
Divided by SF_6 peak	6218.5, 4521.1, 4499.9, 2623.2, 7360.7, 4806.6, 3441.0 cm^{-1}	3.38

Table 3.1.13 SEP obtained for models built from N_2O calibration spectra plus SF_6 as an internal standard

In this case, the results obtained for the CO_2 and N_2O models gave similar errors in prediction, thus at this stage it was inconclusive as to whether the addition of an internal standard was advantageous. Previous work employing N_2O and CH_4 suggested results could be improved, and this had led to the investigations using SF_6.

SF_6 was chosen as an internal standard as none of its absorption peaks overlapped with CO, CO_2, CH_4 or N_2O. SF_6 had no adverse effect on the spectra; hence it continued to be added to gas samples while further investigations were carried out on background variation compensation.

3.1.1.5 Conclusion

The aim of this stage of the work was to determine the most effective data pre-treatment to apply to the spectra prior to building prediction models.

From the analysis of the raw data it was clear some means of correcting for background/baseline variation was required. The approaches investigated were Savitsky-Golay smoothing, reconstruction of the data using PCA and the addition of internal and external standards. To judge the effectiveness of the treatments, comparisons were made between the prediction errors (SEP) of MLR and VS-MLR models built from the data with different pre-treatments applied. As a general rule, applying MLR to the entire spectra will result in poor prediction errors due to co-linearity and singularity problems (as discussed in section 1) therefore lower prediction errors may not necessarily be a true indication of better data, hence VS-MLR results were also compared. The drawback with comparing VS-MLR results was that the SEPs obtained from the prediction models were determined by the final wavenumbers selected, which in turn depended on the performance of the VS algorithm.

The performance of the VS algorithm depended on the settings, particularly what values were assigned to the convergence levels, (which links back to the amount of error and type of data employed). Thus further investigations were carried out in the next stage of the work focusing on the effect of changing the different settings on the VS-MLR program.

To determine whether the data treatment was effective, not just the SEPs were compared, but also the number of wavenumbers selected, and whether they were reasonable i.e. if they corresponded to regions of absorption for the analyte of interest. Comparing the actual wavenumbers selected showed that the Savitsky-Golay and PCA reconstruction gave no significant improvement over the raw data, hence these treatments were no longer applied.

The employment of an internal reference gave improved results in some cases, and little difference in others, so more investigation was required, hence an internal standard was added to further gas mixtures.

The transmission data was not compared against absorption spectra (i.e. spectra obtained by ratioing the scans against background scans), therefore absorption spectra were collected in subsequent work to allow further comparisons of different data treatments.

Another approach not previously investigated was the application of PLS models as opposed to MLR and VS-MLR models, subsequently this was also included in the next stage of the work.

3.1.2 Mixtures of the four gases plus SF_6 as an internal standard

3.1.2.1 Appearance of spectra and effect of division by reference peak

The raw single beam spectra obtained for the different sample mixtures is shown in figure 3.1.16 and the corresponding absorption spectra in figure 3.1.17. As with the previous calibration spectra, variations in the energy transmission can be clearly seen in the raw single beam spectra, hence division by the reference SF_6 peak was applied.

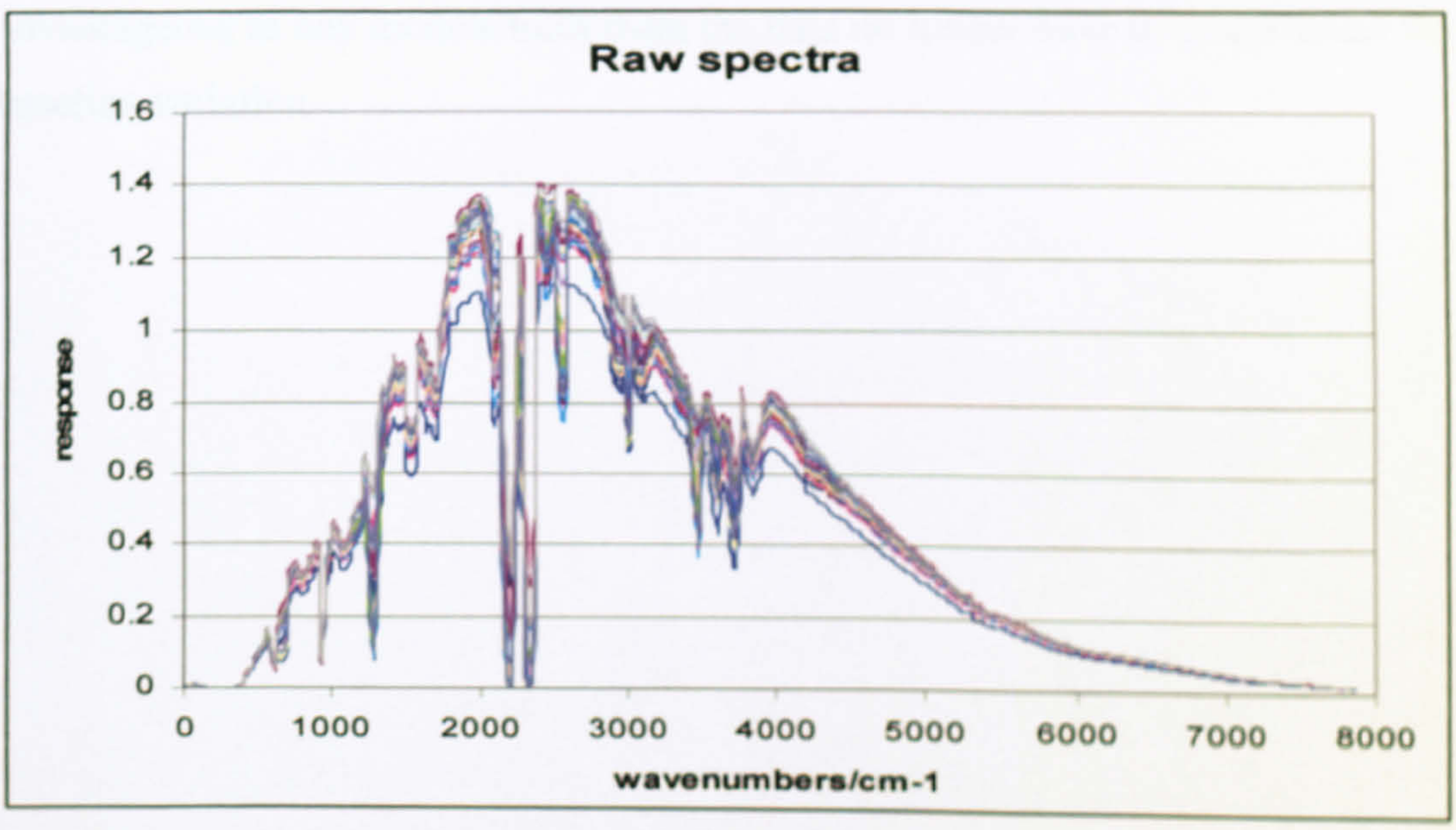


Figure 3.1.16 Raw single beam spectra for mixtures of CH_4, CO, CO_2 and N_2O

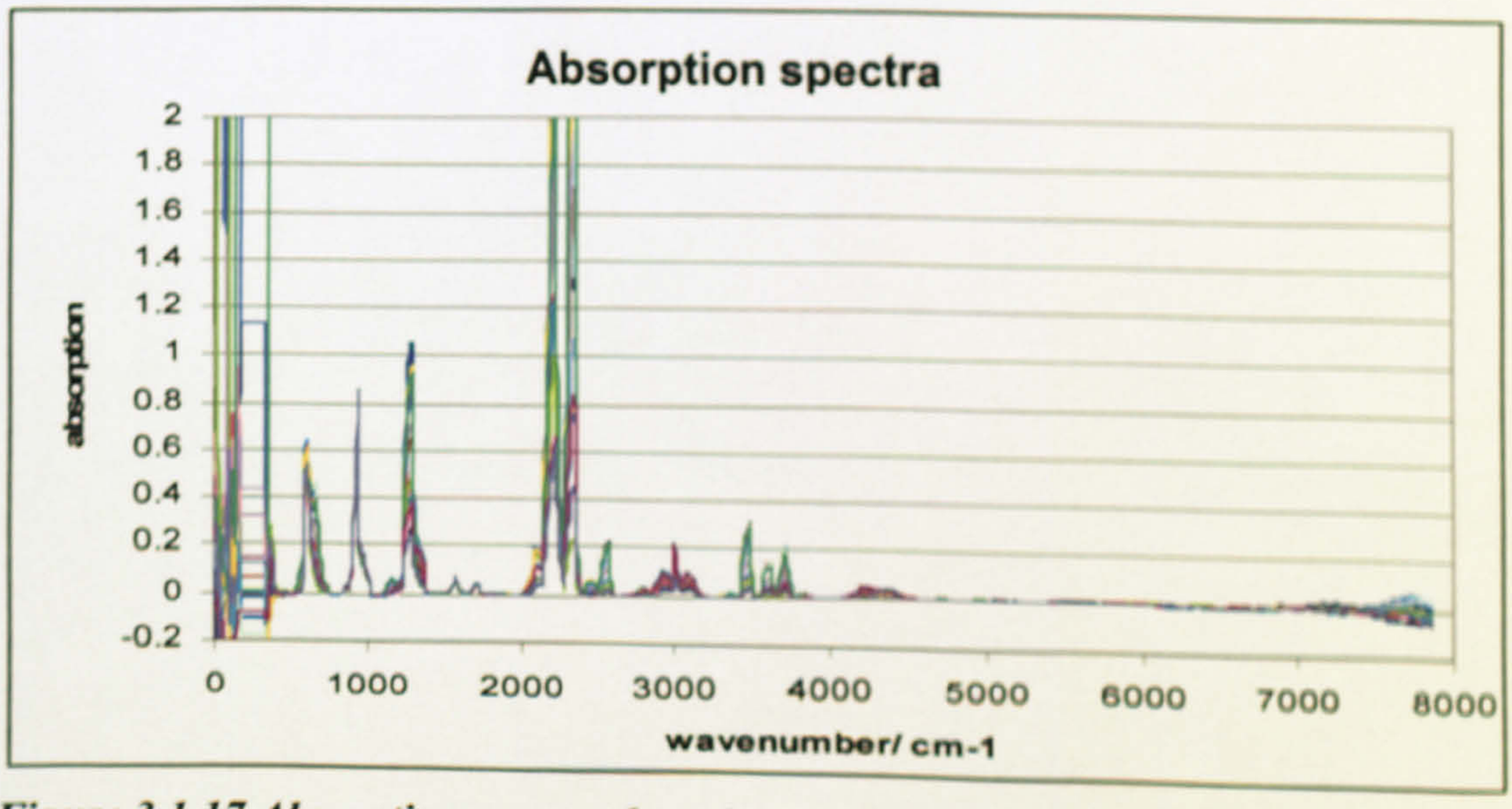


Figure 3.1.17 Absorption spectra for mixtures of CH_4, CO, CO_2, and N_2O

Figure 3.1.18 compares the effect of dividing the spectra by a reference point and normalising. Dividing by the reference point increased the responses by a factor of 10, but still left a high degree of variation in the baseline. With normalisation the point of highest transmission is set to 1 and regions with no transmission set to 0, thus effectively shifting the baselines to the same magnitude. Focusing on the region where the main absorption peaks occur, i.e. 1500-4000 cm^{-1}, (figure 3.1.18) clearly demonstrates the effectiveness of this treatment in removing baseline variations. Comparing the treated spectra with the original raw data, it can been seen that changes in the overall transmission no longer have a significant effect on the spectra and variations due to absorptions by the different gases are accentuated. This is advantageous as any models built from the data no longer have to compensate for baseline variation.

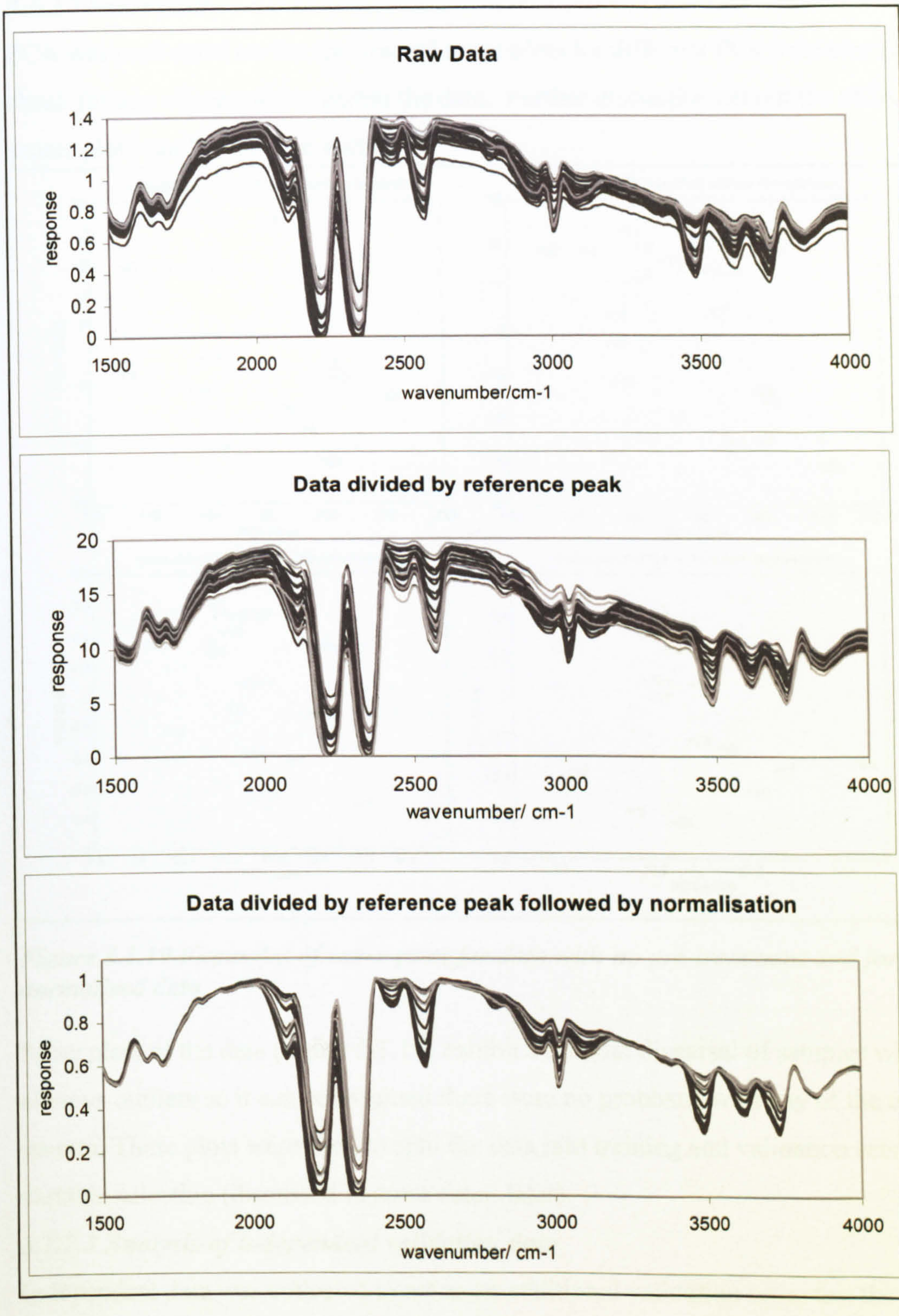


Figure 3.1.18 The effect of data pre-treatment on spectra obtained for mixtures of CO, CO_2, CH_4 and N_2O

3.1.2.2 Score plots

PCA was performed on the spectra and score plots for different PCs were examined to check for any abnormalities within the data. Further discussions about the utilisation of scores plots can be found in section 3.7.

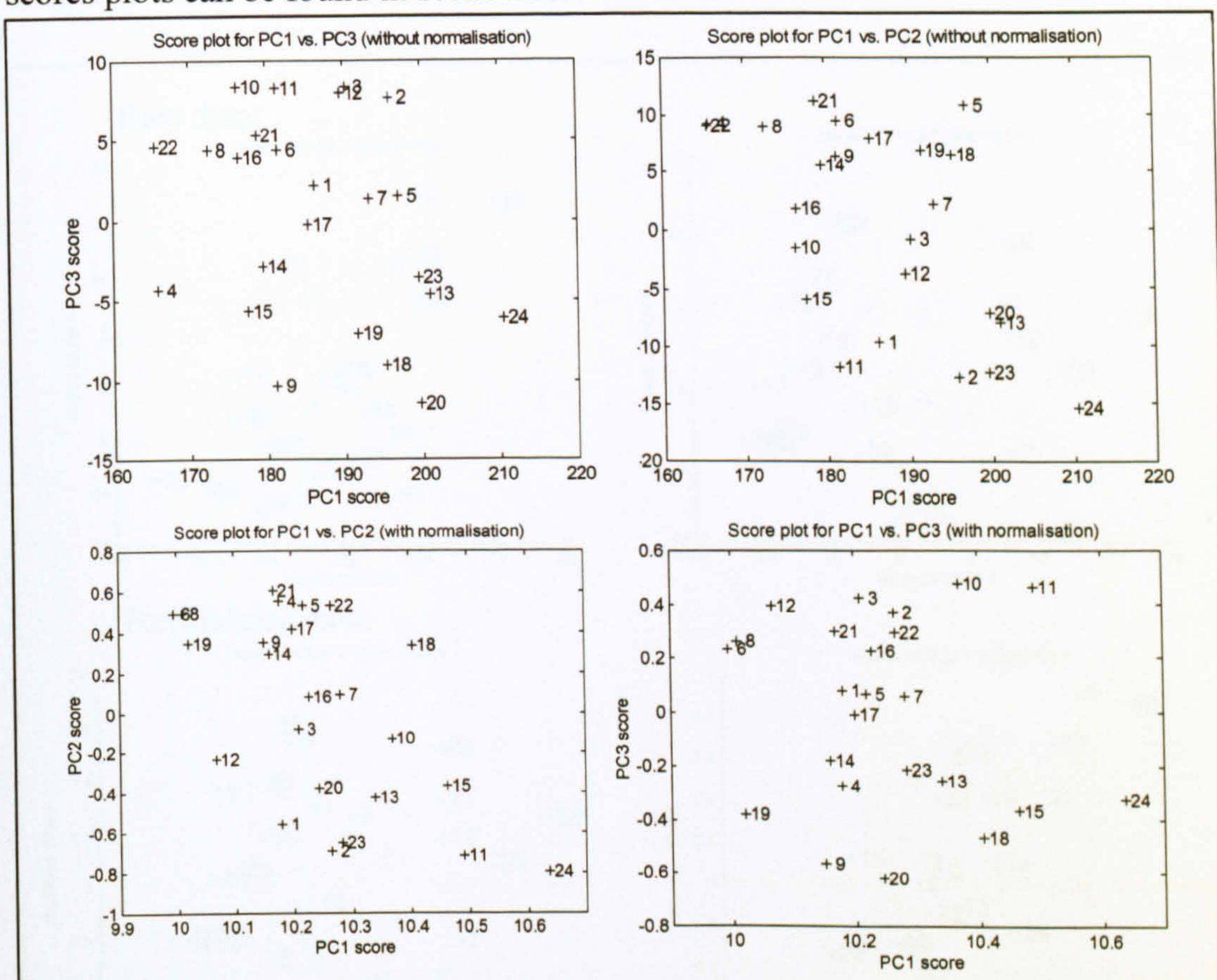


Figure 3.1.19 Examples of score plots for data with no pre-treatment and for normalised data

Score plots of the data (figure 3.1.19) exhibit a random dispersal of samples with no obvious outliers so it can be assumed there were no problems with any of the collected spectra. These plots were used to split the data into training and validation sets for the variable selection (discussed in more detail later).

3.1.2.3 Analysis of independent validation data

Independent data was collected to act as an additional validation set to test the model's abilities to predict new data. PCA was performed on all the data including the new validation data (collected two weeks later).

The initial data had been collected over a period of three days. The new data was collected a couple of weeks later so PCA was performed on all the data to see if there was any significant difference between the data, ie. instrumental drift, artifact peaks, the amount of water vapour, changes in the gas cell mirror alignment, etc.

Score plots for the raw and normalised data are shown in figure 3.1.20. Samples 1 to 6 are the new samples, as these are randomly dispersed in the score plot there is no significant difference between the old and new data. Any major differences would have resulted in clustering of the new samples from the rest.

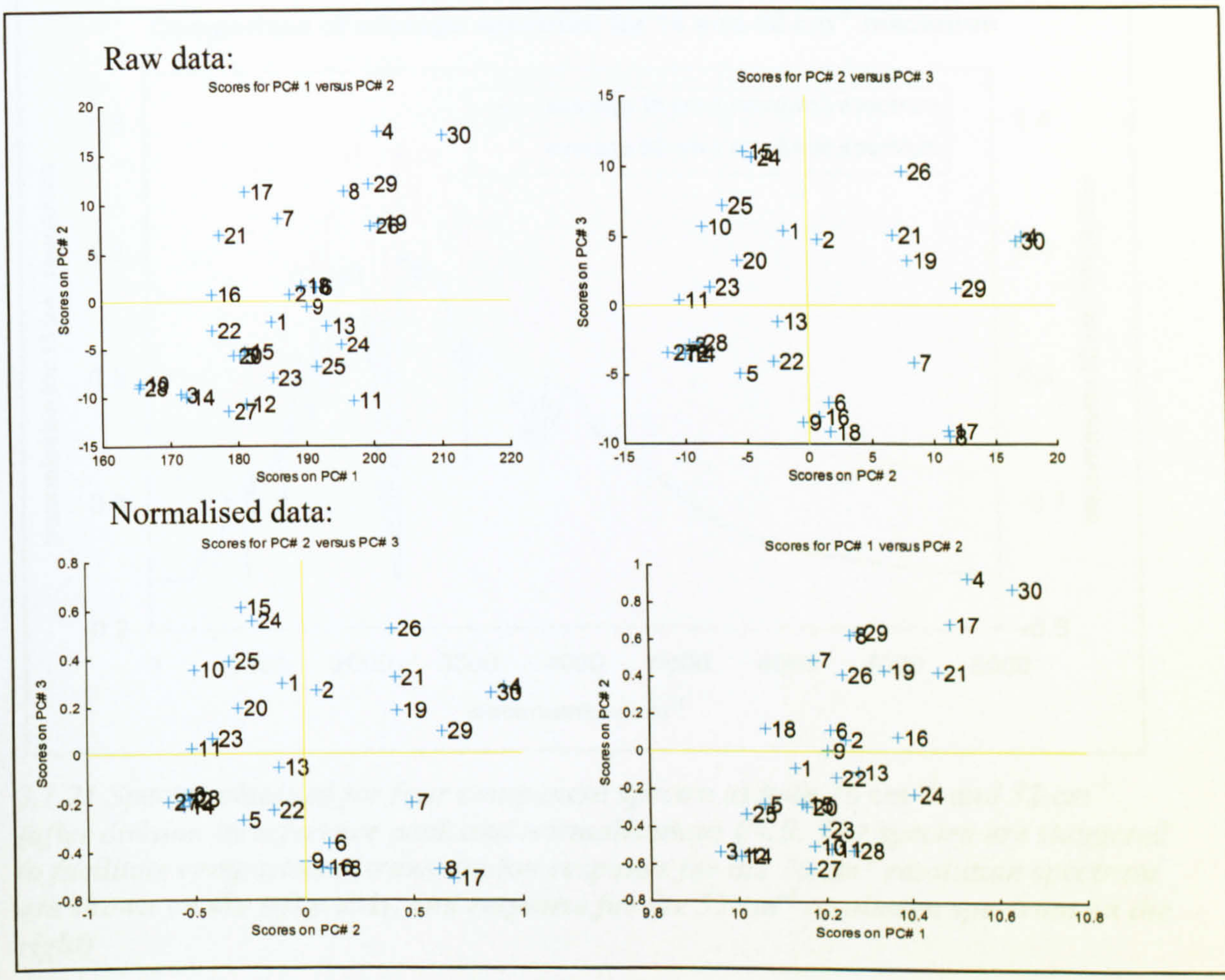
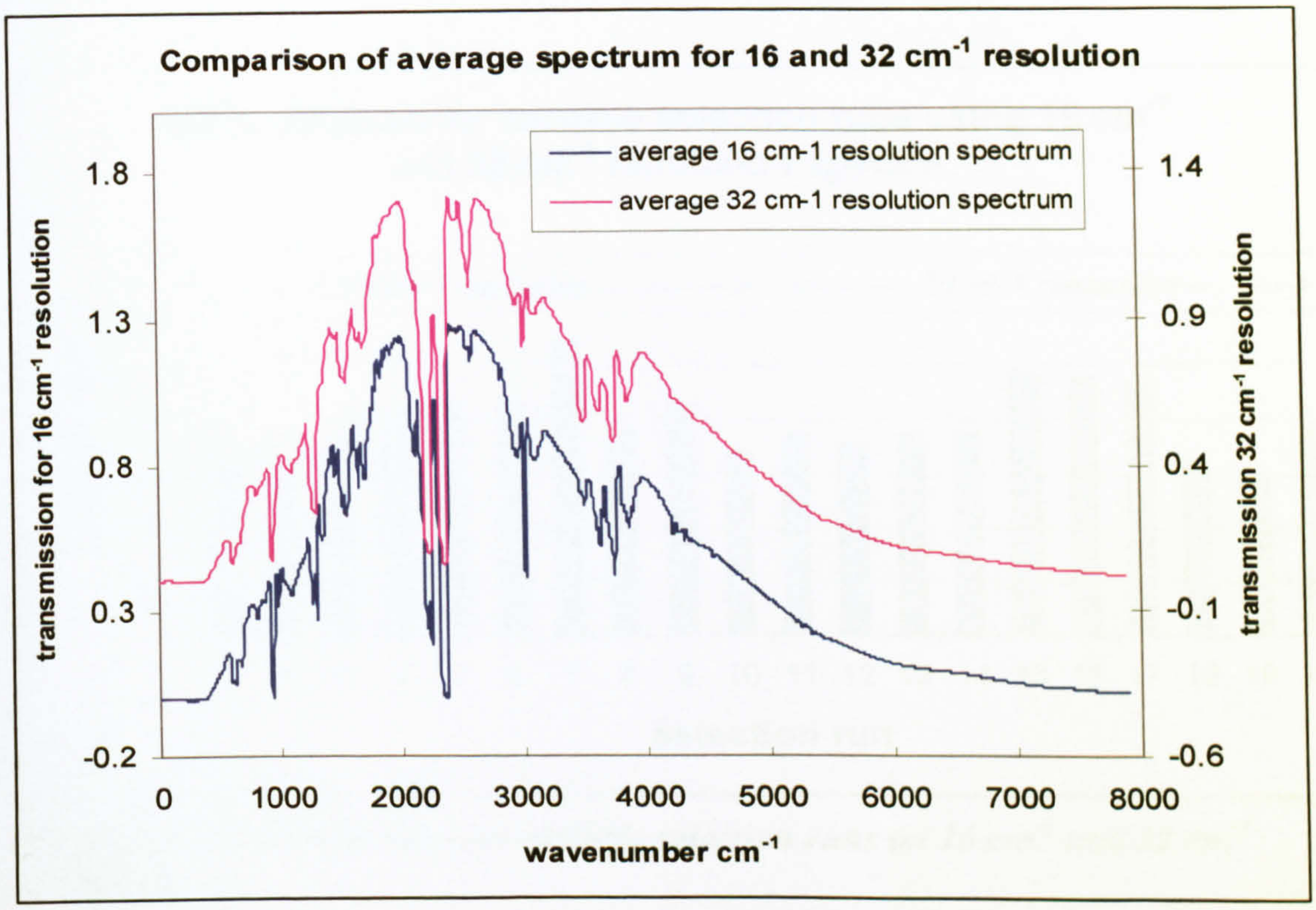


Figure 3.1.20 Score plots with the independent validation samples (samples 1-6)

3.1.2.4 Effect of resolution

An important consideration in infrared spectroscopy is the resolution, this refers to the range of wavenumbers over which each measurement is made. At high resolution, i.e. less than 2 cm^{-1}, each measurement is made over a very narrow range of wavenumbers; at low resolution each measurement covers a larger range of wavenumbers. With gases, absorption bandwidths are narrow (less than 1 cm^{-1}), hence higher signal to noise ratios are obtained using higher resolutions. For the individual component spectra, the highest resolution possible with the FTIR instrument was employed; this corresponded to a resolution of 2 cm^{-1}. However, for the detectors within the planned array the bandwidth measured will be much broader with bandwidths closer to 30 cm^{-1}. To allow for this,

the resolution of the spectra collected for the quaternary gas mixtures was reduced. To compare the effect of reducing resolution, spectra were collected at low resolution using both 16 cm^{-1} and 32 cm^{-1} resolution. Figure 3.1.21 compares the average spectra obtained for 16 cm^{-1} and 32 cm^{-1} resolution.



3.1.21 Spectra obtained for four component spectra at both 16 cm^{-1} and 32 cm^{-1} (after division by reference peak and normalisation) (N.B. The spectra are staggered to facilitate comparisons, transmission response for the 16 cm^{-1} resolution spectrum are shown on the left y-axis, and response for the 32 cm^{-1} resolution spectrum on the right)

At 16 cm^{-1}, measurements were effectively the average response for a window of 15.4 cm^{-1}, with the centre of each window every 7.7 cm^{-1}. This resulted in 1024 measurements between 0 and 8000 cm^{-1}. At 32 cm^{-1} resolution, the windows are 30.9 cm^{-1}, with the centre every 15.4 cm^{-1} giving a total of 512 measurements. (NB. The reference peak used occurred at 941 cm^{-1}; for the 32 cm^{-1} resolution spectra this corresponded to measurement number 62, for 16 cm^{-1} resolution data, the same peak corresponded to measurement number 130.) Figure 3.1.22 displays the average SEP obtained for the four gases from 10 different selection runs for both resolutions, employing the same modelling conditions i.e. the same training and validation split and selection coefficients (the effect of these coefficients is discussed in section 3.2). Fluctuations in the SEP obtained for different runs arose due to different final wavenumbers selected; however, the main conclusion drawn was that prediction models

made with the 16 cm^{-1} and 32 cm^{-1} spectra gave similar standard errors in prediction. It was decided to focus on the lower resolution spectra as the resolution would be closer to the bandwidth of the filters to be used in the detector array and further investigations were carried out on the 32 cm^{-1} resolution, the results of which are discussed in section 3.2.

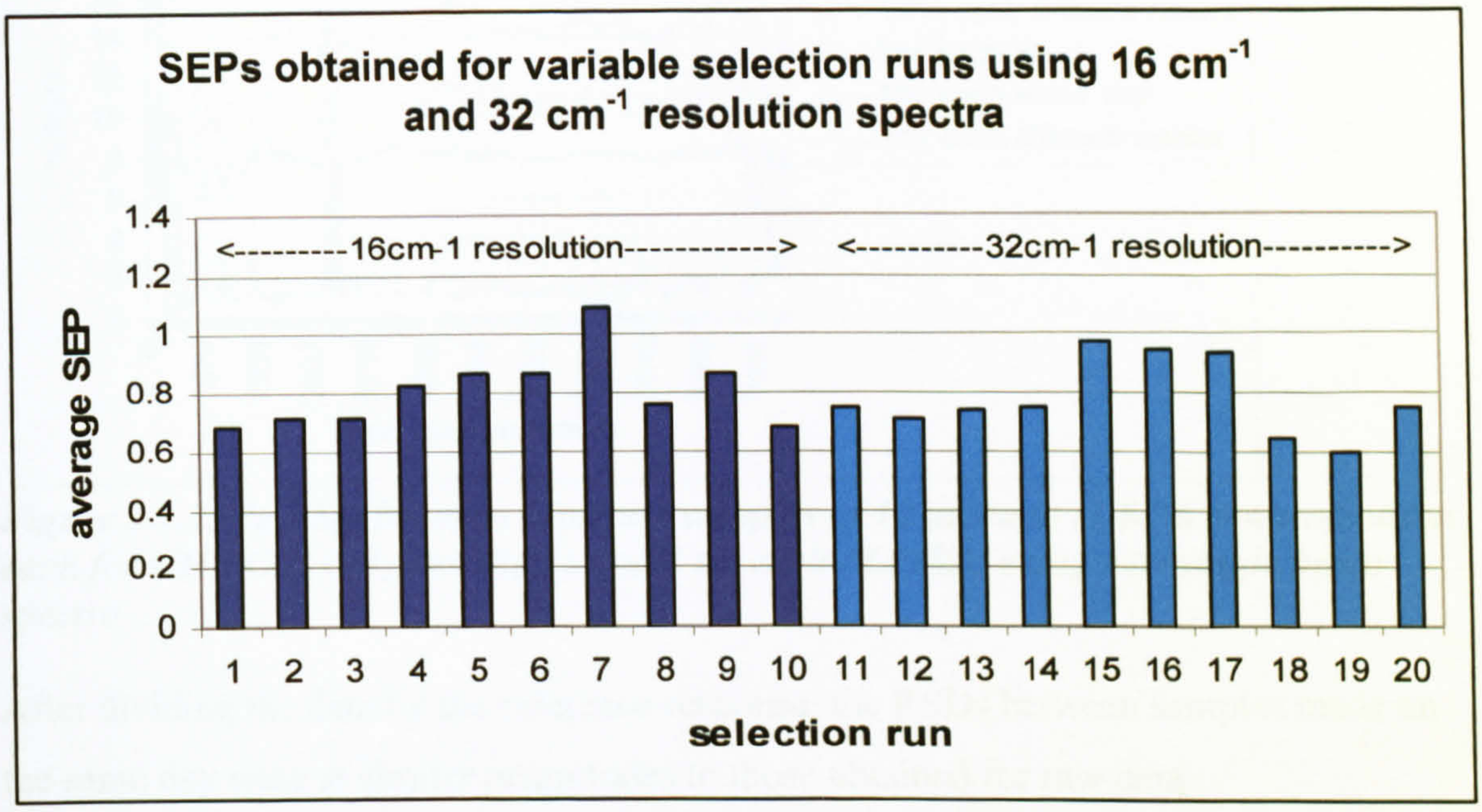


Figure 3.1.22 SEP obtained for variable selection runs on 16 cm^{-1} and 32 cm^{-1} resolution

3.1.2.5 Reproducibility of spectra

The RSDs between samples made on different weeks and between samples made on the same day were compared for both raw data and scans divided by the SF_6 reference peak (fig. 3.1.23). 'Same day' refers to the three samples made on the same day, 'different weeks' refers to the comparison of spectra collected on different weeks (with one spectra from each week), 'all' includes all the replicate samples i.e. the three samples made on the same day plus the two samples made on different weeks.

The regions with high RSD values (>100 %) occur at wavenumbers where the amount of energy transmitted was close to zero. Wavenumbers less than 500 cm^{-1} corresponded to the cut off region and high RSDs between wavenumbers 2100-2200 cm^{-1} was due to large CO_2 absorption peaks arising from atmospheric levels of CO_2 in the air surrounding the gas cell in the sample chamber.

Ignoring the aforementioned regions, the RSD between raw spectra for samples made on the same day (both the 0.19 % (1 ml) each and the 1.88 % (10 ml) each samples) averaged 1.2 % but was higher in absorption regions of interest. This increased to over

10 % RSD (11 % for the 1.88 % each samples and 14 % for the 0.19 % each samples) when comparing spectra from different weeks (see figure 3.1.23).

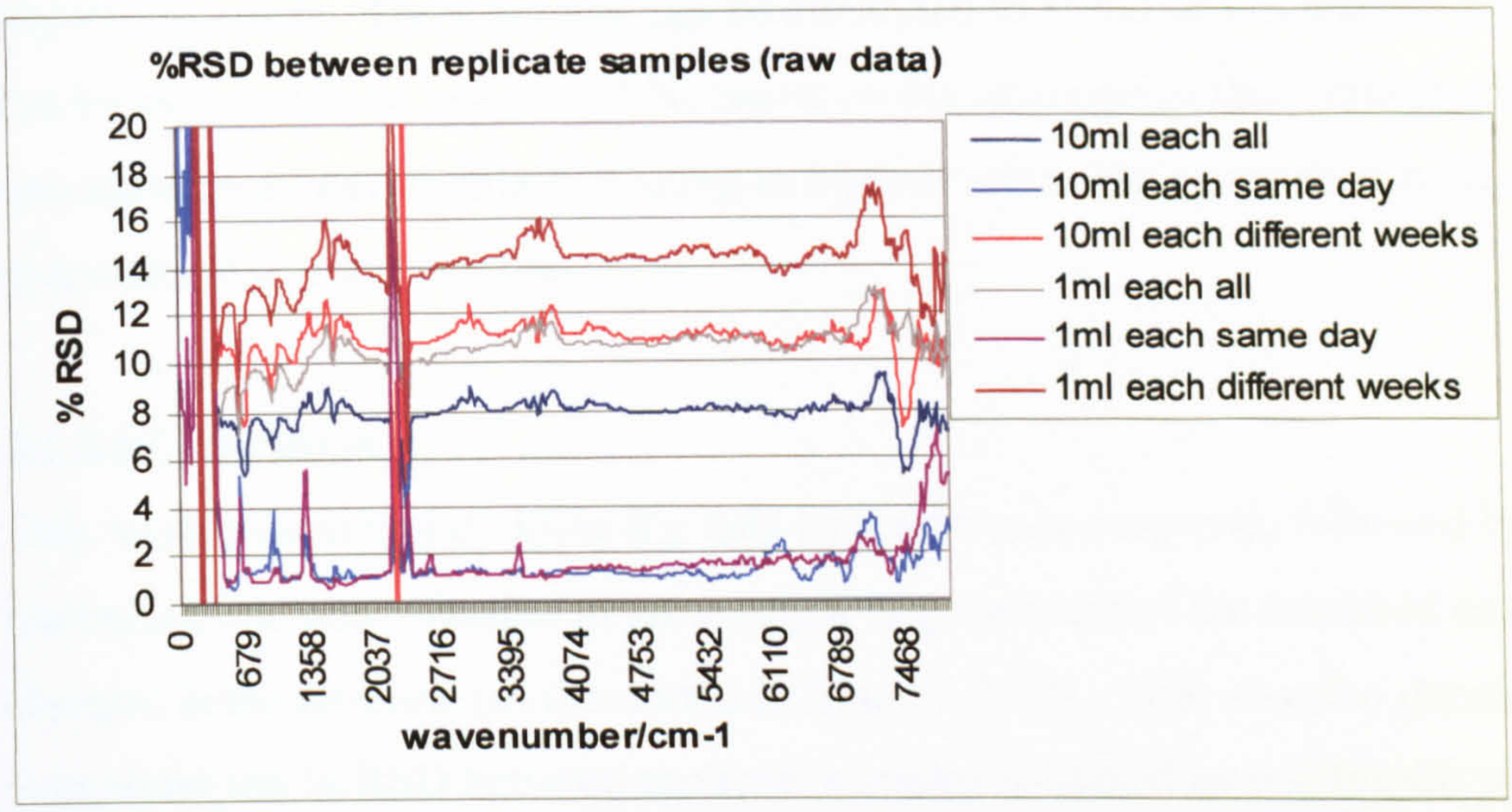


Figure 3.1.23 % RSD between replicate samples of 10 ml each (1.88% concentrations each for CH_4, CO, CO_2 and N_2O) and 1 ml each (0.19%), using raw single beam spectra

After dividing the data by the reference response, the RSDs between samples made on the same day were at similar magnitudes to those obtained for raw data.

The effectiveness of division by the reference peak can be seen when looking at the RSD between samples made on different weeks, where there is no longer any significant difference between same day and different week spectra (refer to figure 3.1.24).

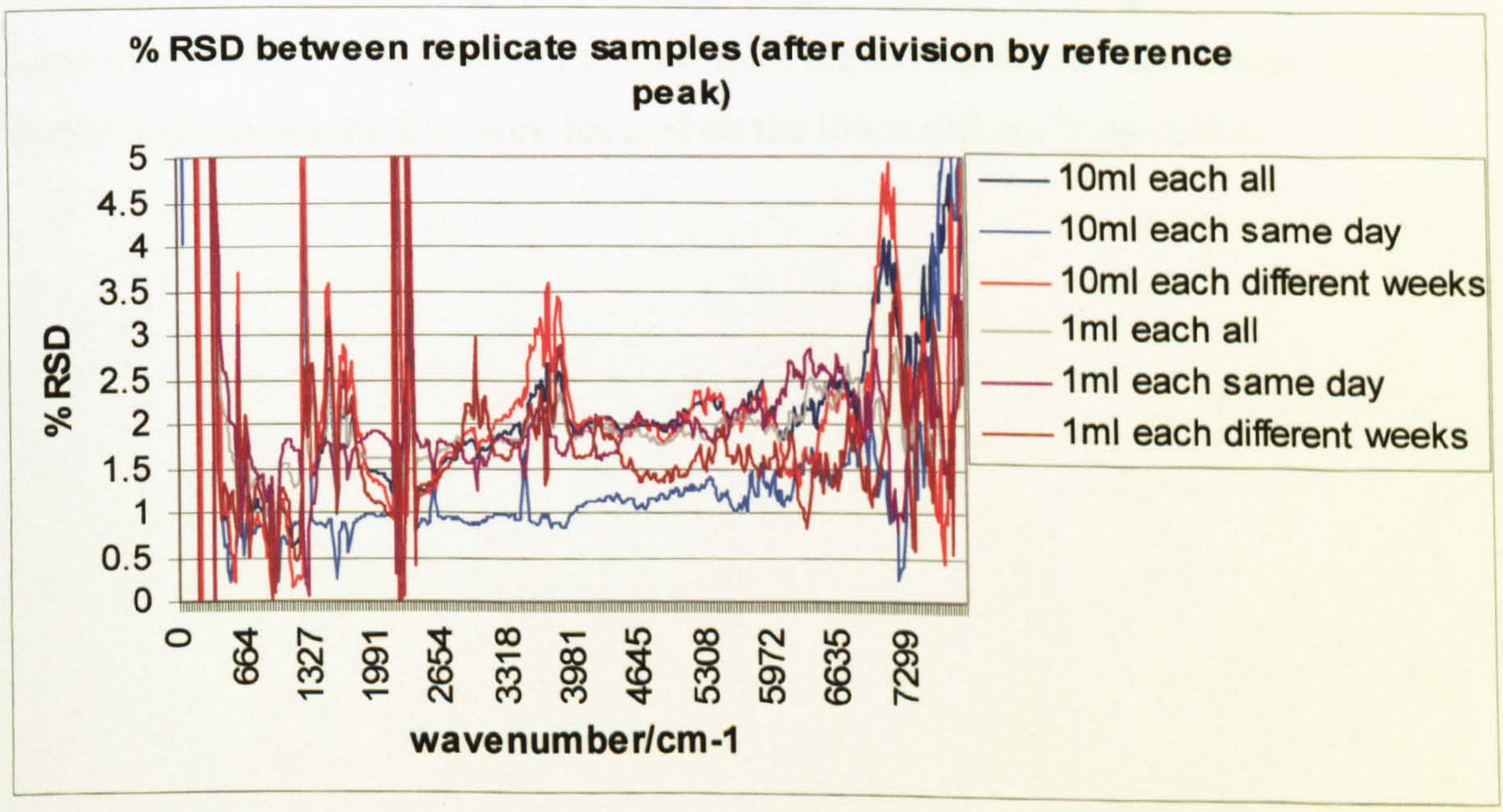


Figure 3.1.24 % RSD between replicate samples, using spectra divided by reference peak

This demonstrates that baseline variation accounts for a high percentage of the overall variation between spectra compared to deviations arising from sample preparation (injection) errors. The RSD that can be attributed to variations in analyte concentration can be assumed to be less than 3 %, based on the assumption that variation in transmission at wavenumbers relating to regions where the gases absorb are directly proportional to concentration.

3.1.2.6 Conclusion

This work proved that dividing the data by a reference response, followed by normalisation, was effective in removing a large amount of the variation caused by changes in the baseline (as illustrated in figure 3.1.18). This was also demonstrated by comparing the % RSD between replicate samples collected over different weeks (figure 3.1.24).

These results lend weight to the idea of employing an internal standard as a means of compensating for background variation, therefore this method was continued.

Also investigated was the effect of resolution, in particular 16 cm^{-1} and 32 cm^{-1}. Ideally high resolution of 1 cm^{-1} or less should be employed for measuring gases, however as mentioned earlier, the resolution of the gas monitoring system being developed in this work is limited by the bandwidth of the QWIP detectors that will be used in the final array. Work by Kumer *et al.* employed lower resolutions of 8 cm^{-1} to monitor stack emissions thus demonstrating the feasibility of successfully utilising lower resolutions in gas monitoring.[32] The results of models built for 16 and 32 cm^{-1} resolution data gave similar SEPs, hence further work focused on the lower (32 cm^{-1}) resolution.

3.2 Multivariate modelling with IR spectra

This section focuses on multivariate models built using the IR spectral data from mixtures of CH_4, CO, CO_2 and N_2O with SF_6 as an internal standard. The aim was to determine sets of wavenumbers, suitable for incorporating into a detector array, that give comparable results to models based on the entire spectra

3.2.1 VS-MLR

The results from section 3.1 suggested reasonable prediction models could be built from 32 cm^{-1} resolution data, hence this resolution was employed for further investigations into the VS-MLR algorithm and for PLS models.

3.2.1.1 Optimisation of the variable selection process

As described in section 3.2, the variable selection process has four different parameters that can be altered; 'maxvars', 'convergence level 1', 'convergence level 2' and the number of iterations. The effect of altering these different parameters was investigated to determine the optimum values, i.e. which values gave selected wavenumber sets with the lowest prediction errors.

Effect of convergence levels on the average SEP value

The convergence levels determine the extent the prediction error has to improve before wavenumbers are retained in the selection model, as a result they have a major impact on the selection process.

The following tables show the effect of changing the convergence levels (cl) on the average SEP obtained for the final chosen set of wavenumbers. The number of final wavenumbers selected for the given convergence levels is also displayed. (N.B. Cl 2 has to be equal or greater than cl 1 otherwise large numbers of wavenumbers are selected since cl 2 ceases to have any effect in reducing the number of wavenumbers, two examples are included in the table to demonstrate this.)

Normalised data

The SEP values and the number of selected variables obtained for different convergence level values are displayed in tables 3.2.1 and 3.2.2.

Convergence level 1 \ Convergence level 2	0.9	1	1.2	1.3	1.4	1.5	1.7	2	3
0.9	0.7572	0.7434	0.6890	0.7103	0.7103	0.7470	0.7095	0.7421	0.7421
1	0.7574	0.7672	2.5414	2.9791	4.5984	3.7673	4.5749	4.5902	4.5888
1.1	3.3303	3.0421	3.3606	4.0067	4.3517	4.3717	4.4208	4.2884	4.5810
1.2	0.7294		0.6468	0.6158	0.6098	0.5984	0.6293	0.7330	0.6500
1.3				0.7852	0.6909	0.6702	0.7426	0.7533	0.6315
1.4					0.5568	0.7057	0.6257	0.6422	0.9345
1.5	0.8594					0.6829	0.6933	0.6016	0.9759
1.7							0.5686	0.7722	0.7228
2								0.8150	0.9434

Table 3.2.1 Effect of convergence levels on the average SEP value (for Normalised data)

Convergence level 1 \ Convergence level 2	0.9	1	1.2	1.3	1.4	1.5	1.7	2	3
0.9	14	12	6	6	6	6	5	5	5
1	12	11	9	7	1	4	1	1	1
1.1	15	11	10	7	2	3	2	2	1
1.2	243		17	16	17	16	16	16	6
1.3				15	14	12	11	5	8
1.4					15	15	12	9	6
1.5						12	6	9	9
1.7							10	9	9
2								9	10

Table 3.2.2 Number of wavenumbers selected for given convergence levels

The cl values entered were between 0.9 and 2 for cl 1 and 0.9 and 3 for cl 2. Cl 1 values of less than 0.9 resulted in no wavenumbers selected, whereas cl 1 values greater than 1.5 significantly increased the length of time the selection process took without any significant improvement on results. Cl 1 values of 1 and 1.1 resulted in average SEP values significantly higher than runs using different values, as well as only selecting a few wavenumbers (this is likely side effect of the normalisation treatment but this was never fully investigated). The minimum SEP value was 0.56 (with cl values 1.4 and 1.4), however this required 15 wavenumbers and ideally a group of only 5 or 6 was preferred, this generally occurred when cl 1 was 0.9.

Non-normalised data

Results for non-normalised data are shown in tables 3.2.3 and 3.2.4.

Convergence level 1	Convergence level 2								
	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
0.9	*	*	*	*	*	*	*	*	*
0.95	0.6560	0.6560	0.5527	0.5825	0.5474	0.5305	0.5968	0.5974	0.6692
1	0.6999	0.7054	0.6100	0.7254	0.7394	0.7041	0.6248	0.8207	0.7968
1.1		0.6354	0.6437	0.7950	0.8696	0.6711	0.9606	0.5399	0.9606
1.2			0.7501	0.6242	0.7876	0.8018	0.7946	0.8259	0.8267
1.3				0.6780	0.8458	0.8421	0.7989	0.8632	0.8347
1.4					0.6953	0.8284	0.7478	0.9409	1.0842
1.5						0.7085	0.7650	0.9330	0.7532
1.6							0.8120	0.8110	0.7989
1.7								0.8371	0.7989

** No variables selected*

Table 3.2.3 Effect of convergence levels on the average SEP value (for non-normalised data

Convergence level 1	Convergence level 2								
	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
0.9	*	*	*	*	*	*	*	*	*
0.95	13	13	7	6	7	6	7	7	7
1	17	17	13	16	17	9	6	11	14
1.1		17	13	11	17	12	14	8	7
1.2			13	18	15	11	9	7	8
1.3				16	15	10	8	15	7
1.4					12	16	7	10	13
1.5						10	12	15	13
1.6							16	12	11
1.7								10	7

Table 3.2.4 Number of wavenumbers selected at given convergence levels (for non-normalised data)

The average SEP value ranged between 0.53 and 1.08, with the lowest SEP obtained when cl values of 0.95 and 1.5 were set. No wavenumbers were selected when cl 1 was less than 0.95 and unlike the normalised data there was no significant increase in prediction error when using a value of 1.1 for cl 1.

Effect of maxvars

'Maxvars' is the maximum number of variables employed in the selection process, i.e. by setting 'maxvar' to 200 the process stops adding any more variables after a total of 200 variables have been used. This has the advantage of lowering the amount of time taken for the selection process to be performed, particularly when large data sets are analysed; however, if this value is set too low, important variables can be missed out of the process.

With the IR data analysed (consisting of 512 wavenumber measurements) the lowest SEP values were obtained when the maximum number of wavenumber measurements used in the variable selection process was greater than 400. Below this, the errors in prediction increased exponentially with decreasing numbers of wavenumbers (figure 3.2.1). For most of the variable selection runs, a minimum of 480 wavenumbers were utilised

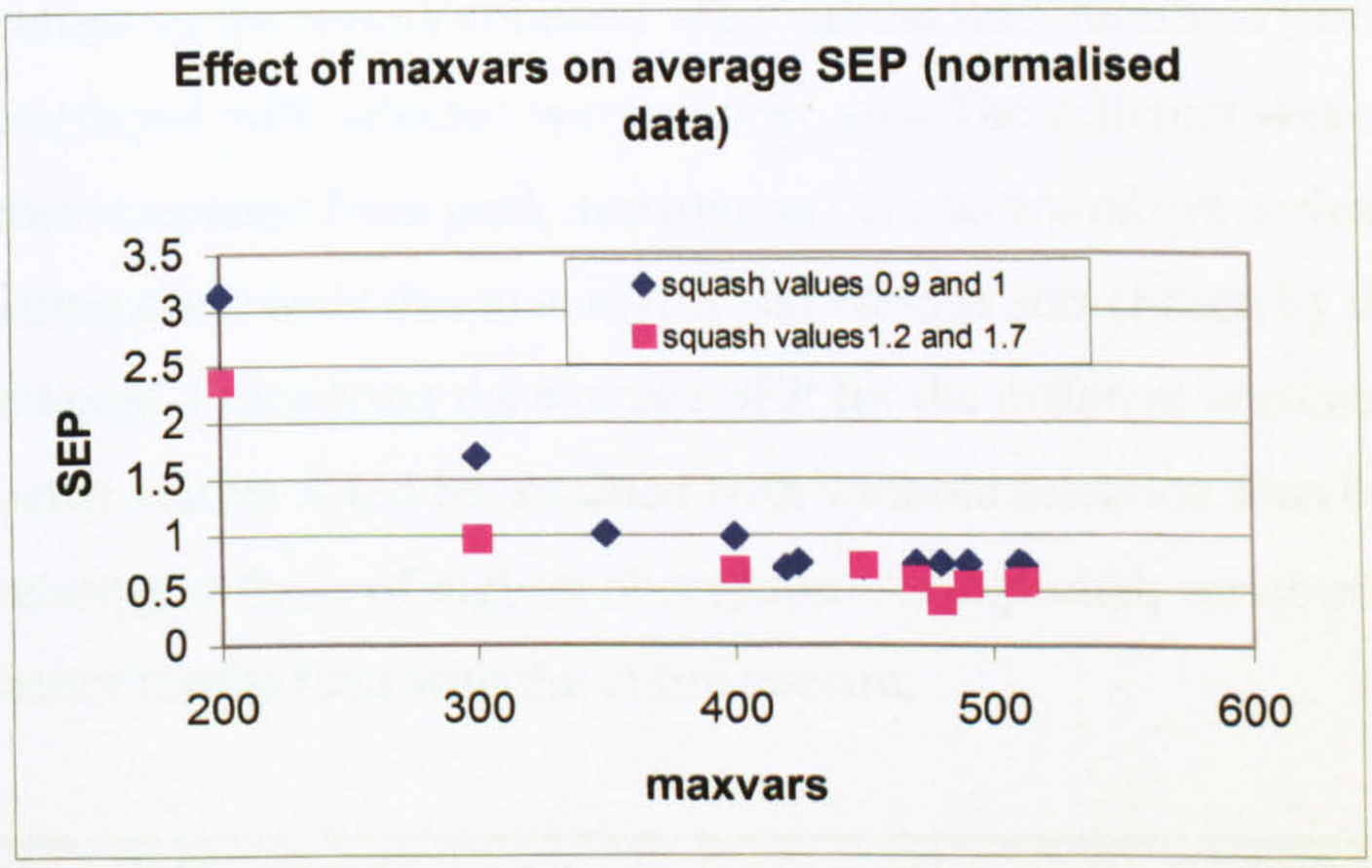


Figure 3.2.1 Effect of the maximum number of wavenumbers used in the variable selection process

Effect of number of iterations

The results showed that the average SEP varied with each run irrespective of the number of iterations per run (figure 3.2.2) and was more dependent on the actual wavenumbers selected. There was no significant change when the number of iterations was increased from 10 to 100, subsequently only 10 iterations were employed as this was faster and allowed more time to be spent investigating the different convergence levels.

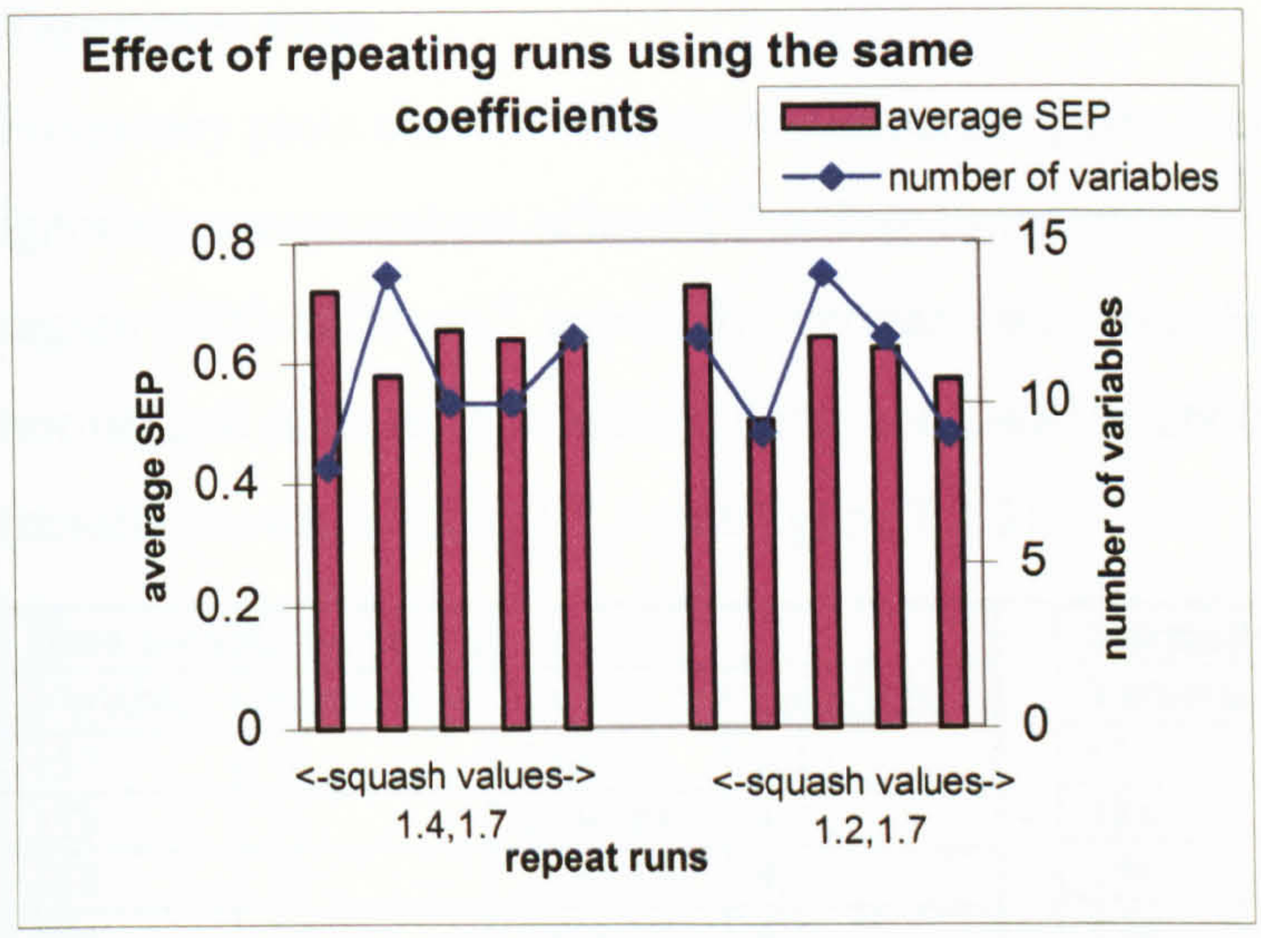


Figure 3.2.2 Effect of number of iterations on the final SEP

3.2.1.2 Errors in predictions for different wavenumber sets

Table 3.2.5 lists the SEP values for some of the different wavenumber sets. This compares the results obtained when all the wavenumbers (the entire spectra) were employed with selected wavenumber sets. The different wavenumber sets included; measurements from peak maximums i.e. wavenumbers corresponding to main absorption bands due to analytes and various sets chosen by the variable selection process. Comparing the average SEP for the different wavenumber sets indicated that better results could be obtained with variable selection than by using wavenumbers relating to those of highest absorption. As expected, variable selection results gave better results than with the entire spectra.

Wavenumbers used in model	**CH_4**	**CO**	**CO_2**	**N_2O**	**Average**
All wavenumbers (normalised)	2.09	1.11	0.45	1.08	1.18
All wavenumbers (no normalisation)	1.97	0.69	0.63	1.33	1.15
Peak maximums at [3009,2577,2469,2345,2222 and 2114cm^{-1}]	1.34	0.71	4.27	0.42	1.69
Peak maximums at [3009,2577,2469,2345,2222,2114,3487,3626 and 3734cm^{-1}]	0.83	1.20	0.45	0.46	0.73
From VS-MLR[2037,2886,3009,2391,2639,2099,3487 and 3657cm^{-1}]	0.43	0.92	0.80	0.80	0.74
From VS-MLR[1404,2114,6110,3704,401,3040,2561cm^{-1}]	0.78	0.79	0.50	0.56	0.66
From VS-MLR[7407,602,3734,4274,3719,3487,2037,1790cm^{-1}]	0.61	0.71	0.40	0.69	0.60
From VS-MLR[2808,1204,2716,2886,3657,2037cm^{-1}]	0.38	0.57	0.46	0.52	0.48

Table 3.2.5 Comparison of SEP values obtained from models built with different wavenumber sets

Frequency Plots

Frequency plots show which wavenumbers/regions were most commonly selected, ignoring wavenumbers selected 2 or less times most of the selections occurred in the region 2000-4600 cm^{-1}, especially between wavenumbers 2000-2200 cm^{-1} (for both normalised and non-normalised data) and 2400 to 2600 cm^{-1} (non-normalised data) (results shown in table 3.2.6 and figure 3.2.3).

Data without normalisation			
Variable	Frequency	cm^{-1}	µm (2d.p.)
157	16	2407.16	4.15
133	15	2036.83	4.91
158	12	2422.59	4.13
159	10	2438.02	4.1
282	10	4335.97	2.31
161	9	2468.88	4.05
170	9	2607.75	3.83
223	8	3425.57	2.92
231	8	3549.01	2.82
190	7	2916.36	3.43
229	7	3518.15	2.84
279	7	4289.68	2.33
49	6	740.66	13.5
61	6	925.83	10.8
62	6	941.26	10.62
287	6	4413.12	2.27
366	6	5632.13	1.78
10	5	138.87	72.01
134	5	2052.26	4.87
189	5	2900.93	3.45
233	5	3579.87	2.79

Normalised data			
Variable	Frequency	cm^{-1}	µm (2d.p.)
133	18	2036.83	4.91
188	10	2885.5	3.47
196	10	3008.95	3.32
156	7	2391.73	4.18
172	7	2638.61	3.79
82	6	1249.87	8
137	6	2098.55	4.77
227	6	3487.29	2.87
278	6	4274.25	2.34
132	5	2021.39	4.95
138	5	2113.98	4.73
224	5	3440.1	2.91
238	5	3657.03	2.73
39	4	586.36	17.05
79	4	1203.58	8.31
136	4	2083.12	4.8
168	4	2576.89	3.88
171	4	2623.18	3.81
199	4	3055.24	3.27
232	4	3564.44	2.81
242	4	3718.75	2.69
244	4	3749.61	2.67
267	4	4104.51	2.44
425	4	6542.53	1.53

Table 3.2.6 Frequency tables of the top selected variables (plus the wavenumber and wavelength they correspond to) for variable selection runs (with SEP <0.7)

Wavenumbers 2000-2200 cm^{-1} correspond to CO absorption peaks so prediction models would require at least one wavenumber selected from this region in order to predict CO concentrations. Comparing the other frequently selected wavenumbers to reference spectra, it is clear that most of these relate to absorption regions (N.B. Reference spectra can be found in the appendix (section 6).

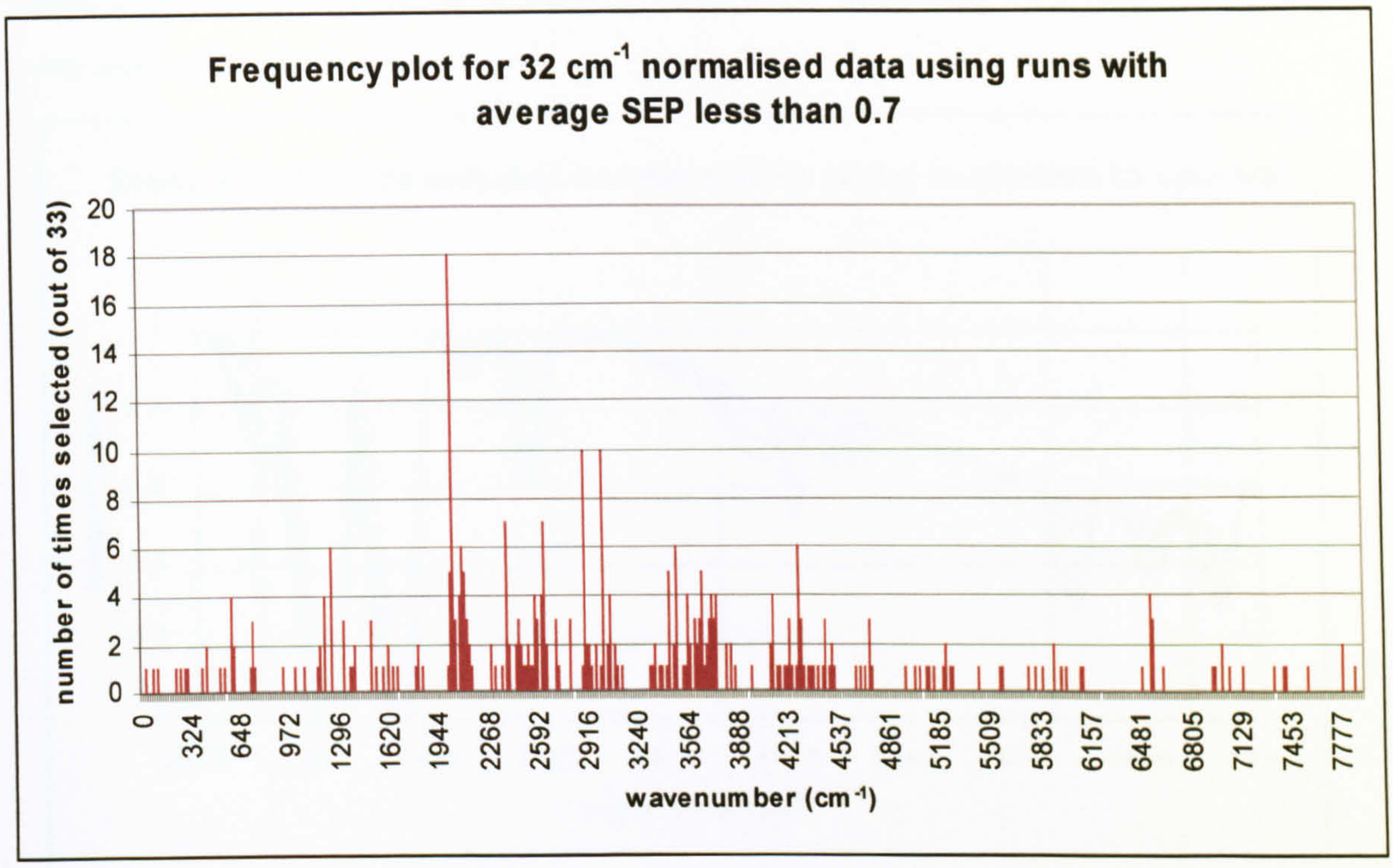


Figure 3.2.3 Frequency plot of wavenumbers selected from variable selection runs performed on 32 cm^{-1} resolution data (with data divided by reference and normalised treatment)

For wavenumbers selected with normalised data, models were built using the top 13 most selected wavenumbers (i.e. those selected 5 or more times), the top 8 and the top 4. Variable selection was performed on the top 13 selected wavenumbers and this resulted in the 2nd, 7th, 11th, 12th and 13th wavenumbers being chosen i.e. wavenumbers 2886, 2099, 2114, 3441 and 3657 cm^{-1}.

The SEP values obtained are shown in table 3.2.7.

Model	CH_4	CO	CO_2	N_2O	Average
Top 13	0.54	0.85	0.92	0.61	0.73
Top 8	0.49	0.80	1.52	0.53	0.84
Top 4	1.10	1.11	1.25	5.54	2.25
Selected set	0.54	0.76	0.65	0.30	0.56

Table 3.2.7 SEPs obtained from models built from top selected wavenumbers

The best results were obtained using the selected set. Figure 3.2.4 relates the selected wavenumbers to the spectra, wavenumbers 2099 and 2114 cm^{-1} mainly relate to CO absorption, 2886 cm^{-1} to CH_4, and 3441 cm^{-1} to N_2O. None of these wavenumbers seem to relate specifically to CO_2 absorption. This was most likely a result of over fitting the

data, hence the effectiveness of different models for predicting new validation spectra was assessed.

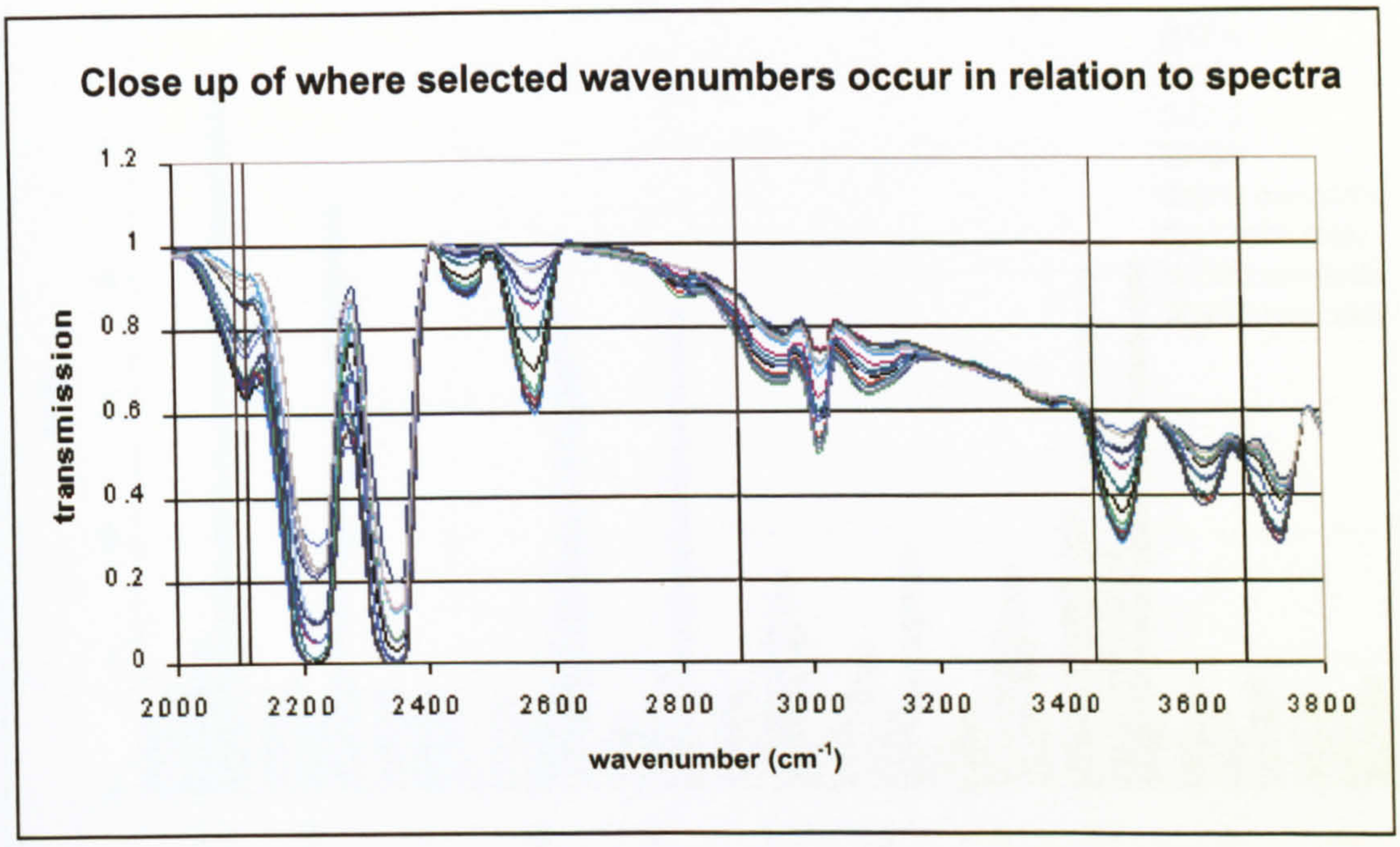
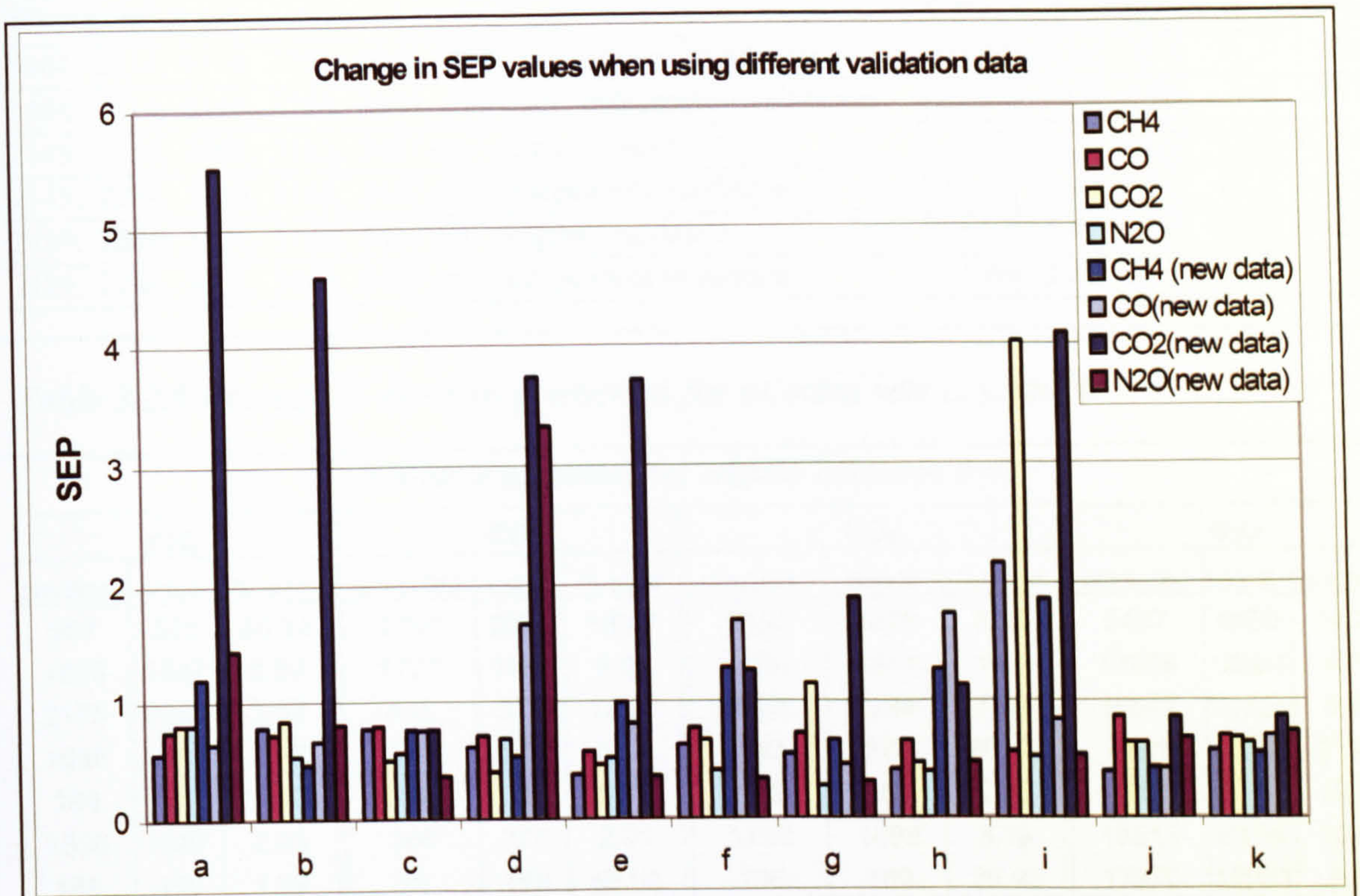


Figure 3.2.4 Plot showing where the selected wavenumbers 2886, 2099, 2114, 3441 and 3657 cm^{-1} occur in relation to spectra

3.2.1.3 Prediction of new validation data

Models built from selection sets that gave low prediction errors were employed to predict concentrations for new spectra (using the same selected wavenumbers). Figure 3.2.5 displays the SEPs obtained for each gas for both the original validation data (used in the selection process) and for the new validation data. The results do indicate over fitting in some of the models, for example in the previously mentioned set consisting of wavenumbers 2886, 2099, 2114, 3441 and 3657 cm^{-1} (set a in figure 3.2.5), none of the wavenumbers specifically relate to CO_2 absorption peaks and this clearly caused an inability to predict CO_2 concentrations for new validation data.

Out of the wavenumber sets tested, only sets c (consisting of wavenumbers 1404, 2114, 6110, 3703, 401, 3040 and 2561 cm^{-1}), j (wavenumbers 2129, 2284, 3071, 3148 and 3487 cm^{-1}) and k (2145, 2284, 3086, 3163 and 3487 cm^{-1}) gave similar SEPs for the two validation sets, thus indicating these models were more robust, and less sensitive to background variations.



set	selected wavenumbers(cm^{-1}) contained within set								
a	2099	2114	3441	2886	3657				
b	2037	3487	3441	2114	2886	3657			
c	1404	2114	6110	3703	401	3040	2561		
d	7407	602	3734	4274	3719	3487	2037	1790	
e	2808	1204	2716	2886	3657	2037			
f	2037	2407	2608	2423	2438	2469	3426	4290	
g	2037	2407	2469	3009	4336	2438	2886	3487	4290
h	2423	2037	3426	2407	2438	2608	4413		
I	1003	926	4413	2639	3564	4290	2423		
j	2129	2284	3071	3148	3487				
k	2145	2284	3086	3163	3487				

Figure 3.2.5 SEP obtained for original and independent validation data for different wavenumber sets

For easy comparisons the average SEP was calculated to compare the selection results. In this case the concentration values were related to the amounts of each gas injected into the gas cell, more commonly percentage error is used to evaluate prediction ability so a closer inspection of the percentage error performed.

Table 3.2.8 gives the average % prediction error obtained with wavenumber sets c, j and k for the original and the new validation data. Wavenumber set k give the most similar % error between the original and new validation data. Table 3.2.9 shows the predicted and actual concentrations, together with the % error of each sample for the model built from set c.

Wavenumber set	CH_4	CO	CO_2	N_2O	Average
1404, 2114, 6110, 3703, 3040, 2561 cm^{-1} original validation	10.98	15.11	10.7	11.1	11.97%
1404, 2114, 6110, 3703, 3040, 2561 cm^{-1} independent validation	23.87	10.9	20	9.5	16.07%
2145, 2284, 3086, 3163, 3487 cm^{-1} original validation	10.2	14.01	12.5	12.5	12.30%
2145, 2284, 3086, 3163, 3487 cm^{-1} independent validation	18.51	19.94	16.1	12.3	16.71%
2129, 2284, 3071, 3148, 3487 cm^{-1} original validation	5.77	25.71	12.6	12.4	14.13%
2129, 2284, 3071, 3148, 3487 cm^{-1} independent validation	14.96	12.28	17.5	9.85	13.64%

Table 3.2.8 Average % error in prediction for selected sets c, j, and k

% error in prediction for original validation data

CH_4			CO			CO_2			N_2O		
predicted	actual	% error	predicted	actual	% error	predicted	actual	% error	predicted	actual	% error
907	566	60.24	1796	2075	13.47	2056	2075	0.94	6467	5660	14.25
1898	1887	0.59	1701	1887	9.83	1924	1887	1.99	17966	18868	4.78
2175	2264	3.92	426	377	12.98	2085	2264	7.93	19542	20755	5.84
1656	1698	2.48	390	377	3.41	492	377	30.38	4591	3774	21.67
198	189	5.10	1786	1698	5.17	195	189	3.28	4165	3774	10.37
1836	1887	2.68	366	377	2.91	1752	1698	3.19	10913	9434	15.68
185	189	1.89	79	189	58.03	138	189	26.93	17870	16981	5.24
average error		10.99	average error		15.11	average error		10.66	average error		11.12

% error in prediction for independent validation data

CH_4			CO			CO_2			N_2O		
predicted	actual	% error	predicted	actual	% error	predicted	actual	% error	predicted	actual	% error
1333	1321	0.96	981	943	3.95	584	377	54.84	11959	11321	5.64
1056	943	11.94	427	377	13.14	566	377	49.97	9959	9434	5.56
1955	1887	3.61	1639	1887	13.15	1852	1887	1.83	17962	18868	4.80
-10	189	105.21	213	189	12.83	197	189	4.23	1267	1887	32.87
1686	1887	10.64	1150	1321	12.96	1926	2075	7.22	12033	11321	6.29
836	755	10.83	855	943	9.36	1854	1887	1.72	5554	5660	1.87
average error		23.87	average error		10.90	average error		19.97	average error		9.51

Table 3.2.9 Percentage error in prediction for validation samples using wavenumber set c (1404, 2114, 6110, 3707, 401, 3040, and 2561cm^{-1}) (with predicted and actual concentrations (ppm))

This shows that the percentage error is generally less than 10 % but is greatly increased at the lower concentrations. It may be that the calibration range was too wide and that the lower concentrations were having a leverage type effect on the data. In order to see if this was true, samples with 189 ppm concentrations were removed from the data set and the percentage errors recalculated; however, this was found to give no significant improvement in prediction errors.

Focusing on wavenumber set k, which gave similar average prediction error for both the original and independent validation data (a breakdown of the % error for each analyte in each sample is given in table 3.2.10), as well as consisting of only five wavenumbers, it

can be seen (figure 3.2.6) that each selected wavenumber relates to the absorption of a specific gas i.e. 2129 cm^{-1} corresponds to CO, 2284 cm^{-1} to CO_2, 3071 and 3148 cm^{-1} both relate to CH_4, and 3487 cm^{-1} relates to N_2O.

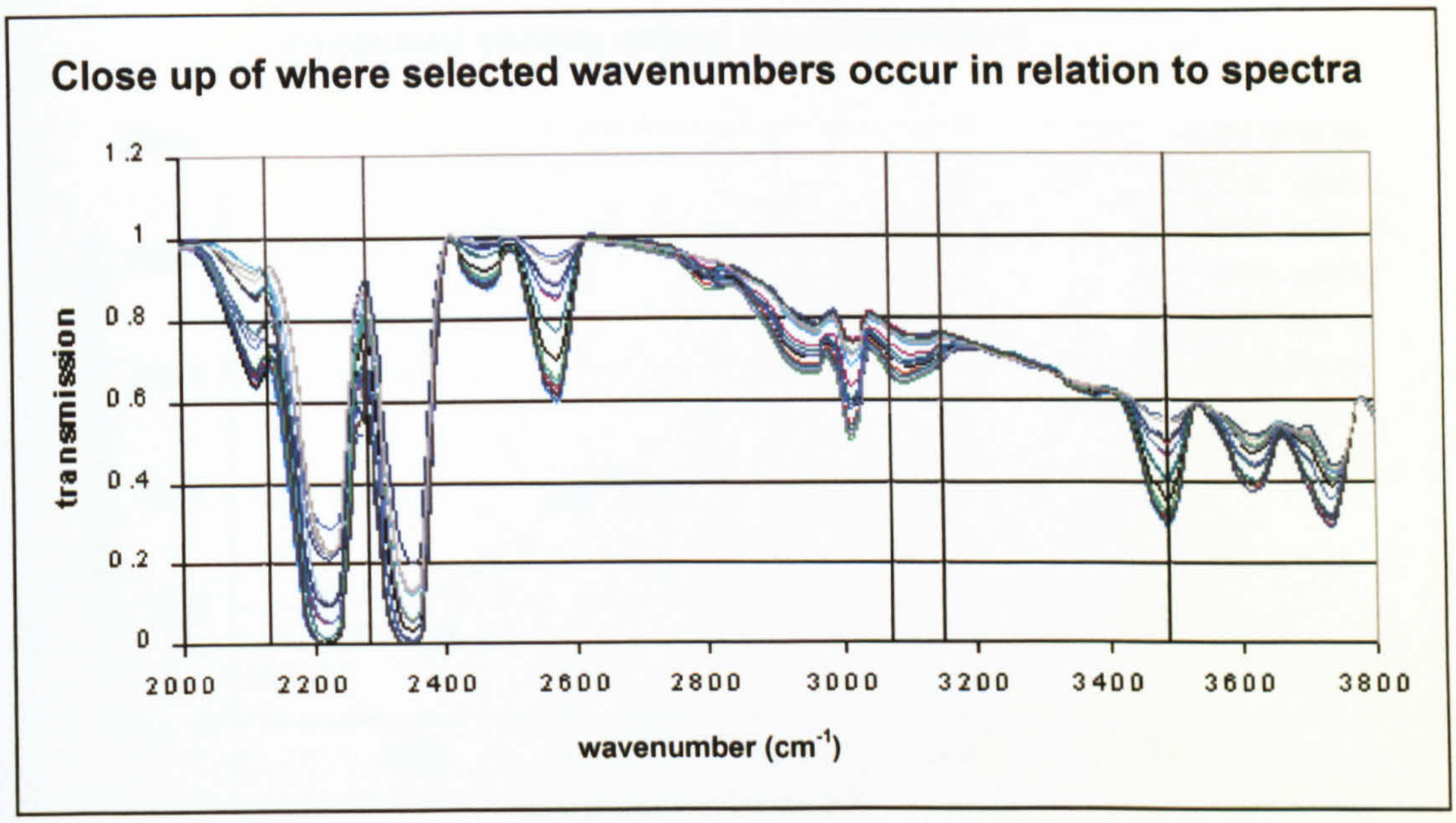


Figure 3.2.6 Plot showing where the selected wavenumbers 2129, 2284, 3071, 3148 and 3487 cm^{-1} occur in relation to spectra

% error in prediction for original validation data

CH_4		CO		CO_2		N_2O	
sample	% error	sample	% error	sample	% error	sample	% error
1	26.80	1	13.26	1	4.64	1	9.11
2	4.81	2	4.79	2	4.33	2	4.70
3	1.76	3	53.52	3	10.42	3	6.45
4	1.31	4	24.65	4	26.62	4	17.00
5	3.09	5	2.97	5	8.05	5	23.88
6	0.91	6	32.09	6	1.96	6	19.40
7	1.71	7	48.72	7	32.21	7	6.55
average	5.77	average	25.71	average	12.60	average	12.44

% error in prediction for newl validation data

CH_4		CO		CO_2		N_2O	
sample	% error	sample	% error	sample	% error	sample	% error
1	7.97	1	1.46	1	9.02	1	13.33
2	3.79	2	12.27	2	7.24	2	14.11
3	1.65	3	7.79	3	11.06	3	3.73
4	67.13	4	50.58	4	58.99	4	0.95
5	0.41	5	0.05	5	11.31	5	16.39
6	8.82	6	1.55	6	7.29	6	10.61
average	14.96	average	12.24	average	17.48	average	9.85

Table 3.2.10 Percentage error in prediction for validation samples using wavenumber set k

Figure 3.2.7 displays the plot of predicted versus actual concentrations and figure 3.2.8 is a plot of the residuals.

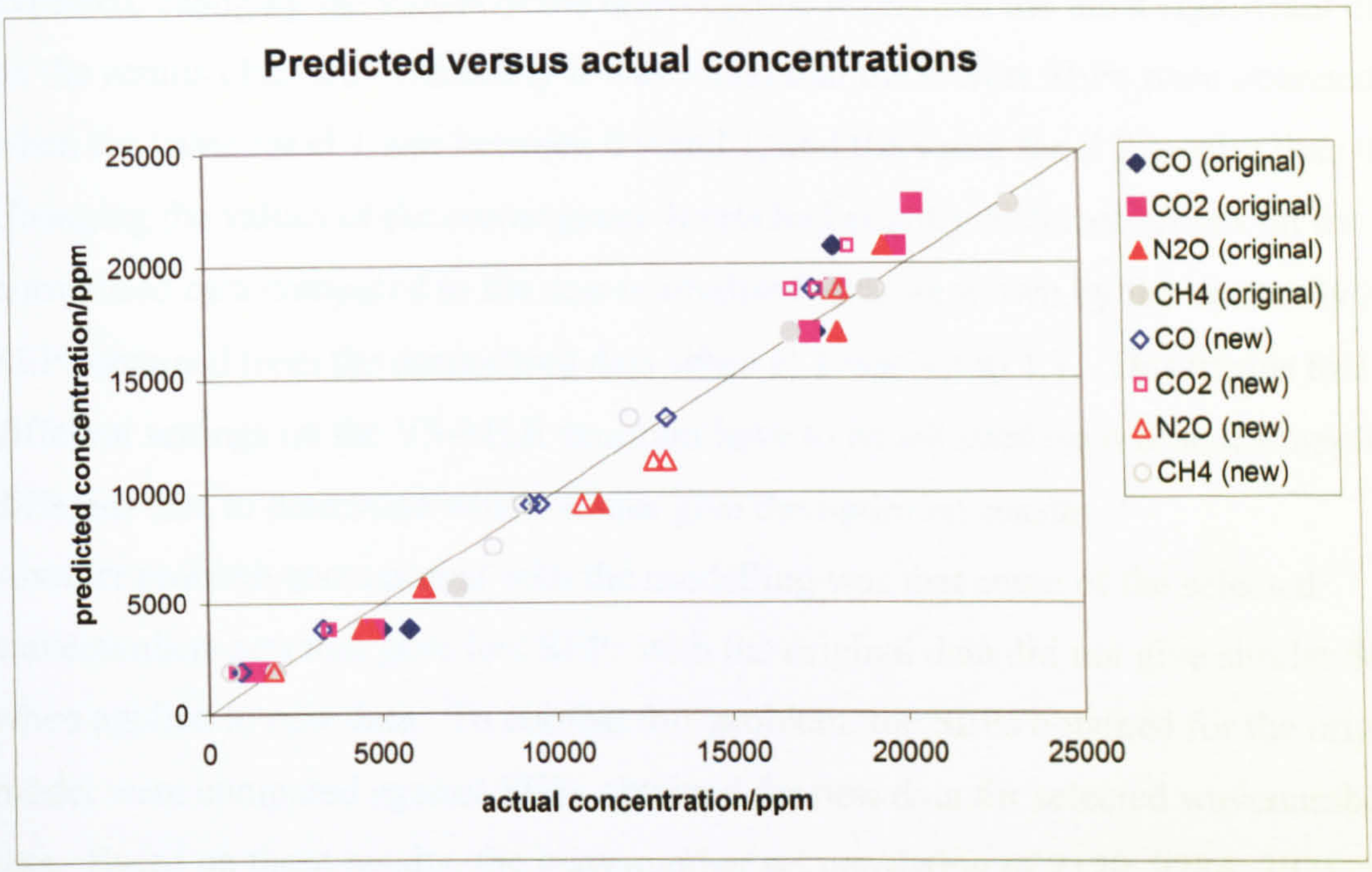


Figure 3.2.7 Plot of predicted versus actual concentrations for model built from wavenumbers 2129, 2284, 3071, 3148 and 3488 cm^{-1}

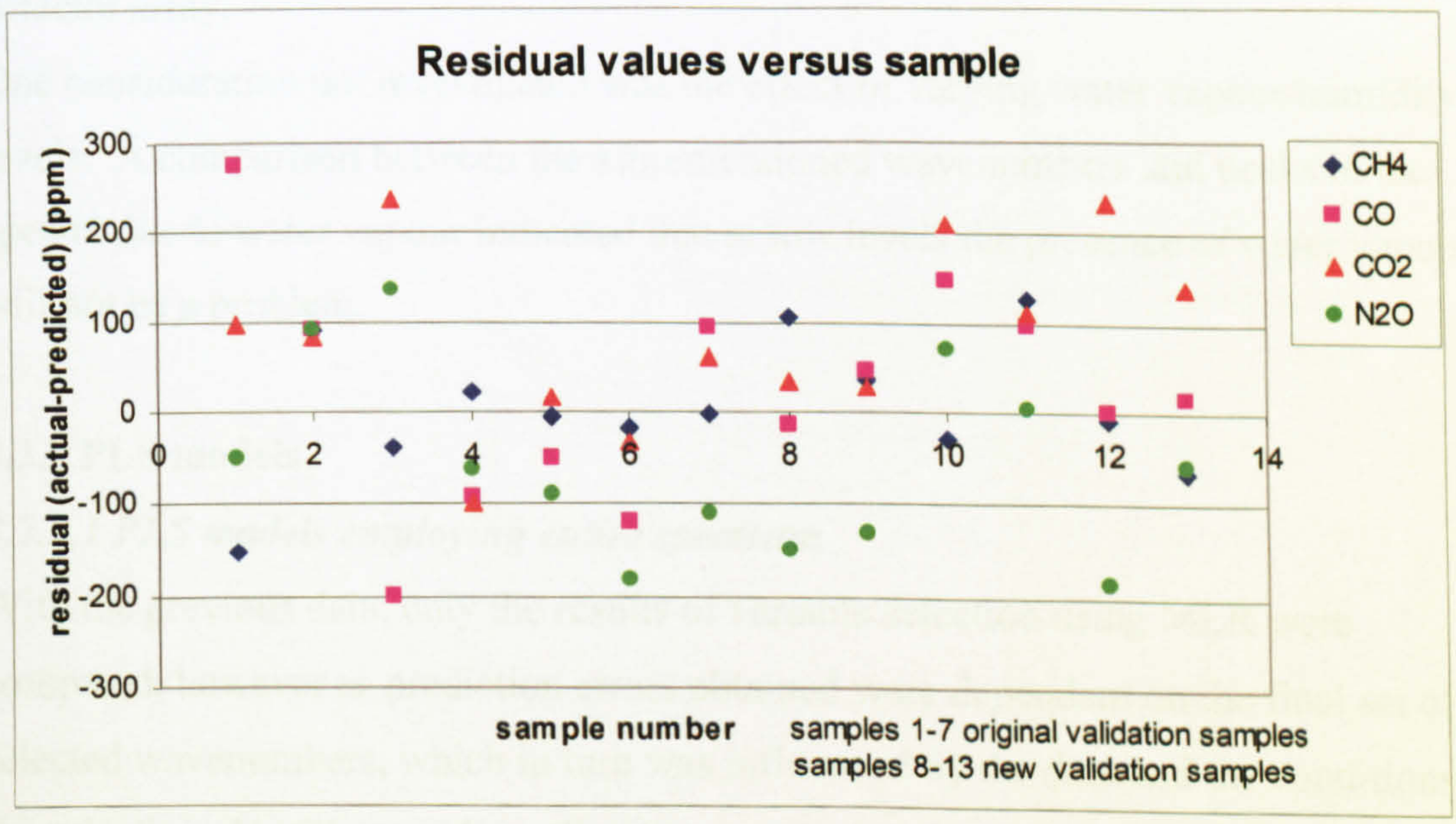


Figure 3.2.8 Plot of residuals for model built from wavenumbers 2129, 2284, 3071, 3148 and 3488 cm^{-1}

3.2.1.4 Conclusion

Part of this work focused on the different settings on the VS-MLR algorithm. As expected, changing the values of the convergence levels had the most significant effect on the results obtained. Generally it was found that the lowest SEPs were obtained when the value for cl 1 was between 0.9 and 1, and the value for cl 2 greater than 0.9. Changing the values of the convergence levels had slightly different effects on the normalised data compared to the non-normalised data, as shown by the unusually high SEPs obtained from the normalised data when cl 1 was set to 1.1. This means that the different settings on the VS-MLR program have to be adjusted each time it is applied to different data to determine which values give the optimum results.

Another problem encountered with the modelling was that some of the selected wavenumbers sets that gave low SEPs with the original data did not give similar SEPs when applied to new data. To combat this problem, the SEPs obtained for the original model were compared against SEPs obtained for new data for selected wavenumber sets. Based on these results, the wavenumber set consisting of 2129, 2284, 3071, 3148 and 3488 cm^{-1} was deemed to be a suitable selection to measure in the planned infrared detector array.

One consideration not investigated was the effect of varying water vapour/humidity levels. A comparison between the aforementioned wavenumbers and peaks in the spectra due to water vapour indicated that at low levels the presence of water vapour will not be a problem.

3.2.2 PLS models

3.2.2.1 PLS models employing entire spectrum

With the previous data, only the results of variable selection using MLR were compared, however as prediction errors obtained were dependant on the final set of selected wavenumbers, which in turn was influenced by the data and the conditions of the selection algorithm used, no direct comparisons were possible. Applying MLR to the entire spectrum results in poor predictions due to the effect of co-linearity within the data, therefore PLS models were built instead.

Unlike MLR, PLS is more equipped to cope with co-linearity within data. In PLS, the data is broken down into scores and loadings, in effect co-linearity is accounted for by the loadings and can actually aid in a model's prediction ability. As a result, PLS

performs better with more wavenumbers and is the preferred technique for building multi-variate calibration models from spectra. In this work, PLS was applied to the different pre-treated spectra to determine how much effect the different data treatments had on the prediction ability.

Figure 3.2.9 contains results obtained for different PLS models based on the different data treatments, these show the average SEP obtained for the four gases from models using different numbers of latent variables.

The data divided by 1975 cm^{-1} followed by normalisation gave the same results as normalised data that had been divided by the SF_6 reference peak, so these results are not shown. Generally, the SEP dropped dramatically for the first five latent variables reaching a minimum around seven latent variables. A minimum of four latent variables would be expected to account for the four gases, thus allowing for variations in water vapour levels and background variations, five latent variables is a reasonable number to employ.

The absorption data initially gave poor results due to the wide variation of responses around the cut off points (see figure 3.1.17); however, the removal of variables relating to measurements made at wavenumbers less than 500 cm^{-1}, i.e. the first 20 measurements, improved the model. Comparing the SEP for five latent variable models built for different data treatments (figure 3.2.10), it can be seen that applying some kind of data treatment improved the prediction error. The best results were obtained for the absorption spectra after the first 20 measurements were removed. Increasing the number of latent variables to seven (figure 3.2.11) almost halved the SEP for data divided by a reference wavenumber and there was no longer any noticeable difference in prediction error between this treatment and absorption spectra.

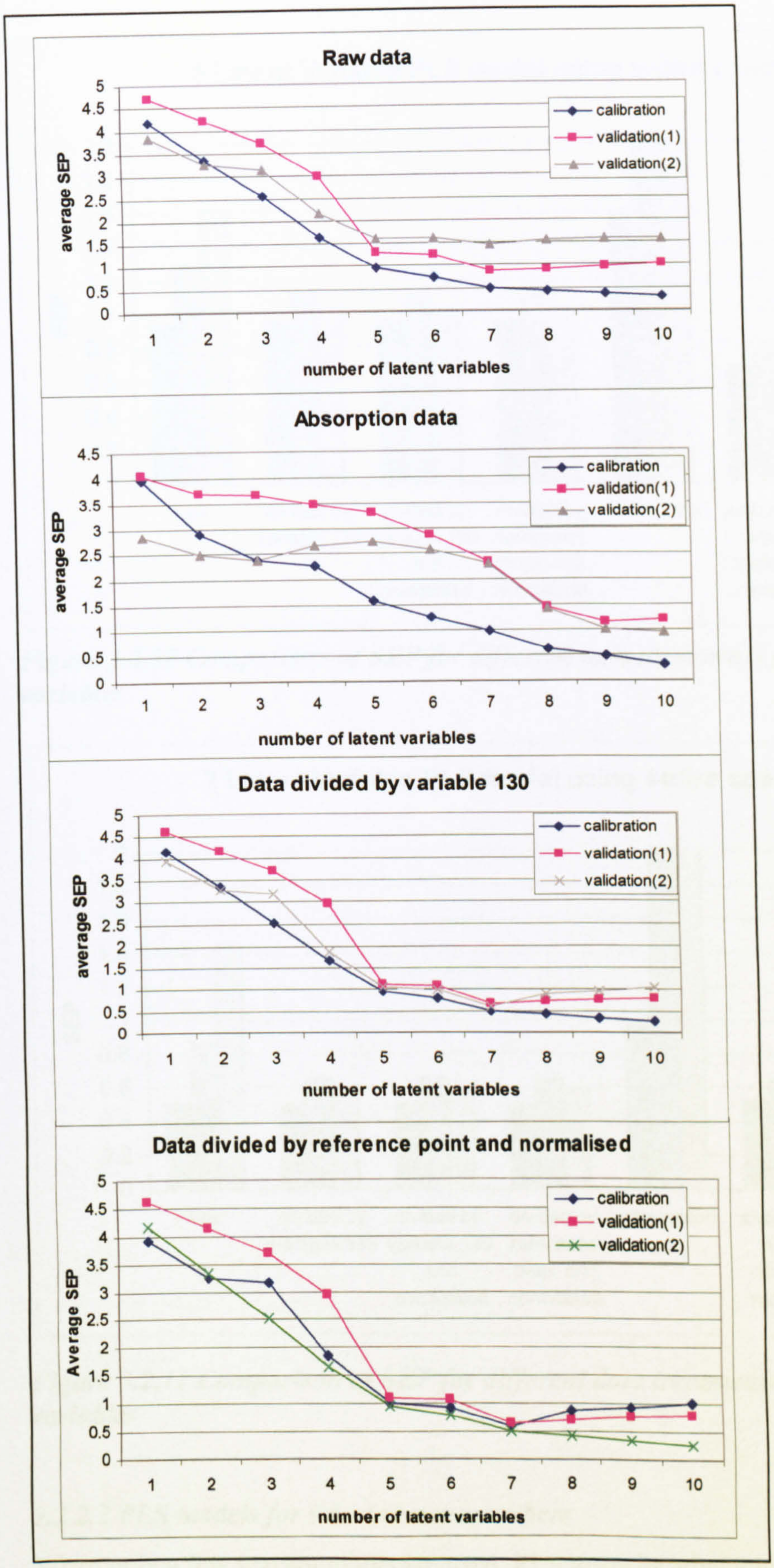


Figure 3.2.9 SEP obtained for PLS models with different data treatments, (NB. Absorption data is spectra minus wavenumbers less than 500 cm^{-1})

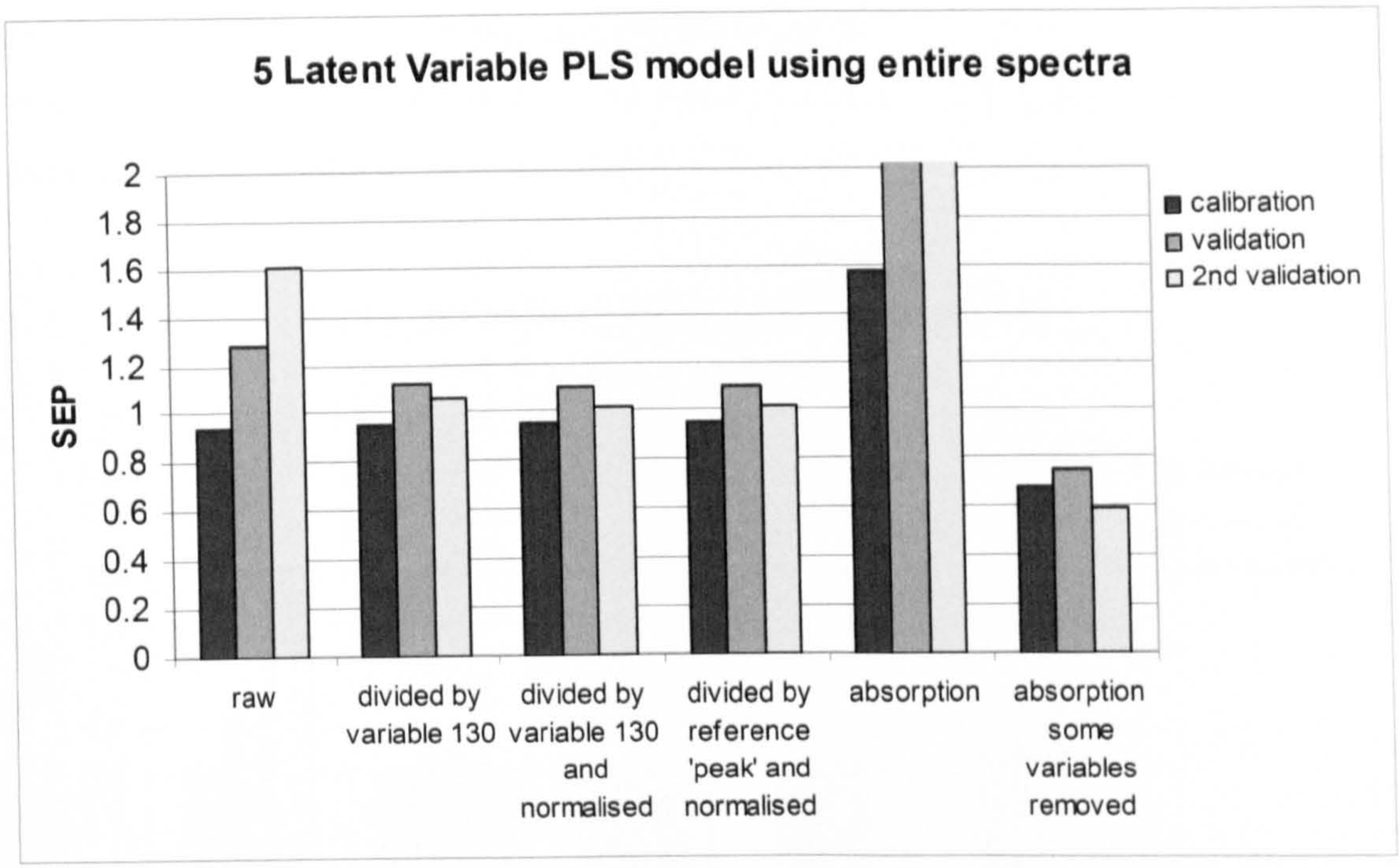


Figure 3.2.10 Comparison of SEP for different data treatments for five latent variables

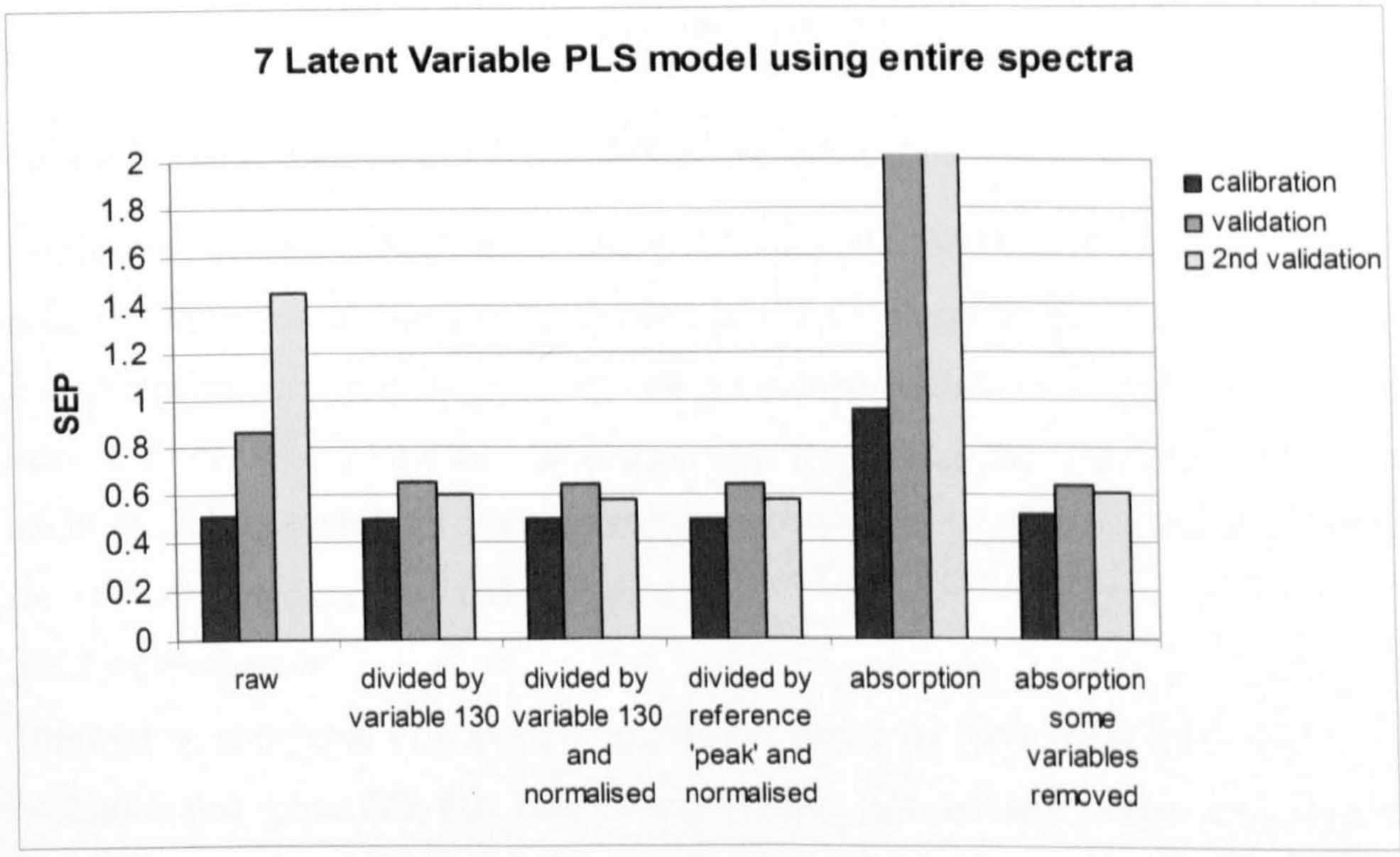


Figure 3.2.11 Comparison of SEP for different data treatments for seven latent variables

3.2.2.2 PLS models for selected wavenumbers

When only a few wavenumbers are used, PLS loses its advantage over MLR models. PLS models were built using the five selected wavenumbers; 2129, 2284, 3071, 3148 and 3488 cm^{-1}. The best prediction results were obtained when five latent variables were

incorporated, the results of which are shown in figure 3.2.12 and as there were only five wavenumbers to start with this gave the same results as MLR models built from the same five wavenumbers.

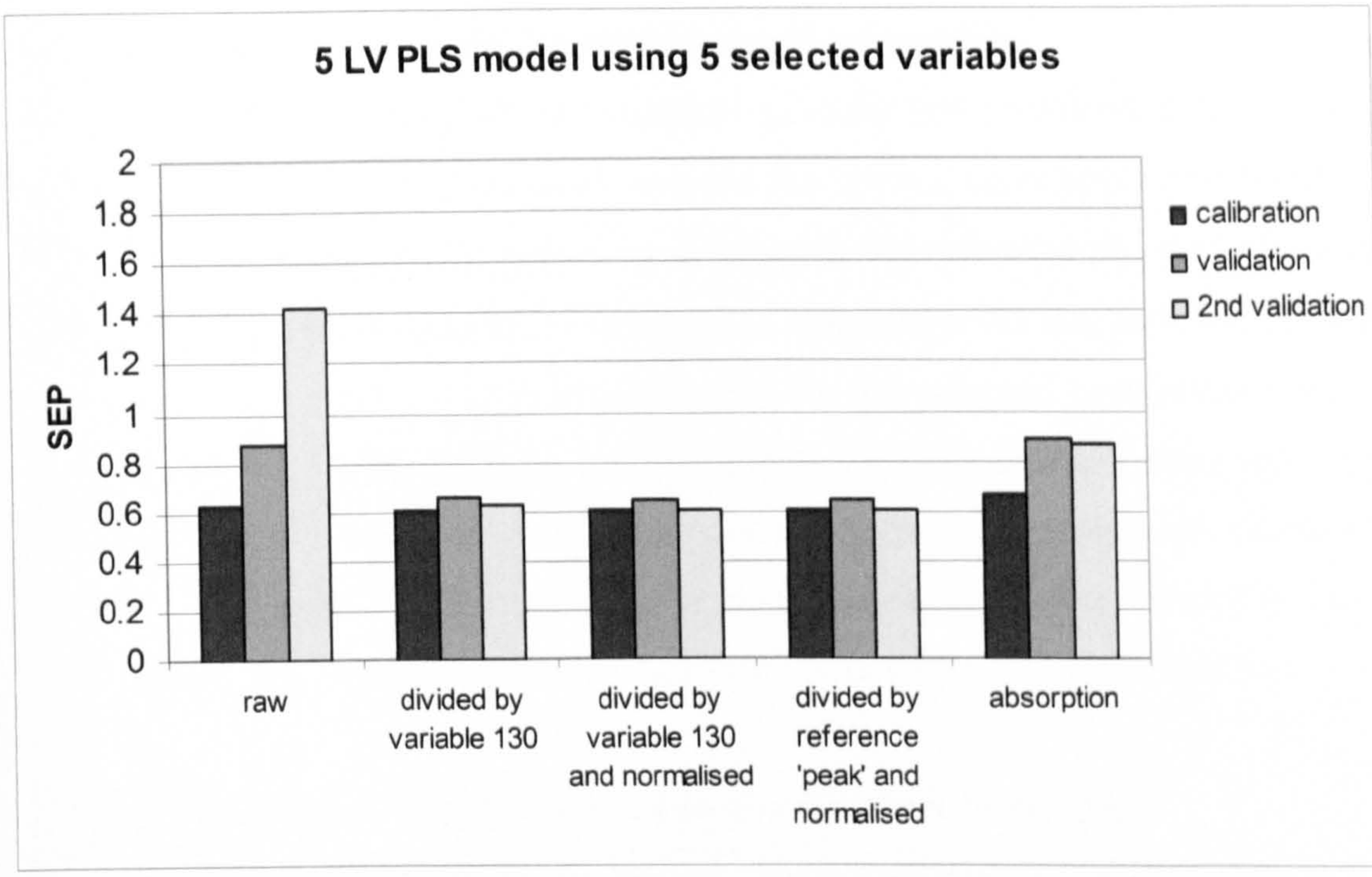


Figure 3.2.12 Results from PLS models using selected wavenumbers

In this case, the lowest SEPs were obtained for data that had been divided by a reference point and the effect of normalising the data seemed to make little difference. For this data treatment, the errors in prediction for the validation data were approximately the same as those obtained for the calibration data. Comparing these to the SEPs from the 7 latent variable models built from the entire spectra it can be seen that they are similar to the SEPs obtained for the validation data.

3.2.2.3 Conclusion

This section compared PLS models built from differently pre-treated data. This compared absorption data (i.e. sample scans ratioed against background scans) against raw (unratioed scans) and data divided by a SF_6 reference peak and data divided by a wavenumber where no absorption occurred (variable 130).

When only five latent variables were employed the absorption data gave the best results (with SEPs of less than 0.8); however, when the number of latent variable were increased to seven there was no longer any significant difference in prediction errors between absorption data and data divided by reference point models. This proved that dividing the data by a reference point was an effective method of data pre-treatment and

has the advantage over absorption data in that there was no requirement to collect background scans.

Also shown from the results was that dividing the data by a point where no absorption occurred was just as effective as dividing the data by a reference SF_6 peak, therefore the addition of SF_6 was dropped from further sample preparation.

Models built from the five selected variables gave the best results for data divided by reference points; however, it should be noted that the wavenumbers were chosen from VS-MLR performed on data treated by dividing by the reference SF_6 peak hence these selected wavenumbers may not necessarily be optimal for the raw and absorption data. What was clear is that the SEPs obtained with the five selected wavenumbers for the data divided by a reference point was lower than the SEPs obtained when employing the entire spectra. From this it can be concluded that variable selection was effective in reducing the number of wavenumbers required down to five without significantly reducing the error in prediction and the objective of this work was achieved.

3.2.3 Conversion of wavenumber measurements to wavelengths

The next stage of the work was to employ narrow band filters to restrict the wavenumbers of light reaching a mid IR detector, hence simulating QWIP detectors operating at the selected wavenumbers. Generally in the field of optics and optical engineering, units in wavelengths are preferred to wavenumber measurements, thus narrow band filters are designed and manufactured in terms of the wavelengths they restrict. In order to determine what wavelength of narrowband filters to employ for the next stage of the work and for future work developing QWIPs measuring at specific wavelengths, the wavenumbers selected in the previous selection models were converted to wavelengths (μm).

Converting selected wavenumbers to their equivalent μm wavelength resulted in numbers to several decimal places. For example the selected wavenumber at 2114 cm^{-1} corresponded to 4.73042 μm and 2129 cm^{-1} corresponded to 4.696121 μm. The interval between measurements was determined by the FTIR which in turn was based on the resolution used, at 32 cm^{-1} resolution this corresponded to measurements made every 15.4 cm^{-1}, this meant that due to the inverse relationship between wavenumbers and wavelengths that the wavelength interval between measurements increased at longer

wavelengths (e.g. the interval between 500 cm^{-1} and 515.4 cm^{-1} corresponds to 0.6 μm, whereas the distance between 5000 cm^{-1} and 5015.4 cm^{-1} corresponds to 0.006 μm).

To find out the effect of rounding μm measurements to two decimal places and to the nearest 0.05, e.g. rounding 4.73042 μm to 4.73 and 4.75 μm, the Bio-Rad win IR software was utilised (by converting the units to wavelengths and using xy coordinates on the corresponding spectral plot) to determine the corresponding response.

Figure 3.2.13 is a plot of the data collected (this included wavelengths 10.6 and 10.62 μm (corresponding to SF_6 absorptions) and 3 μm (region with no absorptions)) and figure 3.2.14 displays the correlation between these measurements and the concentrations of the different gases. For comparison, responses were also collected at wavelengths quoted in other references as being the main absorption bands (based on the HITRAN[98] database). Results using these measurements were compared to those obtained from the selected wavelengths.

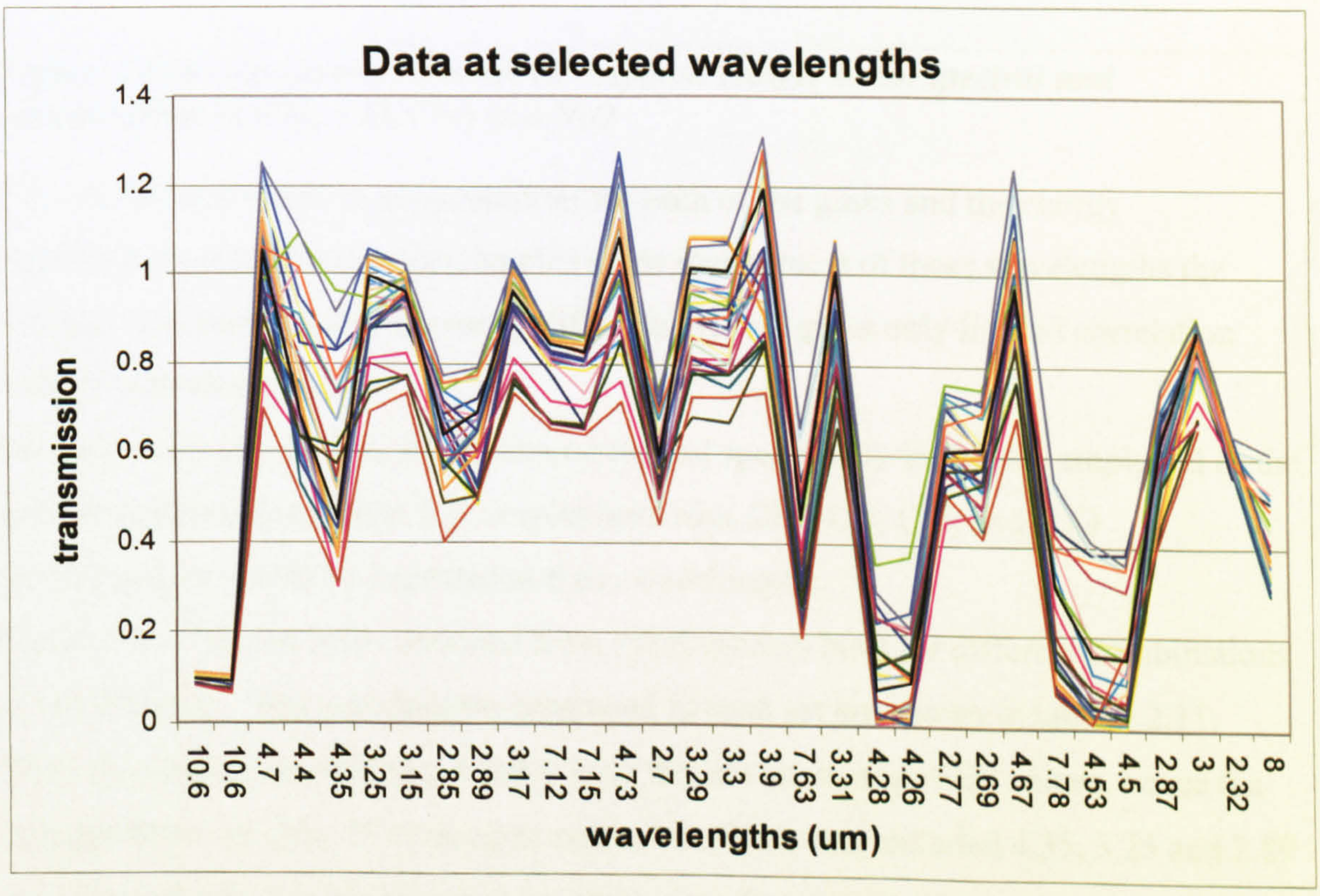


Figure 3.2.13 Plot of IR response (for single beam spectra collected from four component samples mixtures of CH_4, CO, CO_2 and N_2O) for selected wavelengths

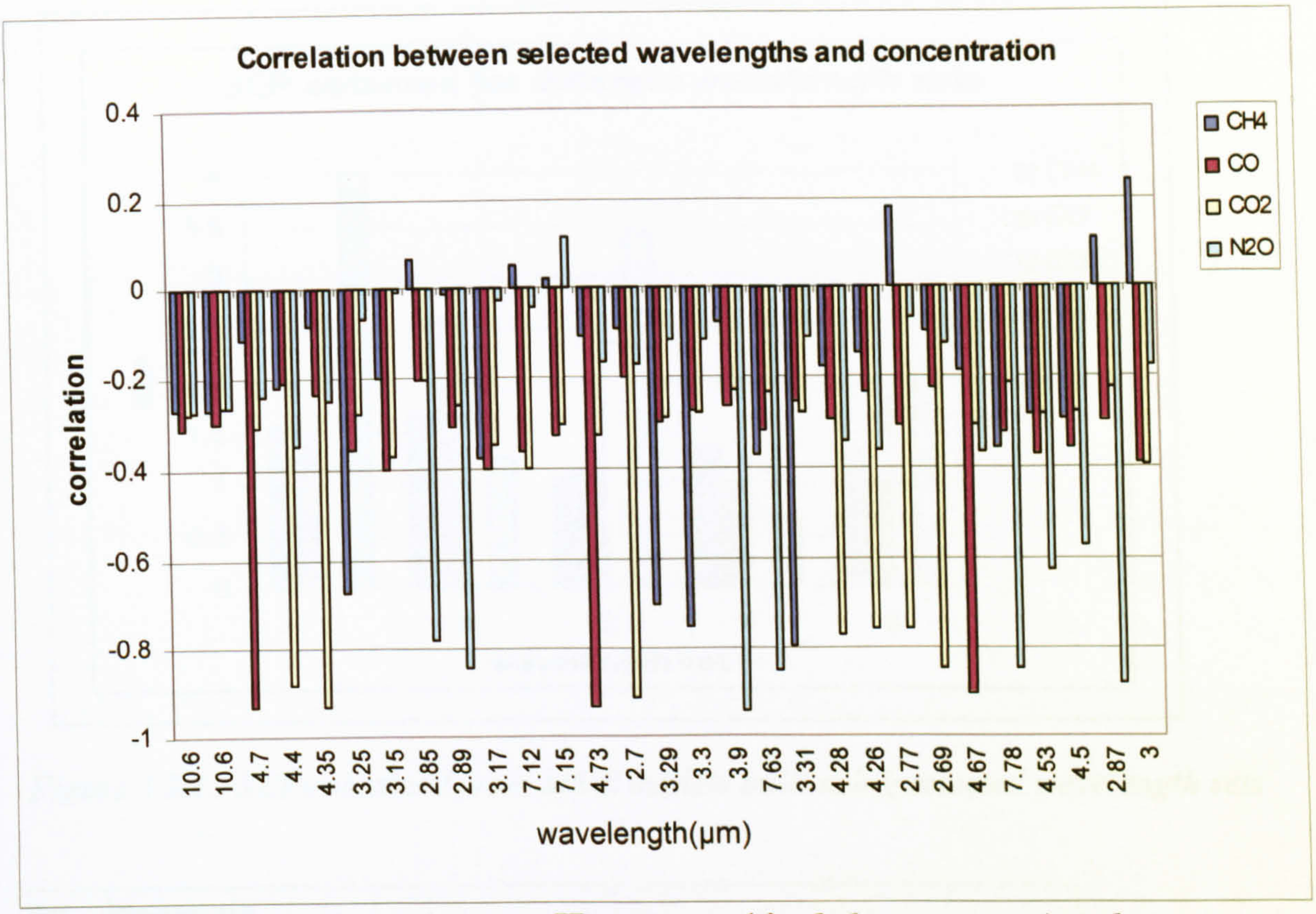


Figure 3.2.14 Correlation between IR response (single beam spectra) and concentration of CH_4, CO, CO_2 and N_2O

The correlations between concentration for each of the gases and the energy transmission at the chosen wavelengths show that at most of these wavelengths the response correlates highly to one specific analyte and gives only limited correlation with the remaining gases.

Measurements at 10.6 and 10.62 μm correspond specifically to the SF_6 employed as the internal reference; therefore low correlations with CH_4, CO, CO_2 and N_2O concentrations would be expected at these wavelengths.

Figure 3.2.15 shows SEPs obtained from MLR models built for different combinations of wavelengths. The wavelengths employed in each set are shown in table 3.2.11. From the results obtained sets 4 and 5 gave the lowest average SEP values, hence the best prediction results. Wavelengths common to both sets included 4.35, 3.25 and 2.89 μm. From figure 3.2.14 it can be seen these wavelengths gave high (negative) correlation to CO_2, CH_4 and N_2O.

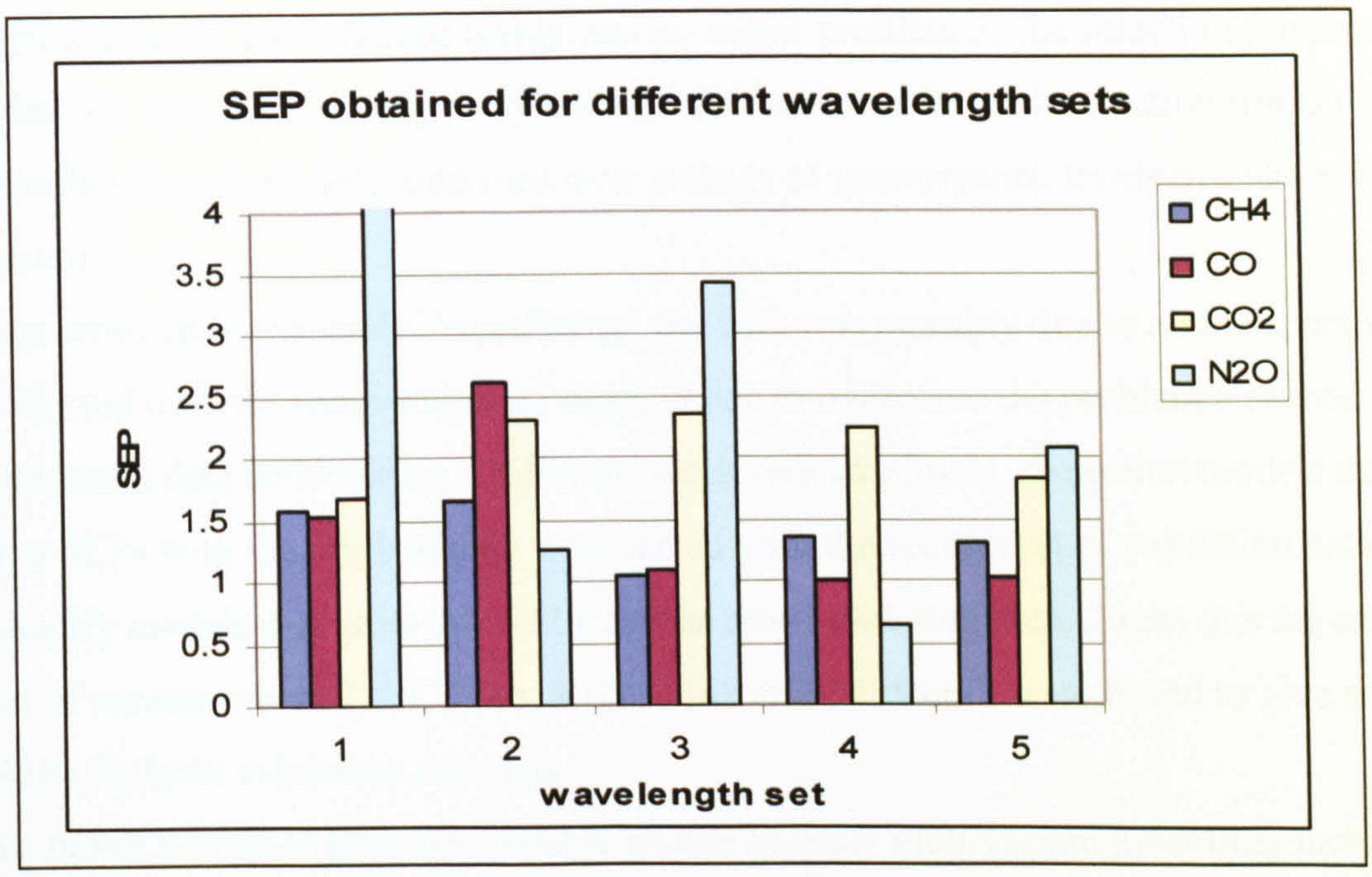


Figure 3.2.15 SEPs obtained from MLR models built using selected wavelength sets

Set	Wavelengths
1	10.6,10.62,4.7,4.4,4.35,3.25,3.15,2.85,2.89,3.17
2	10.6,10.62,4.7,4.4,4.35,3.25,3.15,2.85,2.89,3.17,7.12,7.15,4.73,2.7,3.29,3.3,3.9
3	4.7,4.35,3.25,3.15,2.9
4	4.7,4.35,3.25,3.15,2.89,3.17,7.12,2.7,3.29,3.9
5	4.35,3.25,2.89,7.12,4.73,

Table 3.2.11 Wavelengths employed in selected wavelength sets

Based on these results it was decided to carry out further investigations into the combination of wavelength measurements at 2.9, 3.25, 4.35, 4.73 and 10.62 μm. The results of employing narrowband filters centred at these wavelengths are discussed in section 3.4.

3.2.4 Conclusion

The aim of this stage of the work was to determine sets of wavenumbers, suitable for incorporating into a detector array, that give comparable results to models based on the entire spectra. As demonstrated by the results obtained this aim was achieved. Investigations employing VS-MLR brought out two weaknesses of the process, the first that the values of the convergence levels had a major impact on the final wavenumbers selected and that there was a tendency for the model to 'overfit'.

Optimising the convergence levels was no major problem as the selection process was relatively quick, on average only taking a few minutes for each selection run, so it was possible to perform selection runs over a range of convergence levels in a short time period.

The problem of the model 'overfitting' the data was possibly due to all the spectra being collected over the same couple of days, hence to overcome the problem a second set of validation data collected on a different week was employed. Selection models that gave low SEPs with the original data were tested with the second set of validation data to identify models that gave low SEPs for the new validation data. From this the selected set of wavenumbers 2129, 2284, 3071, 3148 and 3488 cm^{-1} were found to give similar SEPs for both validation data sets.

To prove VS-MLR gave comparable results to other multivariate modelling techniques, PLS was employed. PLS is commonly employed for analysis of FTIR spectra and its effectiveness is shown in work by Griffiths *et al* [89] and Spiegelman *et al.*[91]

PLS models were built for raw data, data divided by the SF_6 reference peak, data divided by a wavenumber where no absorption occurred and for absorption data. Comparing five latent variable models showed that the absorption data gave the best results; however, at seven latent variables the 'dividing by a reference point' treatment gave comparable results to the absorption data, thus proving that the data treatment was effective.

Another important conclusion drawn from the results was that there was very little difference obtained between data divided by the SF_6 reference peak and data divided by a wavenumber where no absorption occurred, thus SF_6 was no longer added to further samples.

3.3 Dilution system

The aim of this work was to set up a gas dilution system incorporating a GC as an independent means of measuring gas concentrations.

The system was based on an exponential dilution system, as mentioned in the introduction, the aim of which is to start with a known concentration of calibrant, which is diluted with a carrier gas at a controlled rate. This allows a range of different concentrations to be produced from just one original standard mixture, and can be used to evaluate detector performance over of wide concentration range and determine limits of detection.

3.3.1 Trial run of dilution system

Initially the system was tested with a gas mixture standard of 1000 ppm each of CH_4, CO, CO_2 and N_2O made up in nitrogen, this was measured with both the GC (with the response measured on chart recorder) and FTIR.

The GC carrier gas employed was helium using TCD detection; hence the response obtained from the GC was a single peak resulting from the high concentration of nitrogen (this was due to the high conductivity of nitrogen compared to helium), which swamped the response to the other gases.

Figure 3.3.1 is a plot of the GC response, using the peak height recorded by the chart recorder, against time. The peak height is directly proportional to the concentration of sample, hence over time the peak height drops exponentially as the sample gets more dilute. Plotting the log of the peak height against dilution time gave a linear plot, with a R^2 correlation of 0.9978 (figure 3.3.2), which indicated the nitrogen in the system was diluted exponentially and it can be assumed the other gases in the sample were similarly diluted.

The IR spectra from the trial run of the dilution system were collected as single beam spectra i.e. transmission, hence as the sample was diluted less light was absorbed, resulting in an increase in the amount of light transmitted. The spectra collected are shown in figure 3.3.3. Plots of IR response versus time for wavenumbers relating specifically to the gases of interest are shown in figure 3.3.4; these give R^2 values of greater than 0.98. The different gradients are due to the different molar absorptivity coefficients at the chosen wavenumbers.

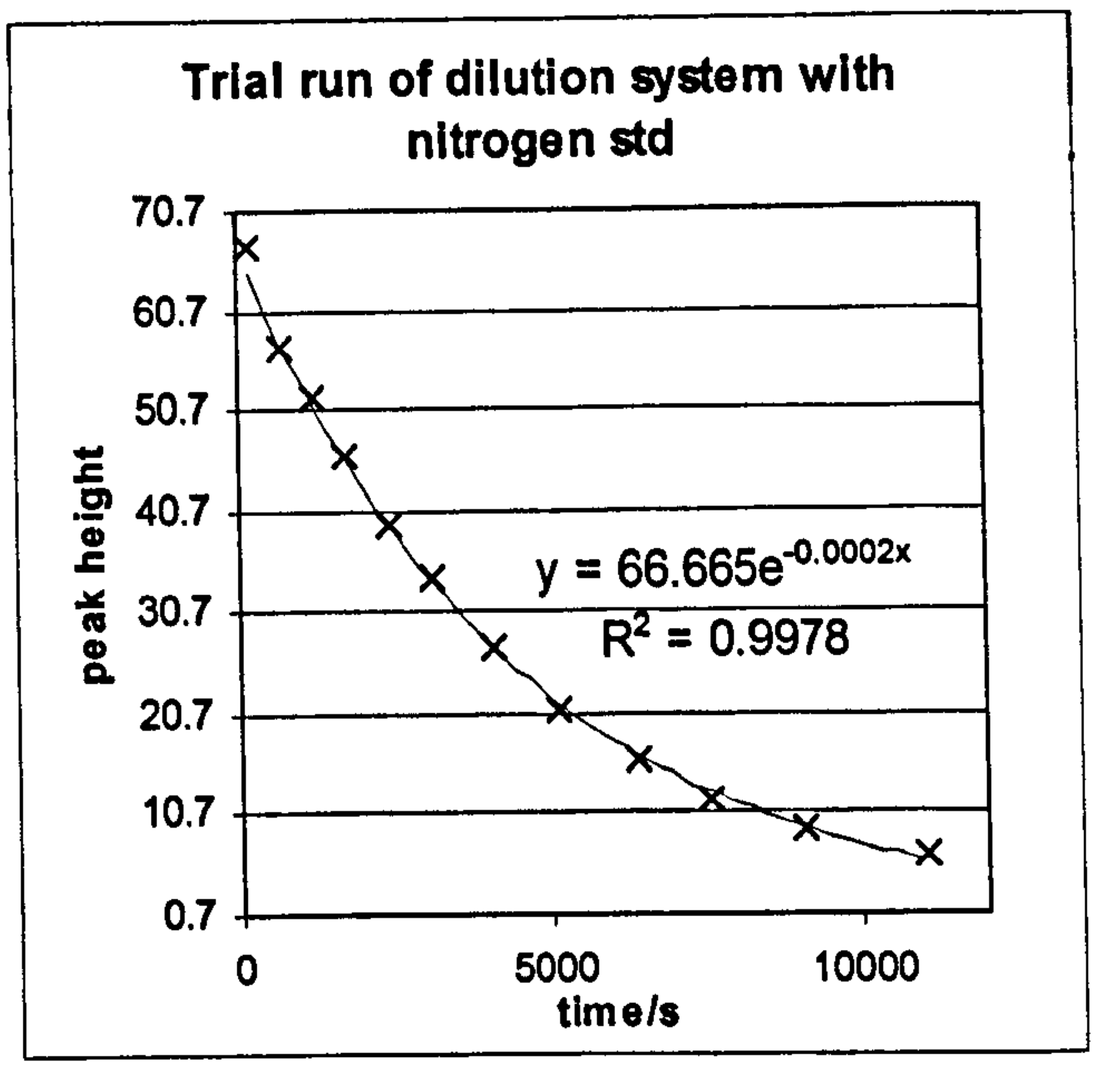


Figure 3.3.1 GC response (peak height) versus time for trial run of the dilution system using the 1,000 ppm gas standard made up in nitrogen

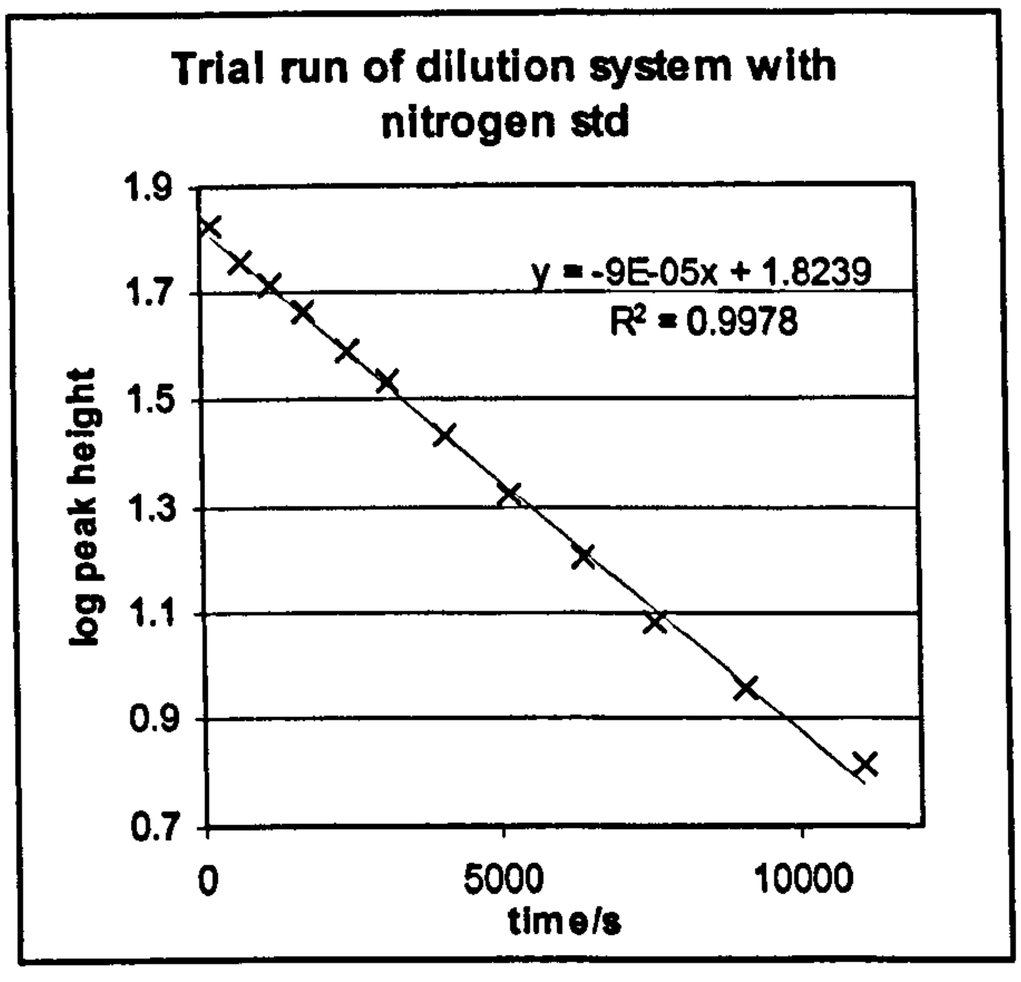


Figure 3.3.2 Log of GC response (peak height) versus time for trial run of the dilution system using the 1,000 ppm gas standard made up in nitrogen

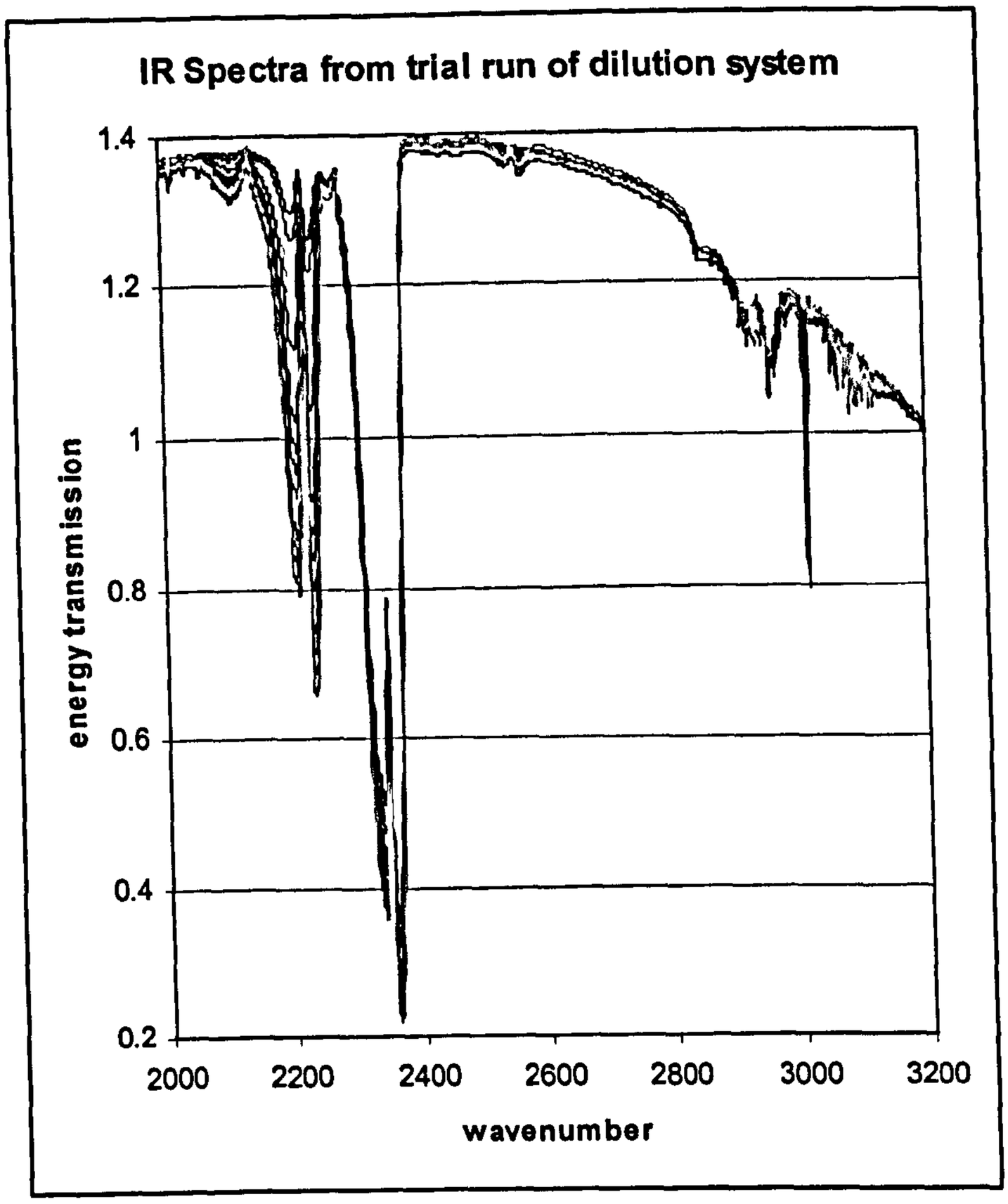


Figure 3.3.3 Close-up of the region 2000-3200 cm^{-1} for IR spectra collected during the trial run of the dilution system

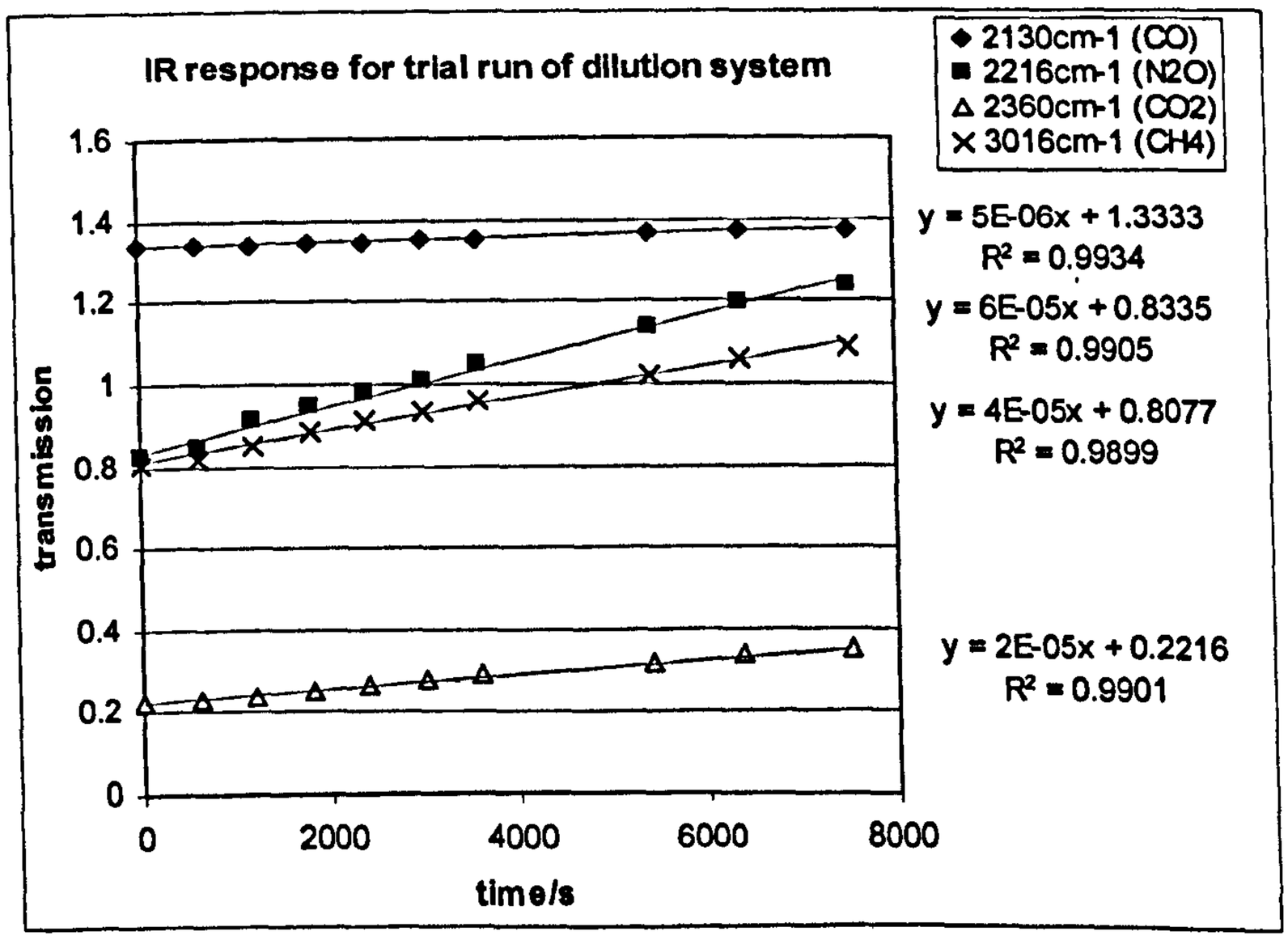


Figure 3.3.4 Response (from trial run of dilution system) versus time for specific wavenumbers relating to gases of interest (NB. The gap in measurements between 4000 and 5000 was due to an interval in taking measurements)

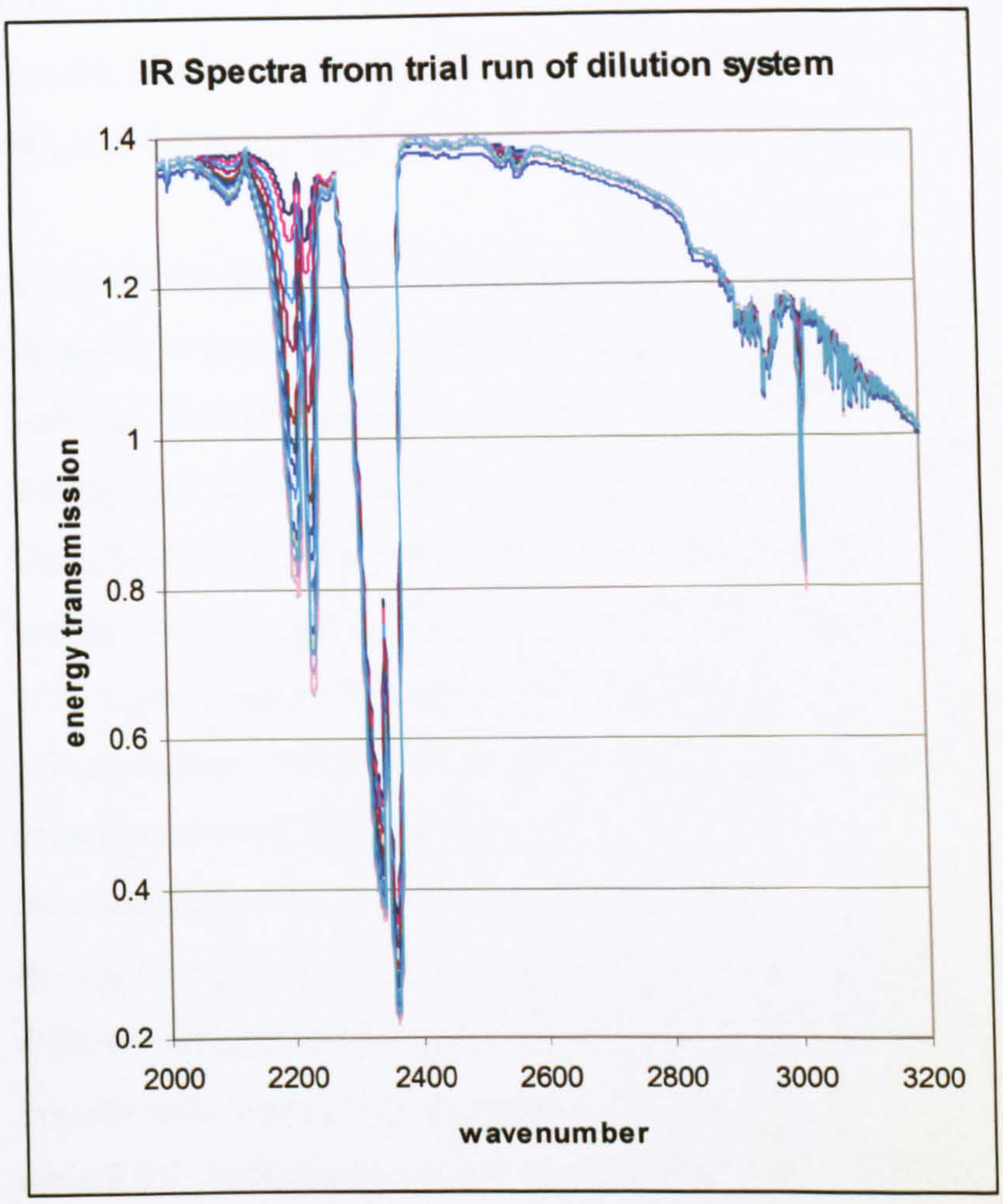


Figure 3.3.3 Close-up of the region 2000-3200 cm^{-1} for IR spectra collected during the trial run of the dilution system

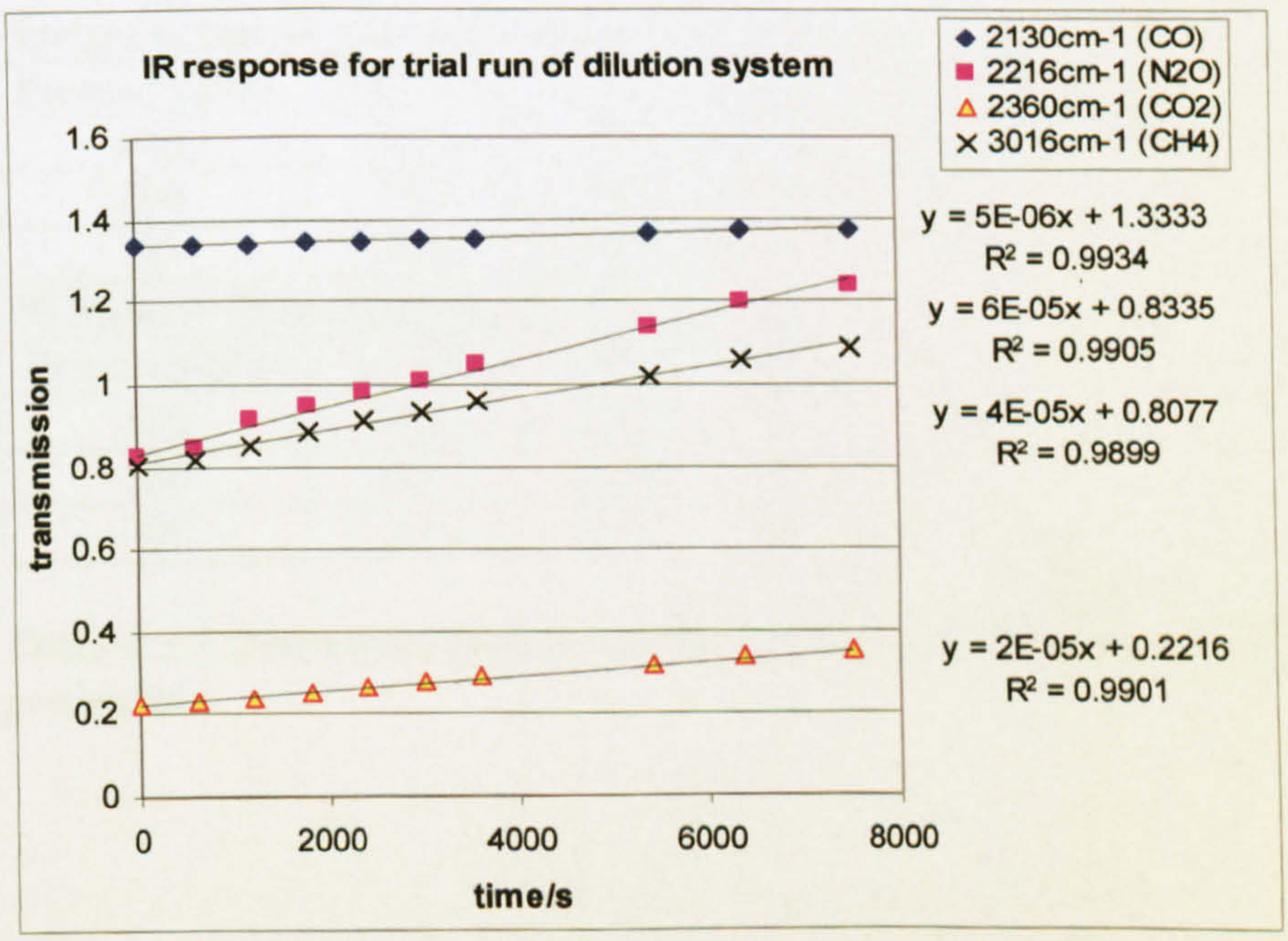


Figure 3.3.4 Response (from trial run of dilution system) versus time for specific wavenumbers relating to gases of interest (NB. The gap in measurements between 4000 and 5000 was due to an interval in taking measurements)

From these results it could be assumed the basic dilution set-up was operational, therefore further work was undertaken to optimise the GC conditions and LabVIEW^{TM} program for data recording using a gas standard mixture made up in helium.

3.3.2 Optimisation of GC conditions

Originally samples of the individual gases were injected directly into the GC, but problems with backpressure made injections difficult, so samples were prepared in the mixing flask using a helium flow to fill the gas-sampling valve on the GC.

This allowed individual retention times to be determined; however there was still a problem with air getting into the system resulting in a GC response to nitrogen.

With a gas standard mixture of the four gases (CH_4, CO, CO_2 and N_2O) made up in helium instead of nitrogen, the effect of different GC conditions could be more fully investigated to optimise resolution.

Increasing the oven temperature of the GC only served to decrease resolution between the gases, therefore the oven was left at room temperature (25°C) and the effect of different column pressures (i.e. flow rates) and detection temperatures were investigated. The change in retention times with different column pressures is shown in table 3.3.1, with some of the GC traces obtained from the chart recorder shown in figure 3.3.5.

Original 1m column filled with Porapak Q				
Pressure(kPa)	N_2/ CO	CH_4	CO_2	N_2O
400	25 s	35	75	100
200	70 s	80	150	200
60	70 s	100	180	230
With new (5m) column				
Pressure(kPa)	N_2/ CO	CH_4	CO_2	N_2O
100	140 s	190	360	460
160	80 s	120	240	310
200	85 s	110	220	280

Table 3.3.1 Retention times (s) obtained for the different gases using different column pressures

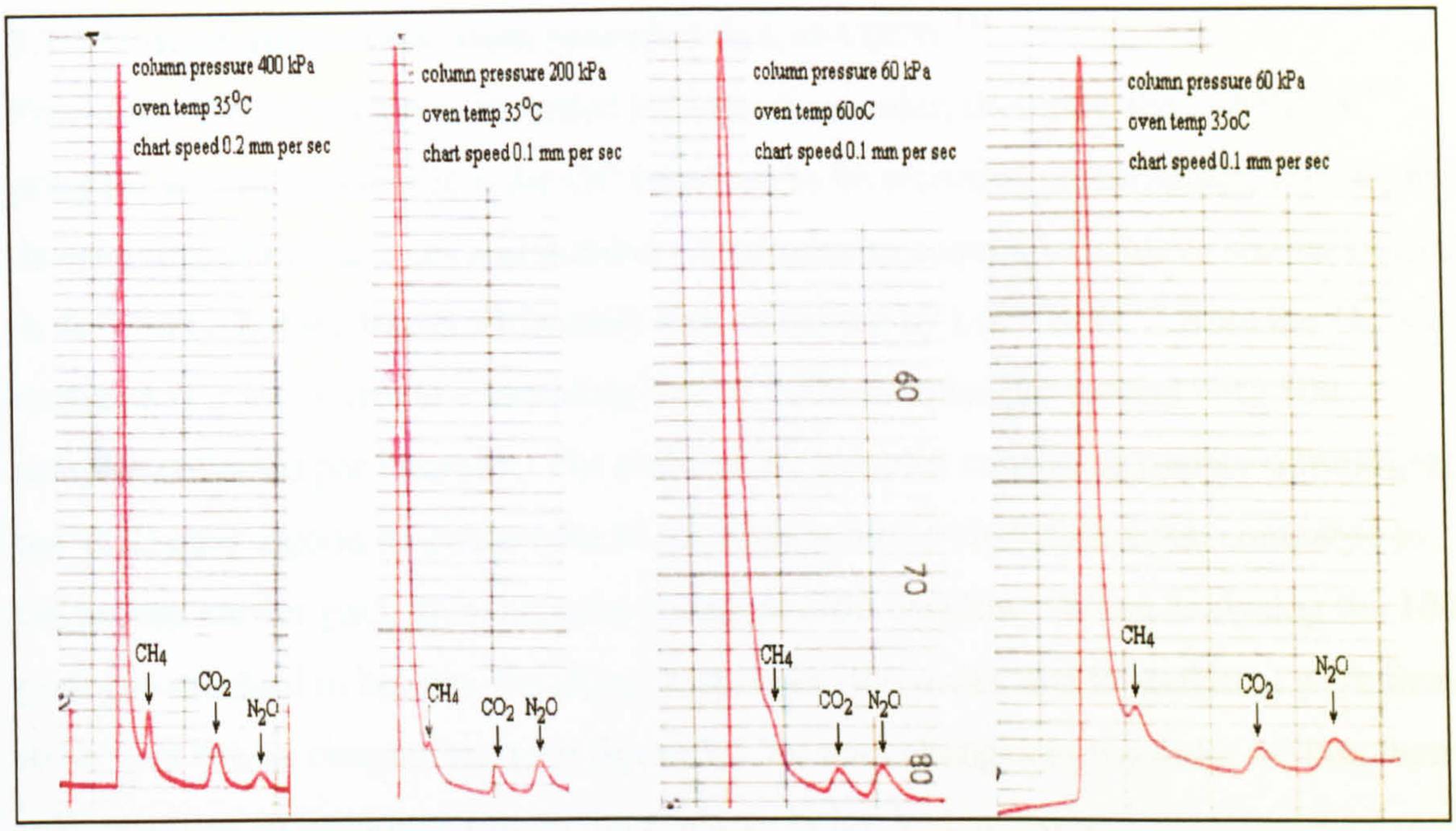


Figure 3.3.5 GC traces obtained using different column pressures, (NB. The first peak is due to both N_2 and CO, hence the high response)

The original sample loop was 3 ml, which was later increased to 10 ml. Overall the best conditions were; oven at room temp (25°C), detector temperature at 180°C, the filament temperature at 230°C, attenuation set to 1 and gain at x10.

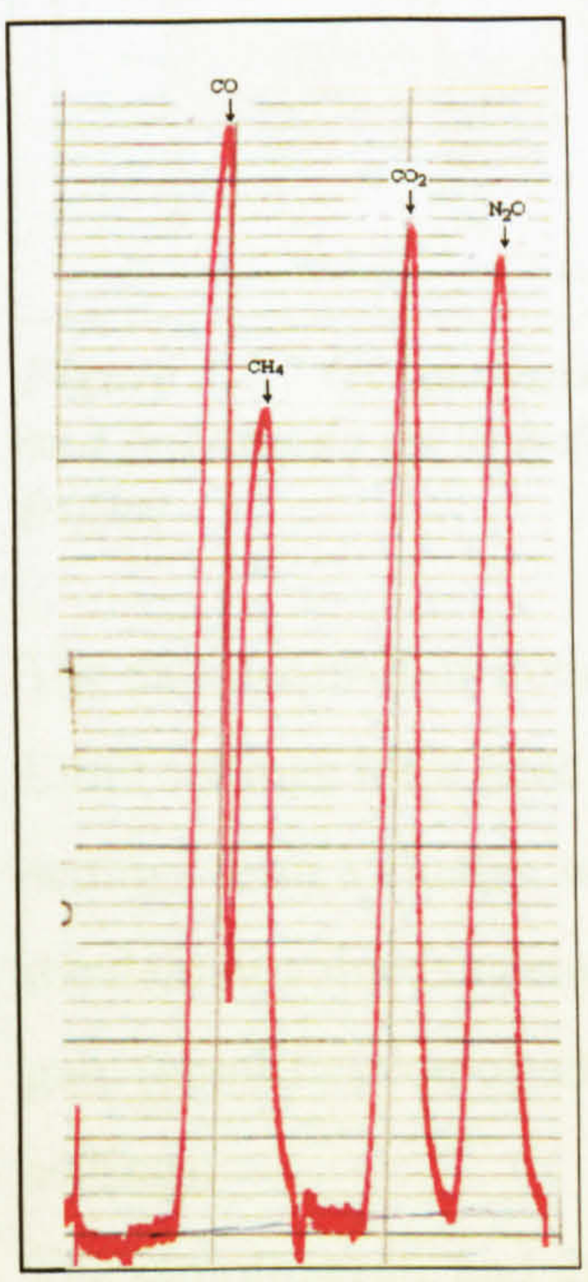


Figure 3.3.6 GC trace obtained for 5 m column

3.3.3 Optimisation of GC data recorded in LabVIEW™

Previously the GC data was recorded on a chart recorder, therefore the LabVIEW™ program was written to allow the GC response to be recorded on computer, this required determining sampling rates and number of samples to acquire to achieve similar results to the chart recorded traces. Originally data collected by LabVIEW™ from the GC was acquired as a waveform at a sampling rate of 1000 samples per second with 100 samples collected per iteration. For ambient air samples consisting mainly of nitrogen, the TCD gave a good response (due to the high conductivity of nitrogen compared to the helium carrier gas), thus the noise to signal ratio was low. When analysing the 1000 ppm gas standard in helium, the signal responses (the peaks due to analytes) were found to be very low in comparison (see figure 3.3.7), with changes in the order of less than 1 mV, resulting in the signal due to the analytes being lost in 'noise'.

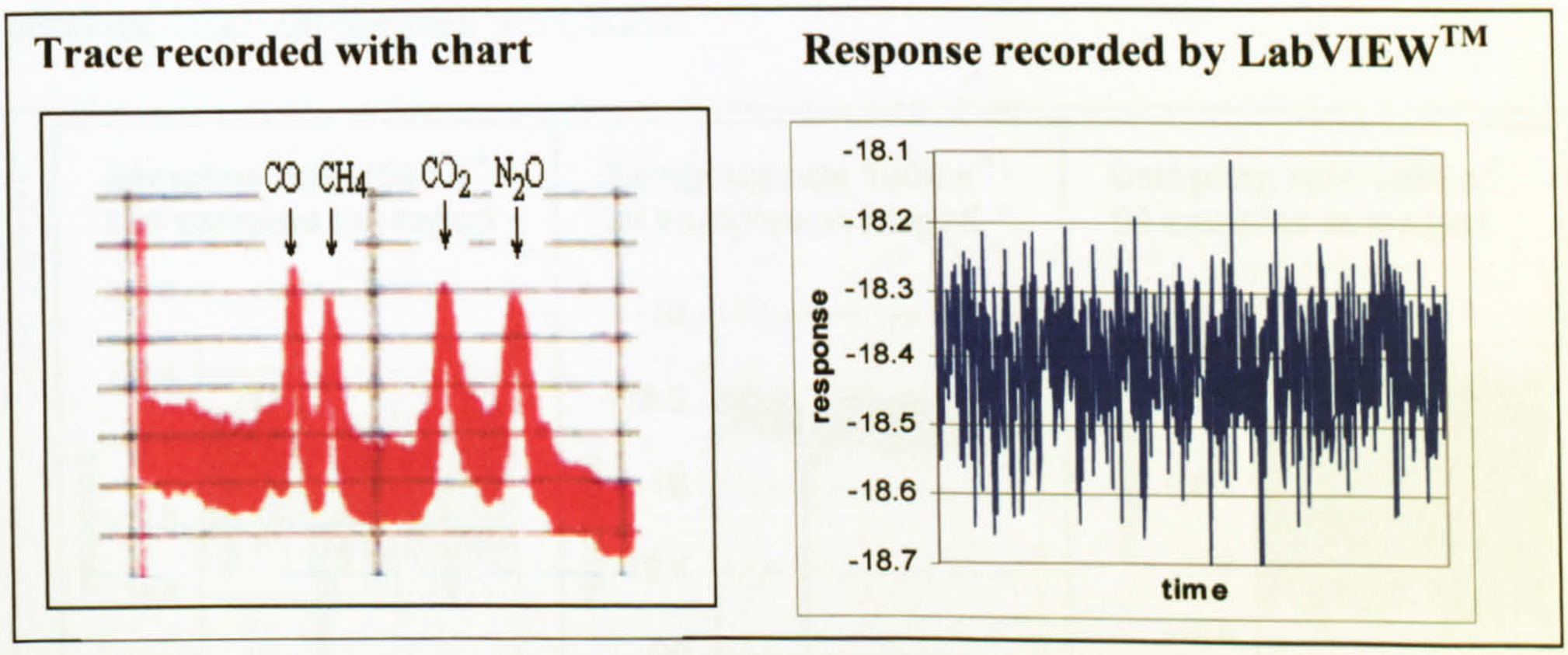


Figure 3.3.7 GC response for standard mixtures of 1,000 ppm each of CO, CO_2, CH_4 and N_2O made up in helium, recorded by chart recorder (left) and in LabVIEW™ (right)

The data acquisition conditions in LabVIEW™ were modified to rectify the situation. Limit settings were used to set the maximum and minimum expected values. The limit settings form a cluster, which assigns limits to the channels in the channel string array; the default value is normally ±10 V, which in this case meant limited sensitivity. The gain (amplification or attenuation) is determined by the programming based on the limit settings.

The sampling rate and the number of samples averaged were both altered in order to find the optimum conditions. Due to the low response to the gas sample mixture, ambient air was let into the system so that the nitrogen peak could be used as an indication of whether the signal to noise ratio had improved.

At high sampling rates with a low number of samples taken, the signals are lost in the high levels of noise. The optimum sampling rate appeared to be 600 samples per second. Increasing the sampling rate resulted in high 'noise' levels, which combined with a low number of samples taken resulted in no peaks being present; using sampling rates of less than 600 also had the effect of increasing the noise levels.

The number of samples taken had a similar effect, i.e. too many or too few samples resulted in a loss of the signal peak. It can be assumed that the noise is due to electrical fluctuations, therefore increasing the number of samples taken helps average out the noise; however, if too many samples are taken the change in signals due to sample peaks also gets 'averaged out'. This is demonstrated in figure 3.3.8, which compares the response obtained for different sampling rates and number of samples taken.

For a sampling rate of 600 samples per seconds the best signal to noise ratio was obtained when 60 samples were taken.

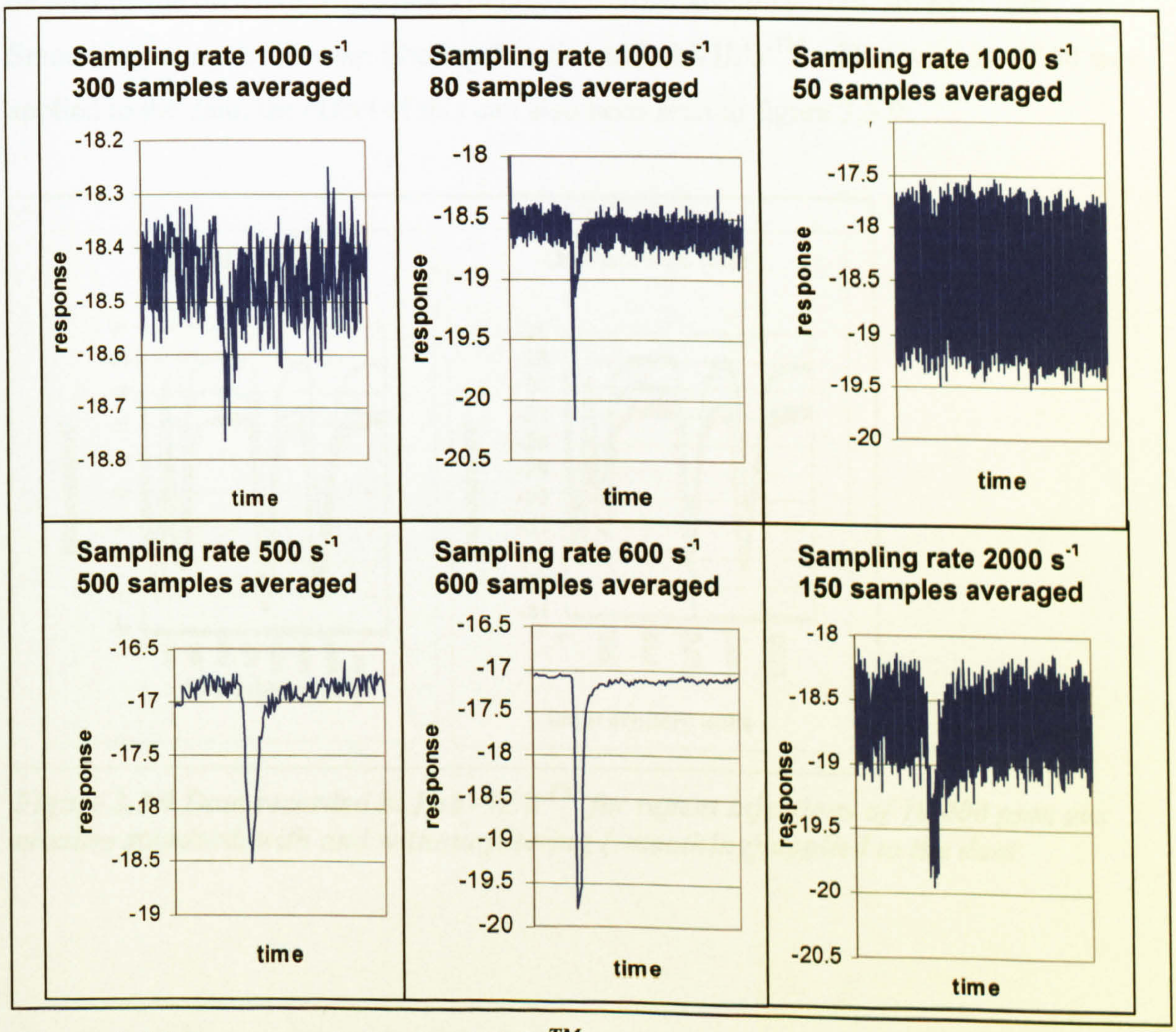


Figure 3.3.8 Data recorded in LabVIEW™ using different sampling rates and number of samples averaged (NB. Response is in mV)

3.3.4 Reproducibility of injections

Repeat injections of the undiluted 1 % standard mixture, were performed to assess the reproducibility of response (using column pressure of 200 kPa, 5 ml sample loop, with sampling rate of 600 s^{-1} and 60 samples averaged for data acquisition in LabVIEW^{TM}). The peak integration and peak detection functions available in the LabVIEW^{TM} software were incorporated into the LabVIEW^{TM} program. These functions required threshold limits to be inputted in order to identify peaks and as the GC employed in this work was prone to baseline shifts this caused major problems. Baseline drifts above the threshold level resulted in an increased number of peaks detected. Subsequently, peak areas were determined using Excel, where baseline shifts could be compensated for before calculating the peak area.

Figure 3.3.9 displays data collected for four repeat injections of the 10,000 ppm gas mixture standard; from this the change in response due to instrument drift over time can be seen by the different magnitudes obtained for the baselines from different runs. Smoothing using the median filtering function in LabVIEW^{TM} with a window of 50 was applied to the data; the effect of this can also been seen in figure 3.3.9.

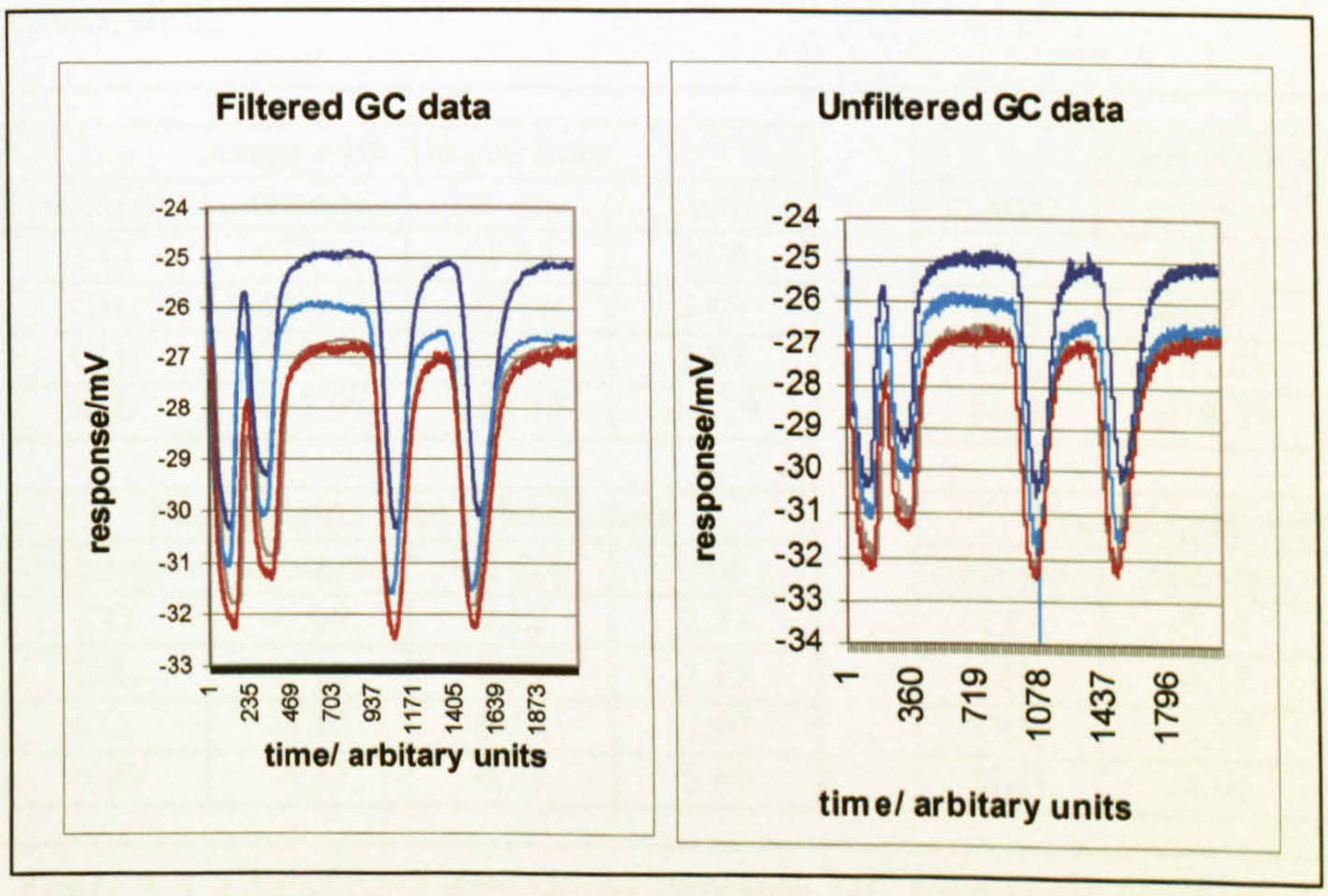


Figure 3.3.9 Data recorded in LabVIEW^{TM} for repeat injections of 10,000 ppm gas mixture standard, with and without filtering (smoothing) applied to the data

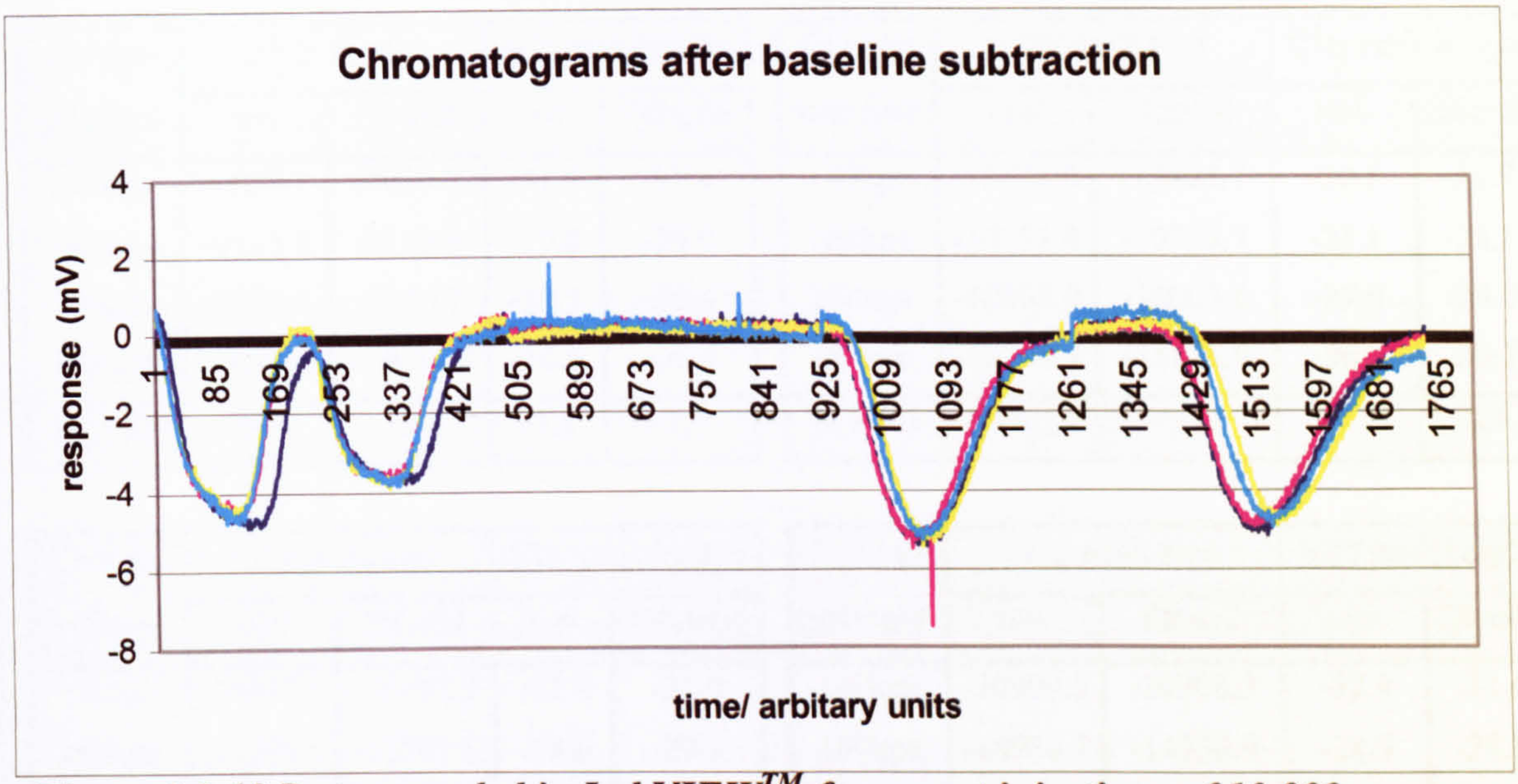


Figure 3.3.10 Data recorded in LabVIEW™ for repeat injections of 10,000 ppm gas mixture standard after baseline correction

Figure 3.3.10 is a plot of the data after baseline correction. The standard deviation and RSD between replicate samples is displayed in table 3.3.2. The filtered data generally gave lower deviation in responses than the unfiltered data, and lower deviations (with one exception) were obtained for peak height measurements than for peak area calculations

Areas with filtered data			
analyte	average	std dev	rsd
CO	-677.73	45.65	6.74
CH_4	-610.52	22.13	3.63
CO_2	-770.90	20.61	2.67
N_2O	-807.97	117.87	14.59

Areas with non-filtered data			
analyte	average	Std dev	rsd
CO	-679.46	52.40	7.71
CH_4	-620.77	28.03	4.52
CO_2	-761.01	29.41	3.86
N_2O	-819.09	111.40	13.60

Peak heights with filtered data			
analyte	average	std dev	rsd
CO	-5.09	0.12	2.37
CH_4	-4.14	0.09	2.15
CO_2	-5.12	0.15	2.95
N_2O	-4.87	0.28	5.69

Peak heights with non-filtered data			
analyte	average	std dev	rsd
CO	-5.14	0.14	2.72
CH_4	-4.17	0.08	1.99
CO_2	-5.17	0.18	3.44
N_2O	-4.91	0.17	3.53

Table 3.3.2 Standard deviations between GC responses (collected in LabVIEW™) for repeat injections of 10,000 ppm standard mixture

Column pressure	CO peak area		CO peak height	
	raw	filtered	raw	Filtered
140kpa	-9923.1	-9929.7	-31.5	-31.4
160kpa	-9145.8	-9150.9	-29.0	-29.0
200kpa	-8124.0	-8130.7	-30.5	-30.4
260kpa	-13048.1	-13048.4	-30.6	-30.6
% RSD	-21.1	-21.1	-3.4	-3.3

Column pressure	CH_4 peak area		CH_4 peak height	
	raw	filtered	raw	filtered
140kpa	-12824.0	-12824.7	-30.5	-30.7
160kpa	-10751.9	-10753.7	-28.1	-28.1
200kpa	-10002.9	-10003.6	-29.6	-29.5
260kpa	-15216.9	-15216.9	-30.2	-30.2
% RSD	-19.2	-19.2	-3.6	-3.8

Column pressure	CO_2 peak area		CO_2 peak height	
	raw	filtered	raw	Filtered
140kpa	-16997.6	-16997.2	-32.0	-32.0
160kpa	-13709.7	-13708.6	-29.0	-29.1
200kpa	-46960.3	-46966.9	-29.3	-29.2
260kpa	-17800.1	-17797.8	-31.3	-31.3
% RSD	-64.9	-64.9	-5.0	-4.9

Column pressure	N_2O peak area		N_2O peak height	
	raw	filtered	raw	filtered
140kpa	-20909.5	-20908.3	-32.4	-32.4
160kpa	-14938.7	-14936.9	-28.9	-28.8
200kpa	-11240.5	-11234.6	-29.0	-28.8
260kpa	-22746.9	-22746.3	-31.4	-31.4
% RSD	-30.5	-30.5	-5.8	-5.9

Table 3.3.3 The effect of different column pressures (Helium flow rates) on the peak areas and heights

Comparing the peak areas and peak heights obtained for different column pressures, the % RSD for peak areas ranged from 19.2 to 64.9 %; whereas, peak heights gave lower RSD, ranging from between 3 and 6%.

3.3.5 Dilution using 1 % mixture with set-up 2

Three dilutions of the 1 % standard mixture were performed using set-up 2 (refer to section 2.4.5). Results from one of the dilution runs are shown in figure 3.3.11, this displays traces obtained with and without mean filtering.

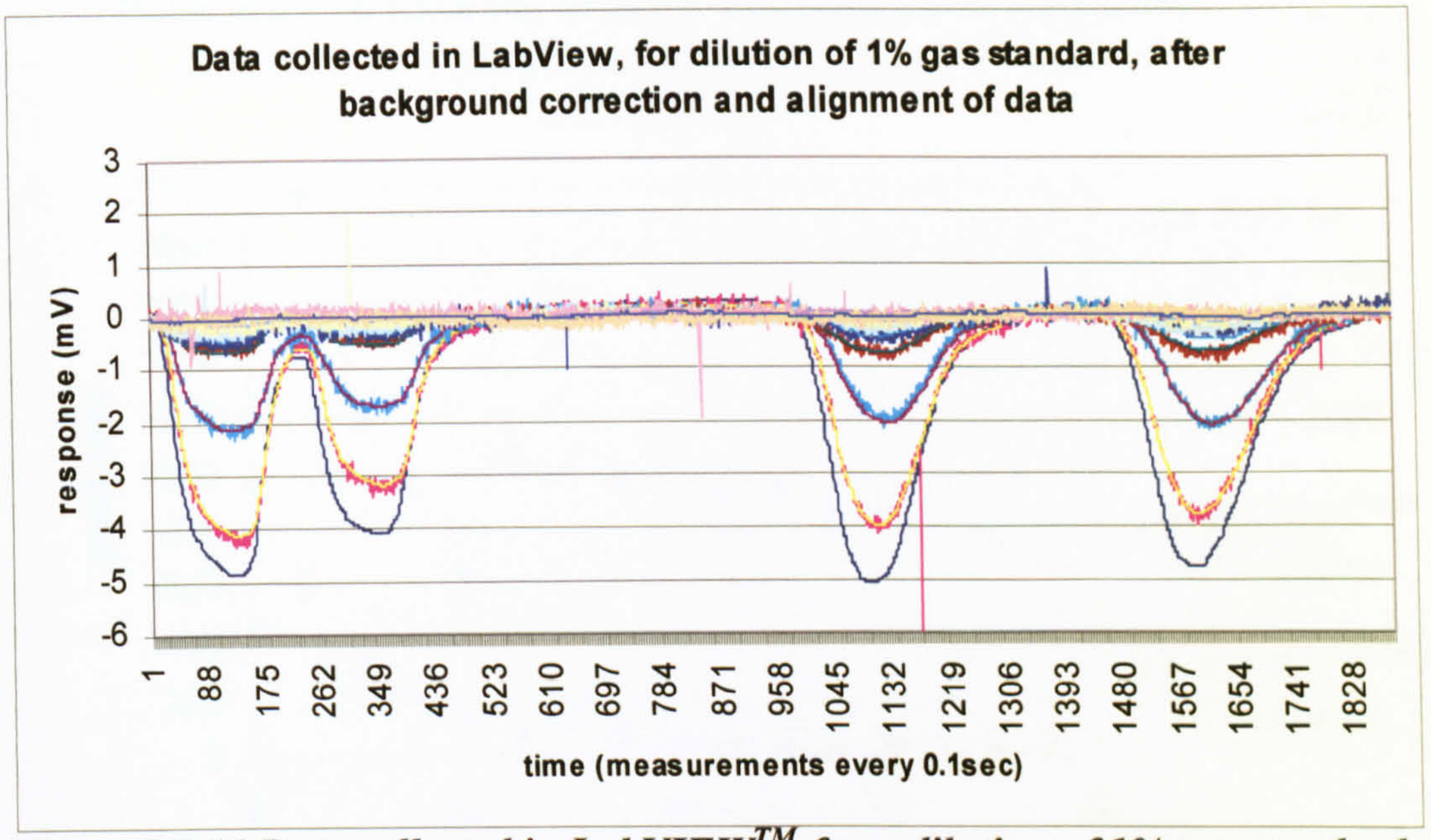


Figure 3.3.11 Data collected in LabVIEW™ from dilution of 1% gas standard mixture using dilution system set-up 2. (NB.Background correction was applied to remove the effect of baseline drift from the instrument and peaks were aligned for comparison purposes)

Plots of peak area versus dilution time are shown in figure 3.3.12, these gave good correlations (greater than 0.95) for exponential trend lines.

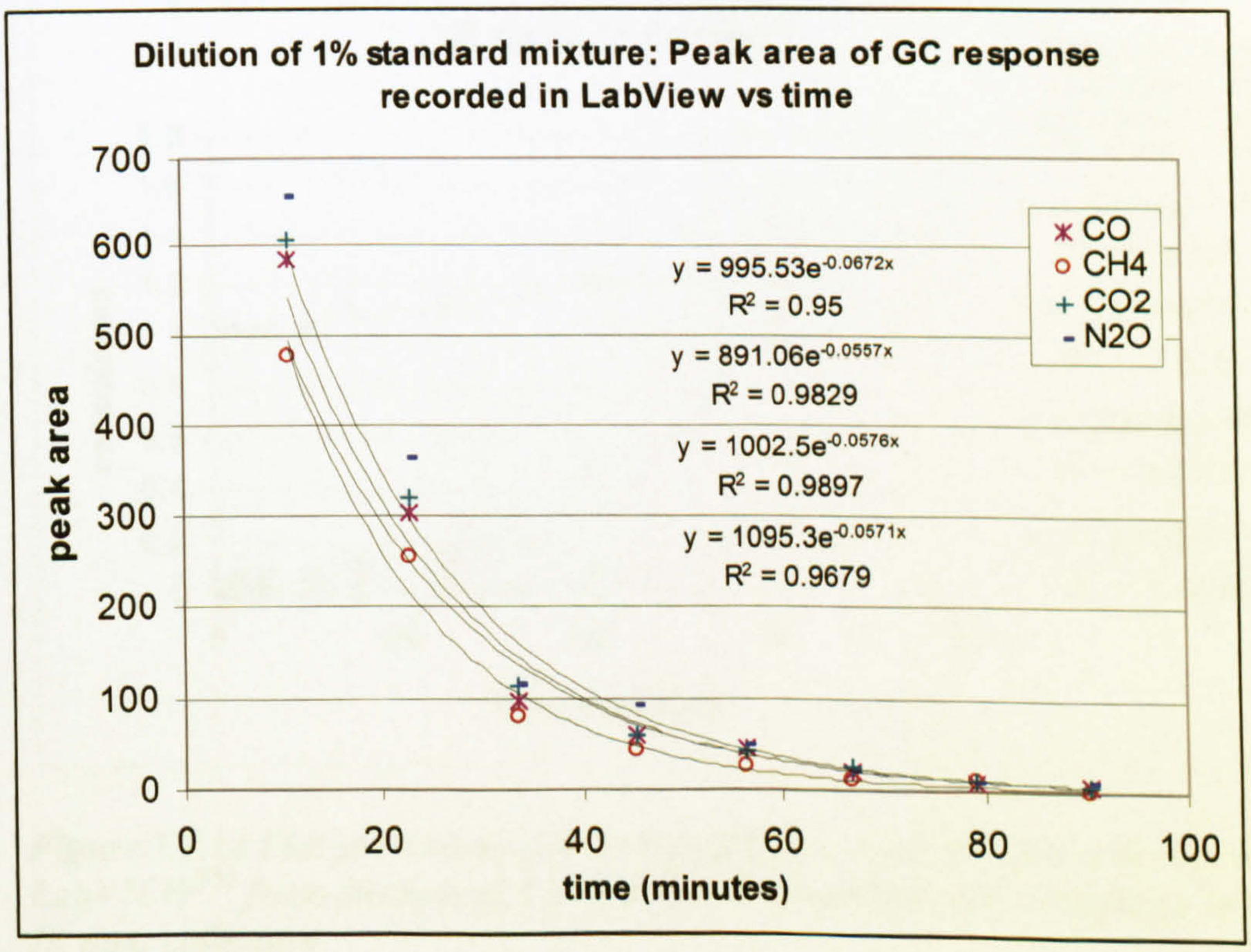


Figure 3.3.12 Plot of peak area versus dilution time for data collected in LabVIEW™ from dilution of 1 % gas standard mixture using dilution set-up 2.

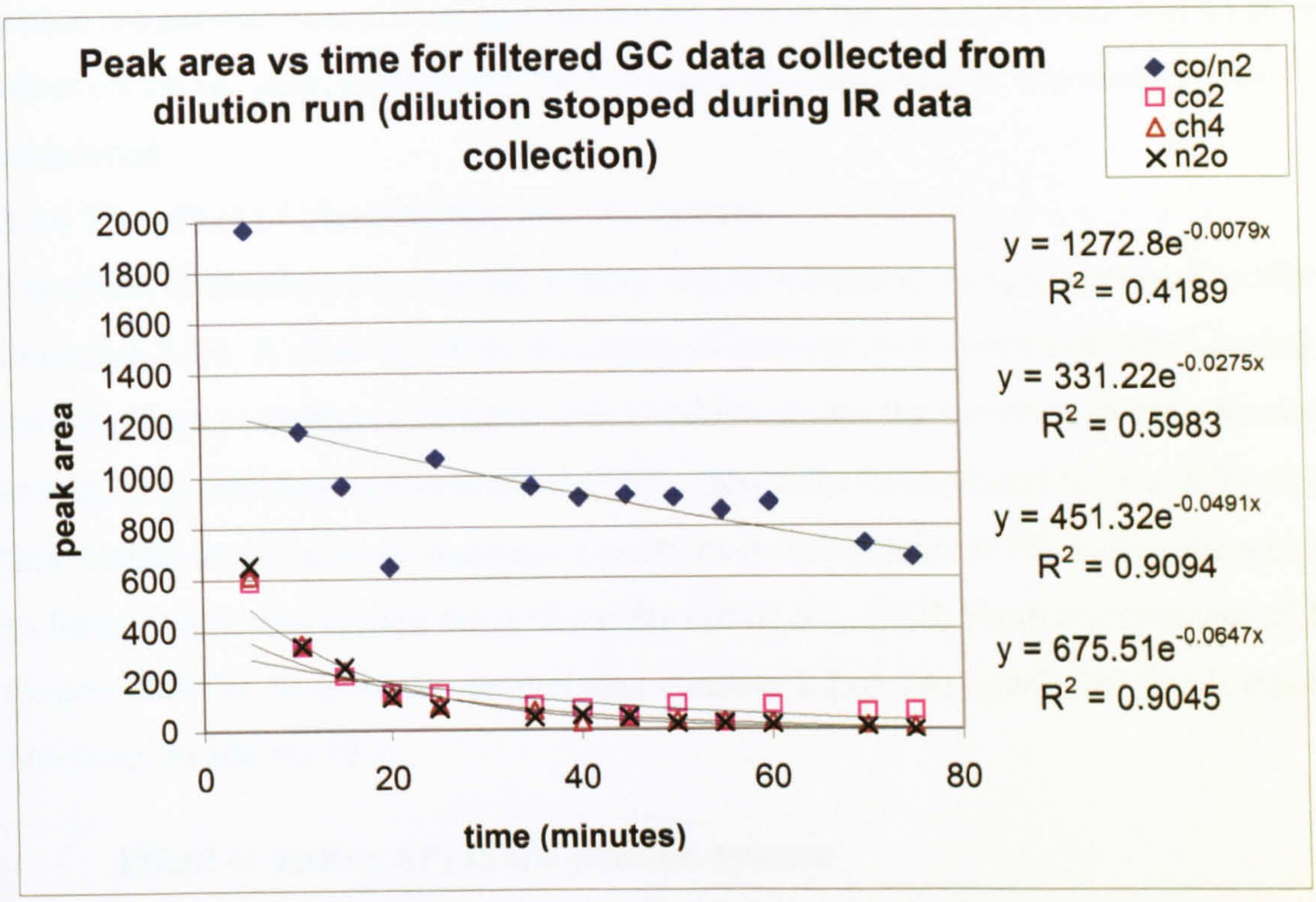


Figure 3.3.13 Plot of peak area versus dilution time for data collected in LabVIEW™ from dilution of 1 % gas standard mixture with stoppages in dilution for IR data collection

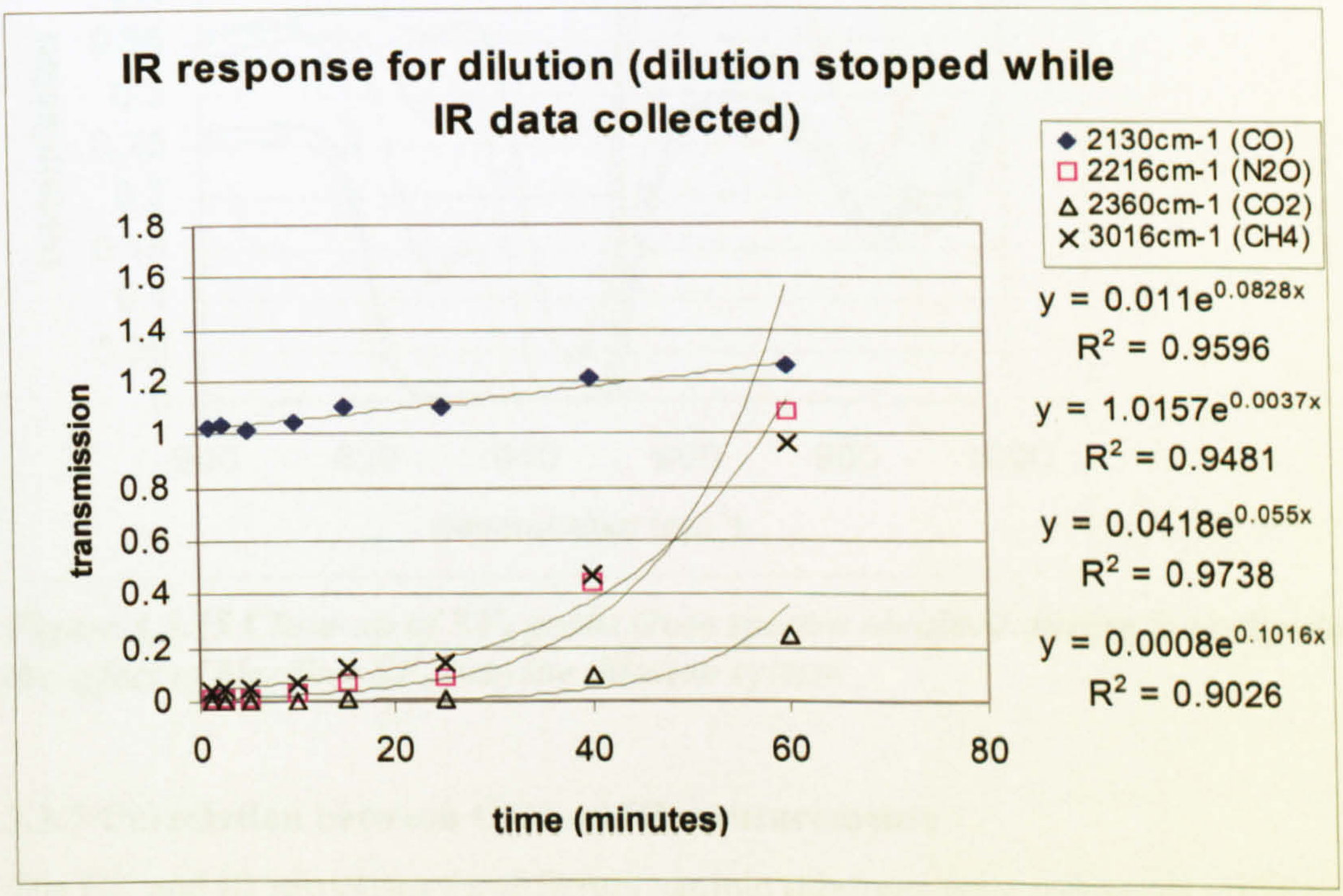


Figure 3.3.14 Plot of transmission versus dilution time for data collected in LabVIEW™ from dilution of 1 % gas standard mixture with stoppages in dilution for IR data collection

Figures 3.3.13 and 3.3.14 display the GC and IR results obtained when the dilution was stopped while collecting IR data, rather then collecting data while the sample contained

within the gas cell was still being diluted. As seen in figure 3.3.13 there was an adverse effect on the GC data, particularly for CO and CH_4 responses, so this method was abandoned.

3.3.6 The effect of bleeding SF_6 into the system

The effect of bleeding SF_6 into the system was investigated using the set-up described in section 2.4.6. A close-up of the IR region of interest from spectra obtained during the investigations is displayed in figure 3.3.15 which shows the variation in transmission arising from differing concentrations of SF_6, generally it was found to be a difficult and time consuming process to maintain a constant concentration of SF_6 within the system, so further work was carried out without the use of SF_6. (N.B. From comparisons of PLS models built for different pre-treated data (section 3.2) it was concluded that it was not necessary to add the SF_6)

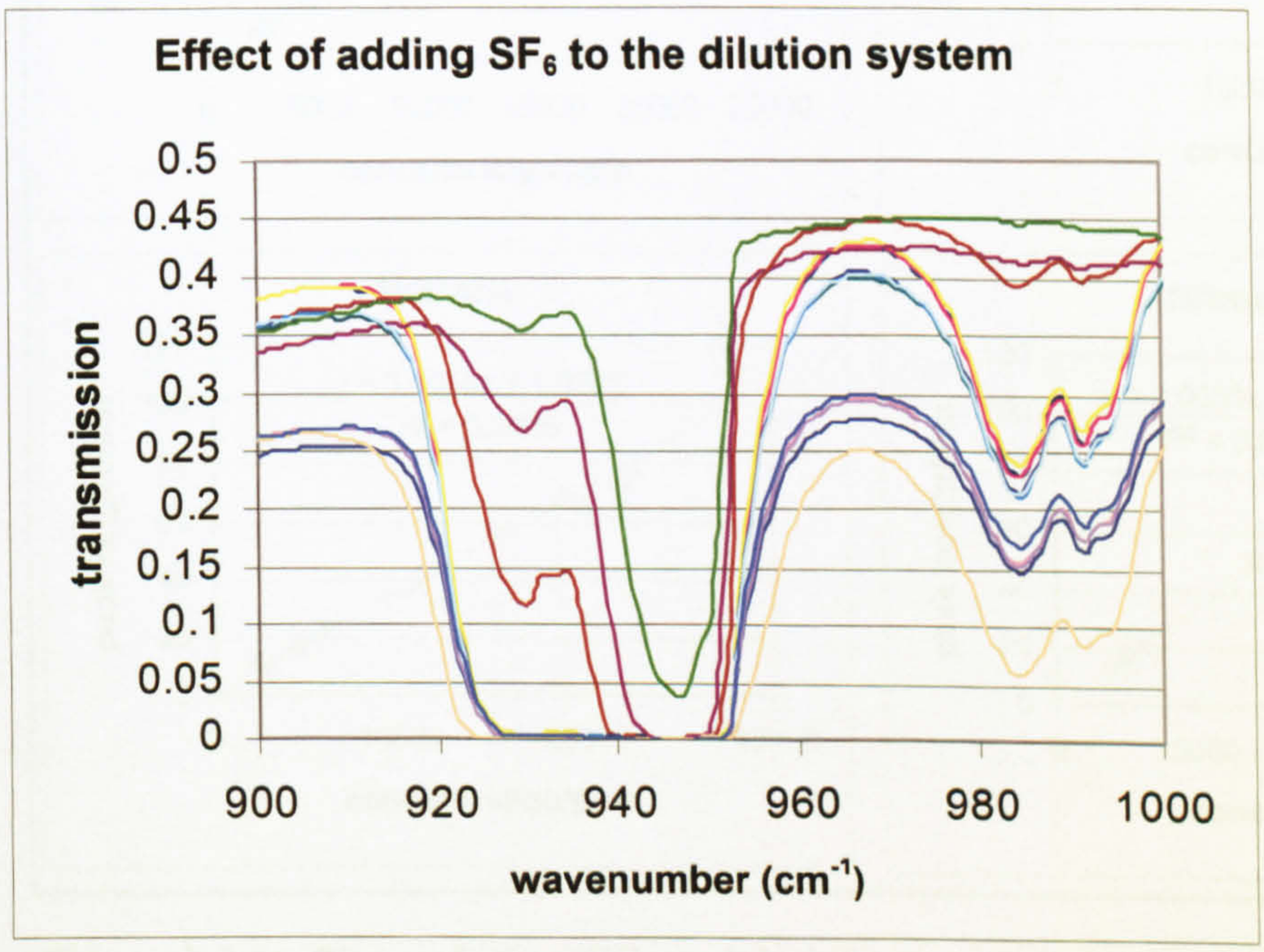


Figure 3.3.15 Close-up of SF_6 peaks from spectra obtained during investigations in to the effect of bleeding SF_6 into the dilution system

3.3.7 Correlation between GC and IR measurements

The GC and IR responses for different sample mixtures were measured utilising the dilution system to fill the GC sampling loop. This was performed to determine the correlation between IR and GC measurements so that in later work developing the IR detector array the GC could be used as reference method to check the composition of gas samples.

The following plots relate the GC responses to the corresponding concentration of gas in the sample (after taking into account the dilution with helium employed to fill the sample loop). The correlation for CH_4, CO_2 and N_2O was greater than 0.98 with the main source of deviation assumed to be due to experimental technique. The correlation for CO was poor, this was mainly due to small amount of air entering the gas cell as samples were prepared and as mentioned previously, the GC response for N_2 is the same as CO.

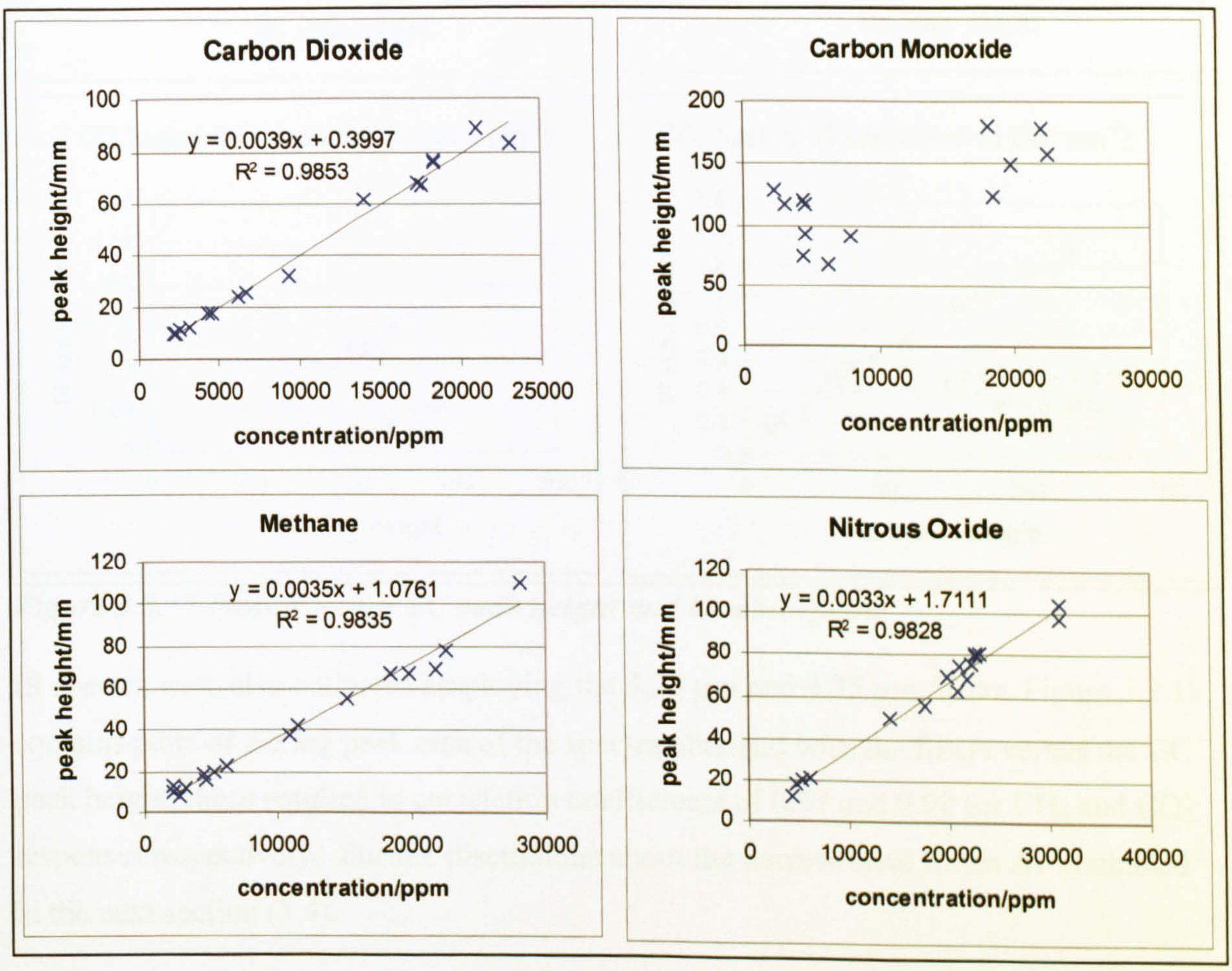


Figure 3.3.16 Plots relating GC peak height to concentration

The GC peak heights were plotted against absorption at wavenumbers known to relate to the gases of interest, i.e. 3088 cm^{-1} for CH_4, 2075 cm^{-1} for CO, 692 cm^{-1} for CO_2 and 2536 cm^{-1} for N_2O. Apart from CO, the correlations obtained were greater than 0.988.

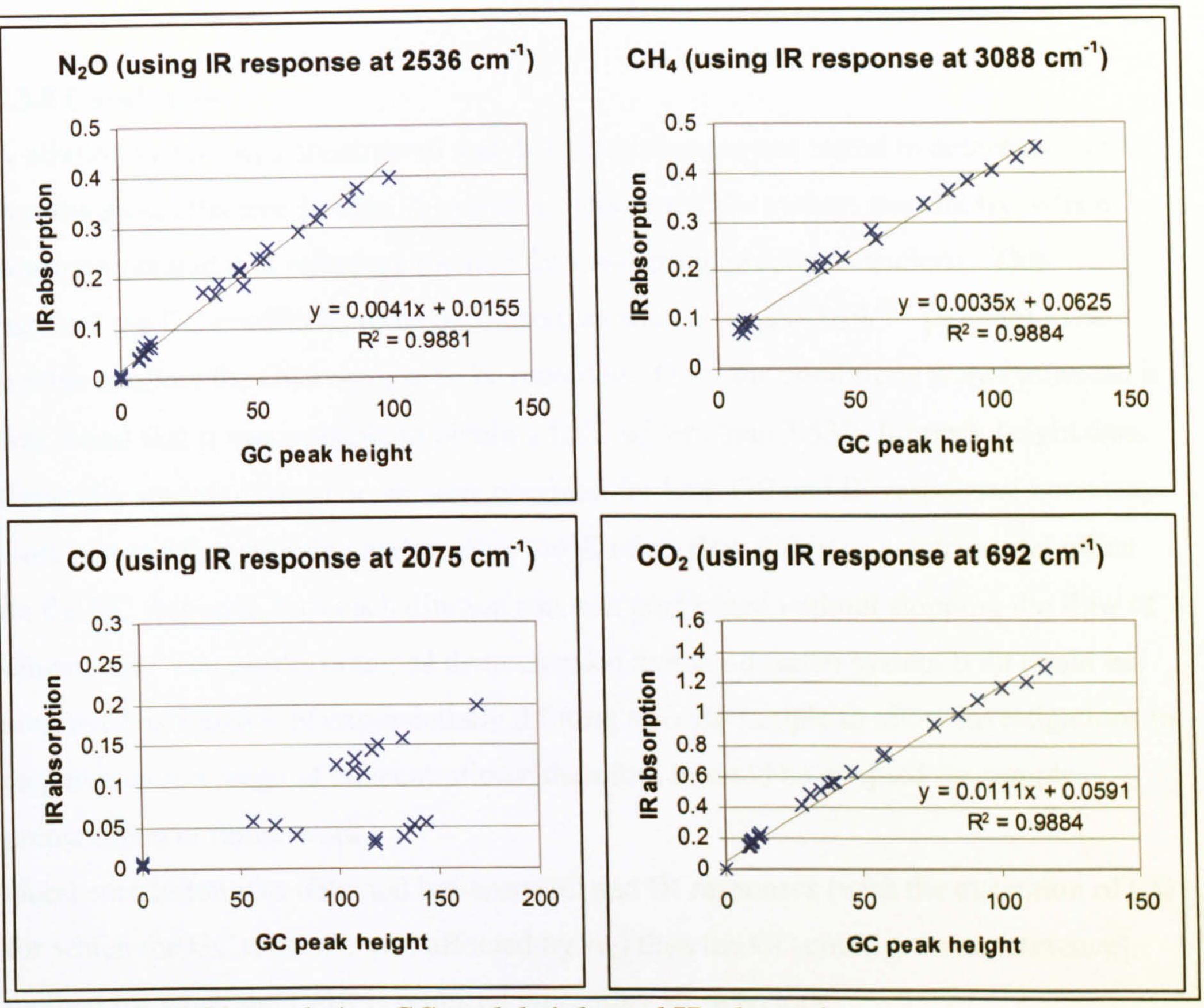


Figure 3.3.17 Plots relating GC peak height and IR absorption

IR spectra were also collected employing the 3.25 μm and 4.35 μm filters. Figure 3.3.18 contains plots of the log peak area of the spectra obtained with the filters versus the GC peak height; these resulted in correlation coefficients of 0.97 and 0.92 for CH_4 and CO_2 responses respectively. Further discussions about the narrow band filters are contained in the next section (3.4).

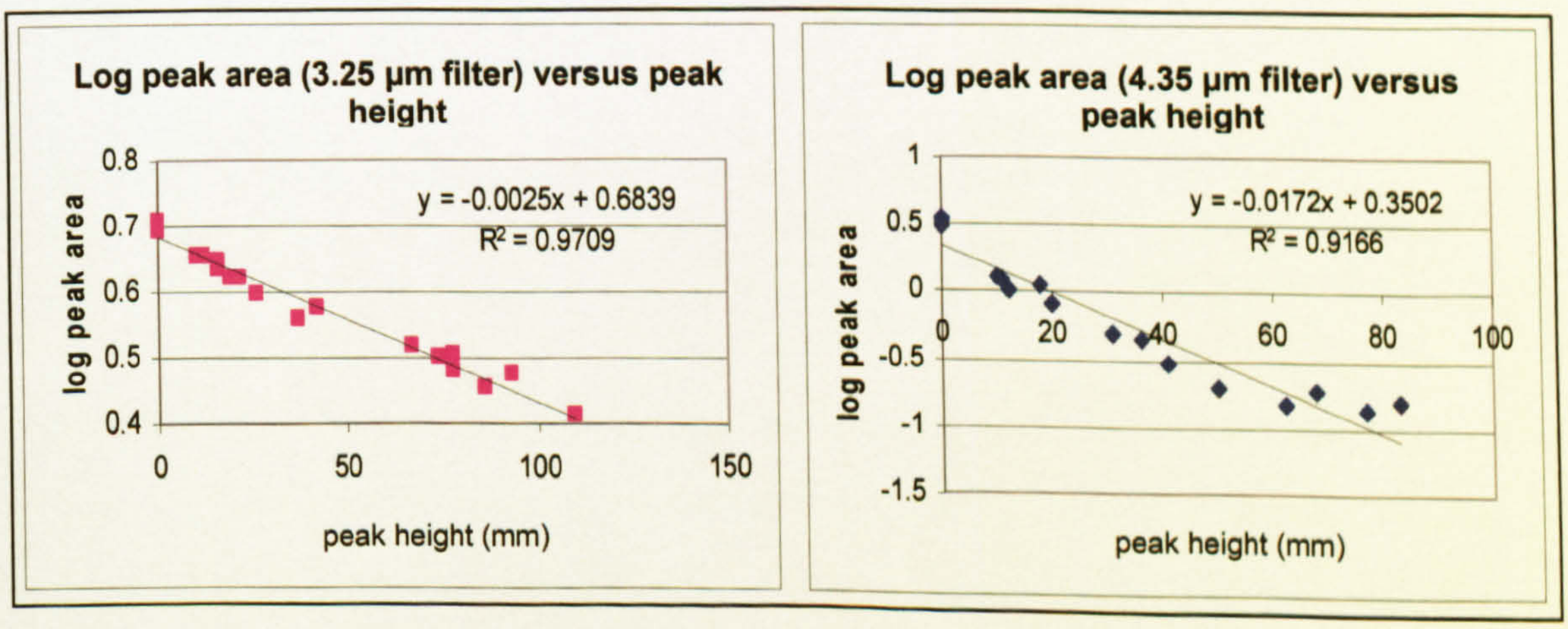


Figure 3.3.18 Plots relation GC peak height to IR response with 3.25μm and 4.35 μm filters

3.3.8 Conclusion

A dilution system was constructed and different layouts were tested to determine what was the most effective design. A major component of the system was the GC which was incorporated as a reference method for measuring gas concentrations. This required the GC conditions to be optimised, as well as a LabVIEW™ program to be written to allow the GC response to be recorded. Once the conditions were optimised it was found that it was possible to obtain a RSD of less than 3.53% for peak height data. Generally smooth decay curves were obtained for both GC and IR responses; however, work showed that stopping and starting the dilution flow did have a detrimental effect on the GC response, thus each dilution run was performed without stopping the flow of diluent gas. The results obtained demonstrated that the dilution system built could be employed as a means of exponentially diluting a single sample to allow investigations in response over a range of concentrations; therefore it could be adopted for sample preparations in future work.

Good correlation was obtained between GC and IR responses (with the exception of CO for which the GC response was affected by N_2) thus the GC could act as a reference method for when the FTIR is replaced with other IR detectors.

3.4 Narrow band filters with FTIR

To simulate the effect of photodetectors measuring at selected wavelengths, narrowband filters that restricted the wavelengths of light, so that only light at the chosen wavelengths were transmitted to the mid IR detector in the FTIR, were employed.

3.4.1 Characterisation of 3.25 μm and 4.35 μm filters

The following plots (3.4.1) show the single beam transmission spectra obtained with the filters for different gas concentrations; it can be seen that these filters successfully blocked light except for the chosen wavelengths. For the 4.35 μm filter, the light transmitted through the filter ranged from 2280 to 2320 cm^{-1}; for the 3.25 μm filter the range was 3048-3100 cm^{-1}.

The response of photodetectors measuring at these wavelengths would be single measurements corresponding to the total amount of light detected, thus it is assumed that such responses would correspond to the area of the spectra obtained when the equivalent filter is employed.

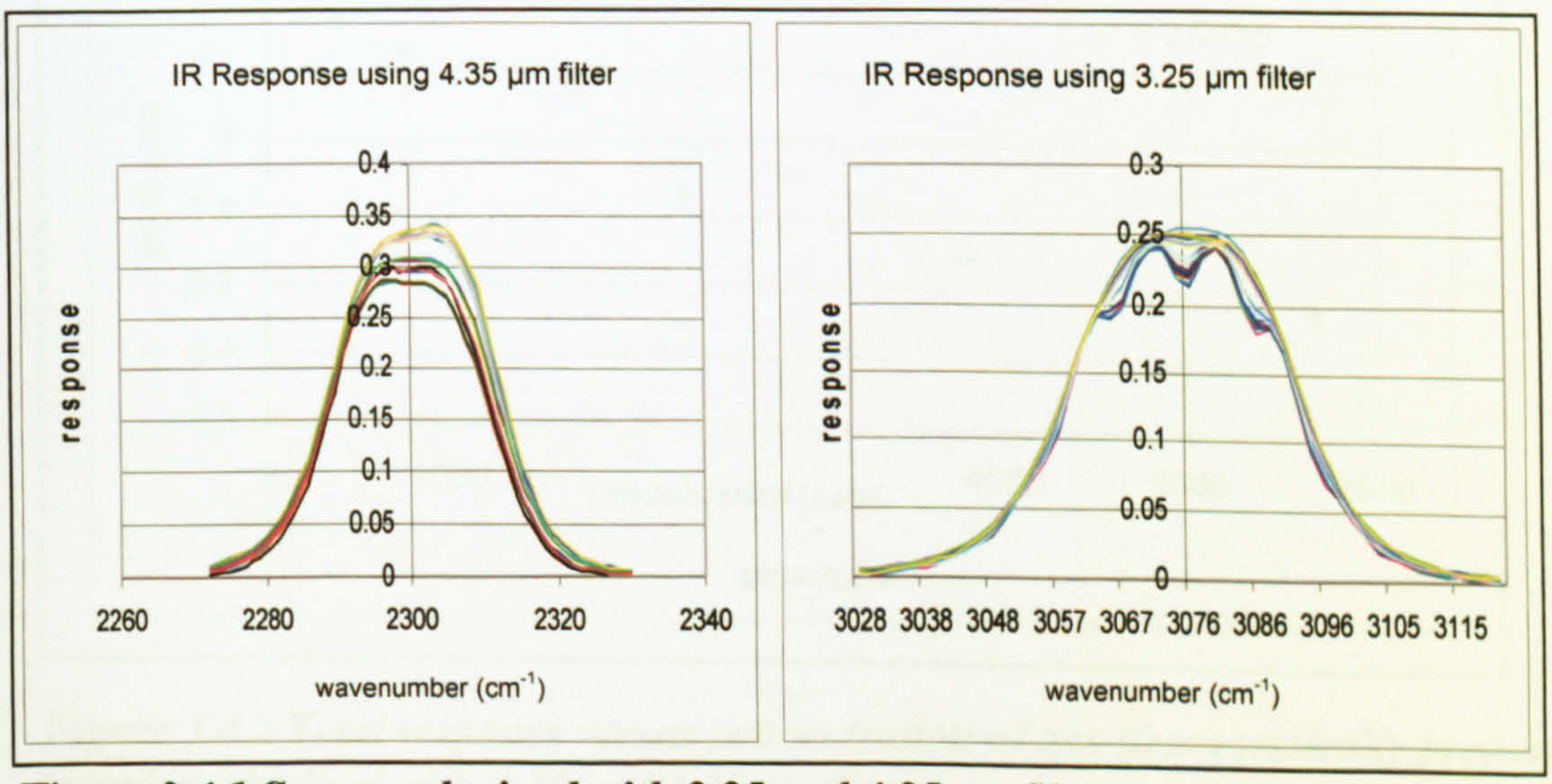


Figure 3.4.1 Spectra obtained with 3.25 and 4.35 μm filters
(NB. due to the range of light transmitted through the filter this can include several gas absorption peaks, as can be seen by the 3.25 μm filter response which shows three CH_4 absorption peaks)

Calibration plots obtained for the two filters measuring statically prepared samples are displayed in figure 3.4.2, these showed R^2 values of 0.95 or greater for total response (area under 'peak') versus concentration.

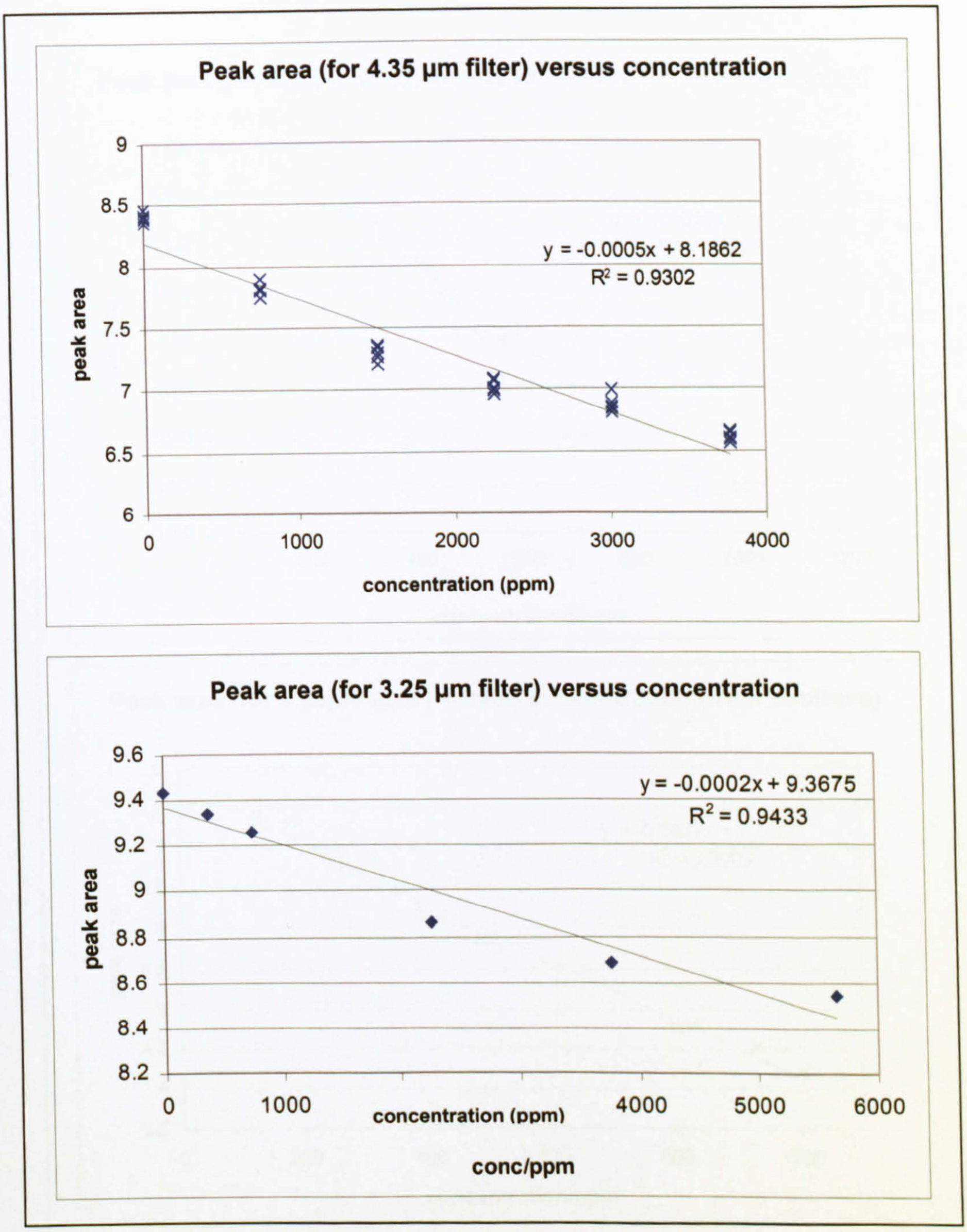


Figure 3.4.2 Total response versus concentration of gas (from statically prepared samples) obtained with 3.25 and 4.35µm filters

Figure 3.4.3 displays the responses obtained using the dilution system; these also gave good correlations between response with filter and concentration. Responses with these two filters were promising; therefore further filters were obtained with band centres at 2.9, 4.73 and 10.62 µm to investigate responses to N_2O and CO, plus the 10.62 µm filter to act as a reference wavelength.

3.4.2 Characterisation of 2.9, 3.25, 4.35, 4.73, and 10.62 μm filters

Figure 3.4.4 compares transmission spectra obtained with and without the narrowband filters, with close-ups of the individual filters shown in figure 3.4.5, the main points to be noted are:

a) The amount of light transmitted through the filters is significantly lower than without filters (see figure 3.4.6),

b) At this resolution the 3.25 μm filter covers several CH_4 absorption peaks

c) There is no SF_6 in the samples scanned; hence, there is no absorption at 10.62 μm

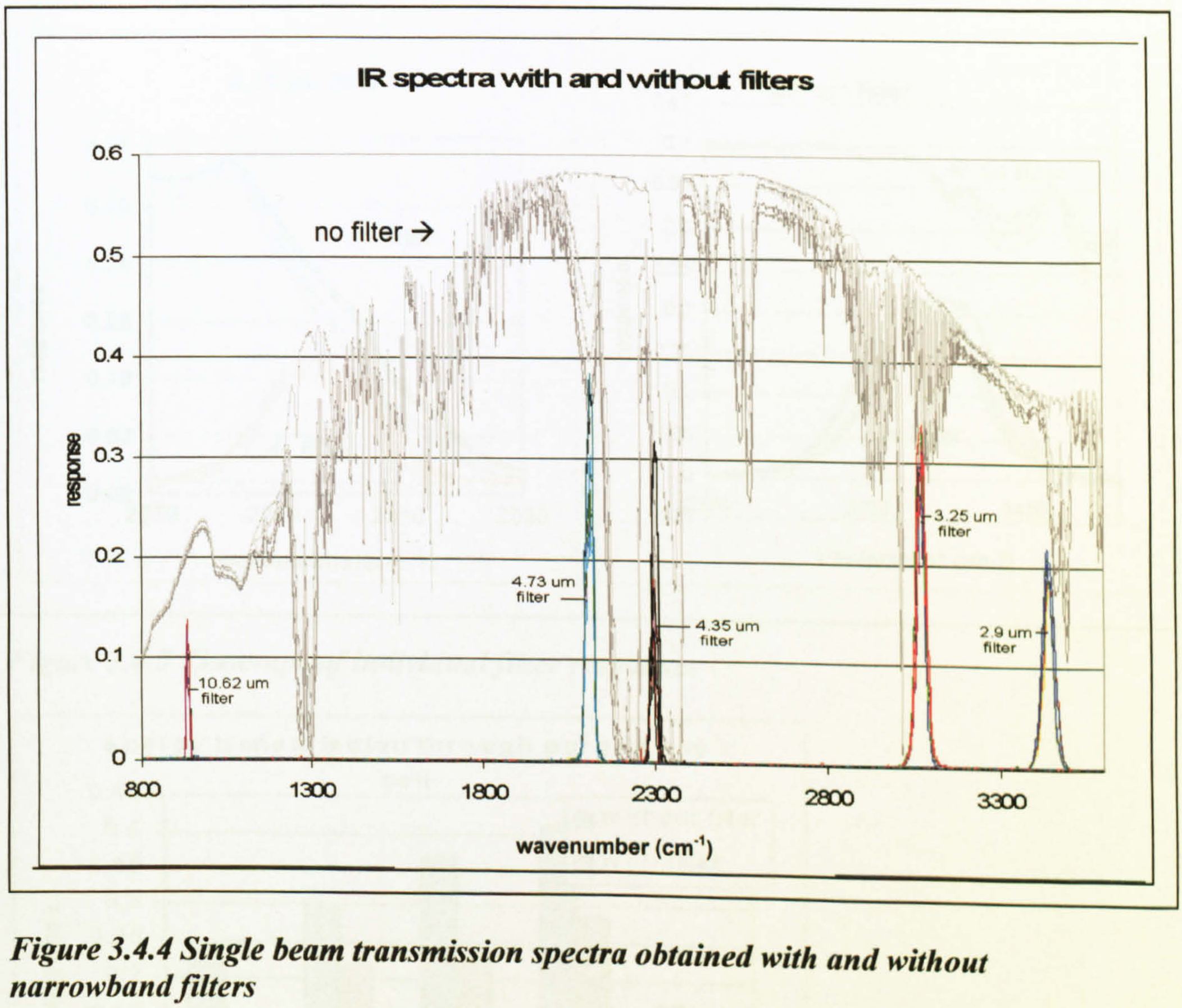


Figure 3.4.4 Single beam transmission spectra obtained with and without narrowband filters

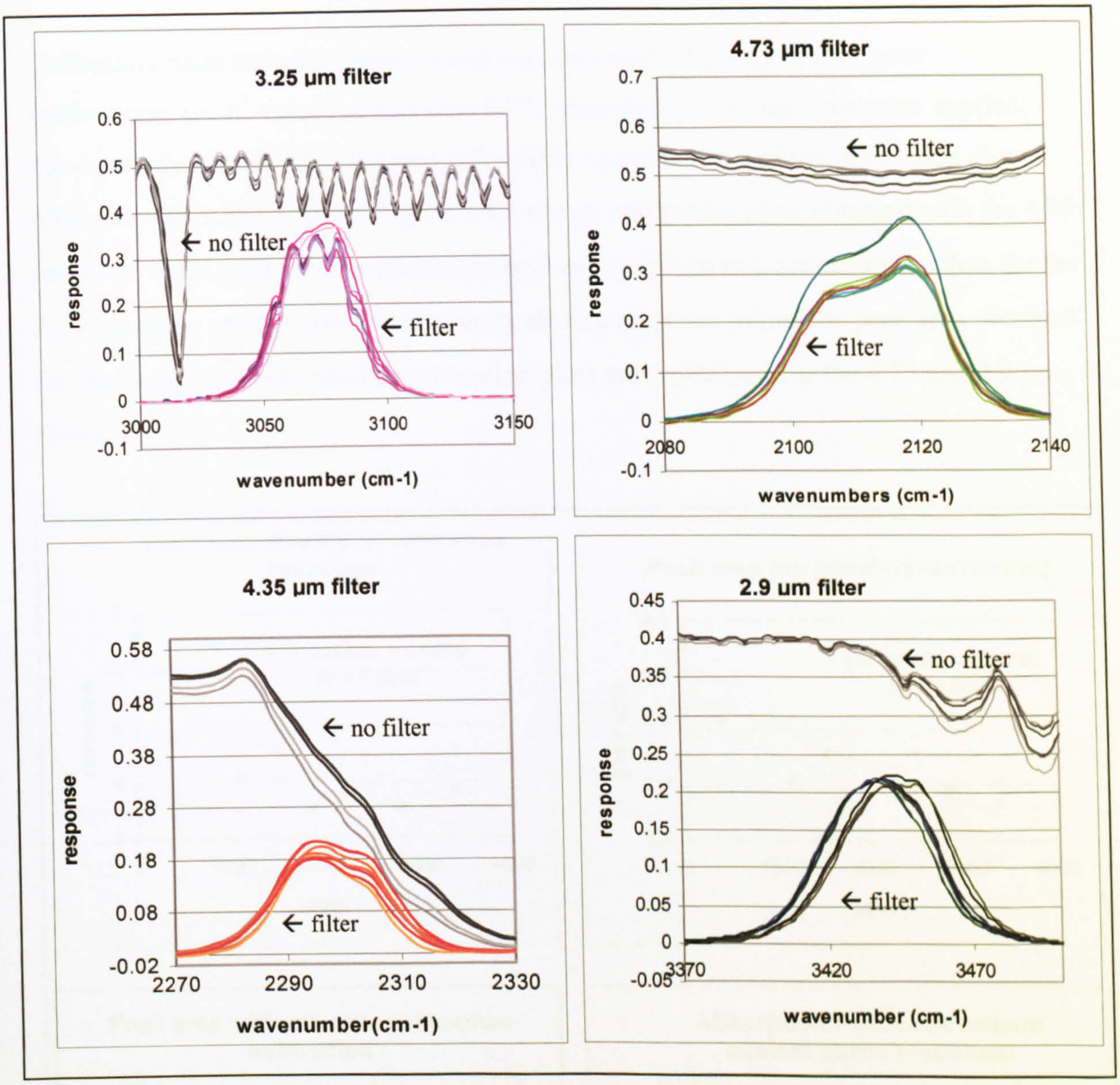


Figure 3.4.5 Close-ups of individual filter responses

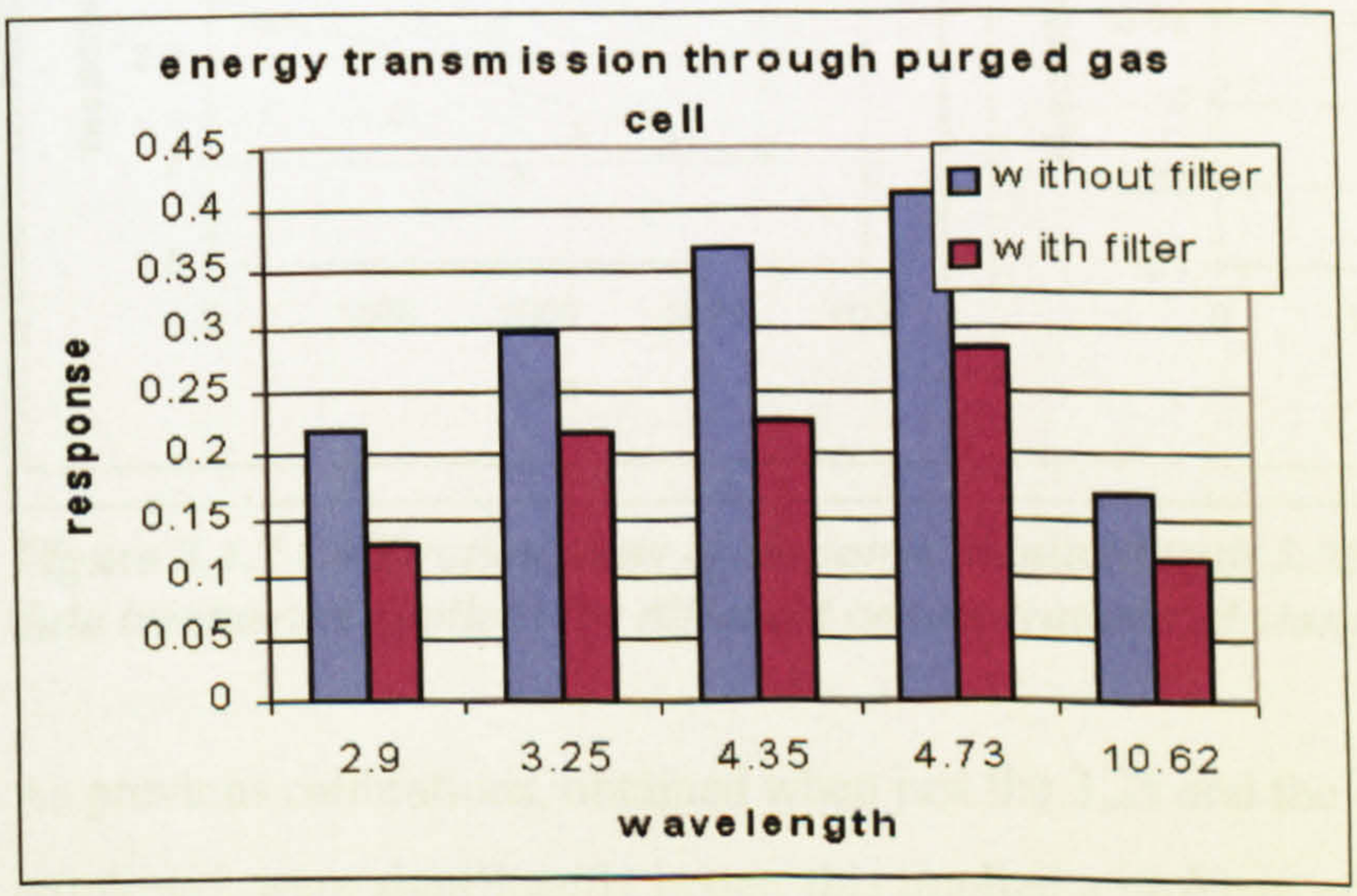


Figure 3.4.6 Comparison of energy transmission between responses with and without filters at chosen wavelengths

Calibration plots between response and concentration of gas gave very poor calibrations, i.e. R^2 values of less than 0.71, regardless of the data treatment applied; this is clearly illustrated in figure 3.4.7, which contains some of the calibration plots obtained with 3.25 μm filter. Figure 3.4.8 shows calibration plots obtained with the 4.35 μm filter; these gave better correlations between response and concentration than for the 3.25 μm filter, particularly when using peak area response minus the peak area obtained for the purged gas cell. Similar calibration plots were obtained for the 4.73 and 2.9 μm filters.

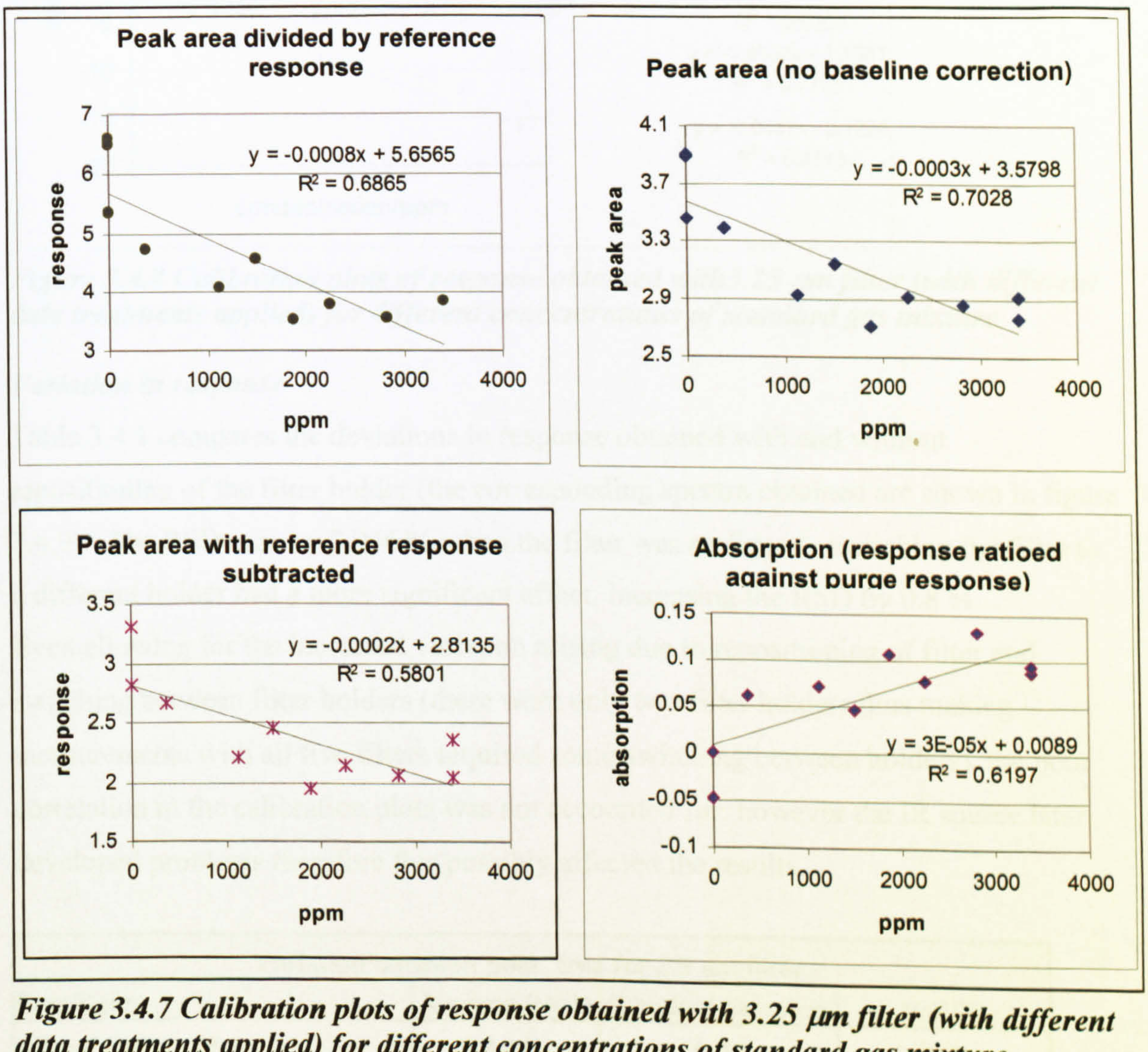


Figure 3.4.7 Calibration plots of response obtained with 3.25 μm filter (with different data treatments applied) for different concentrations of standard gas mixture

As previous calibrations, obtained when just the 3.25 and the 4.35 μm filter were employed, were significantly better, this implied a problem either with the FTIR or

more variation in response arising due to switching between the different filters (possibly resulting in slight changes in alignment causing changes in response).

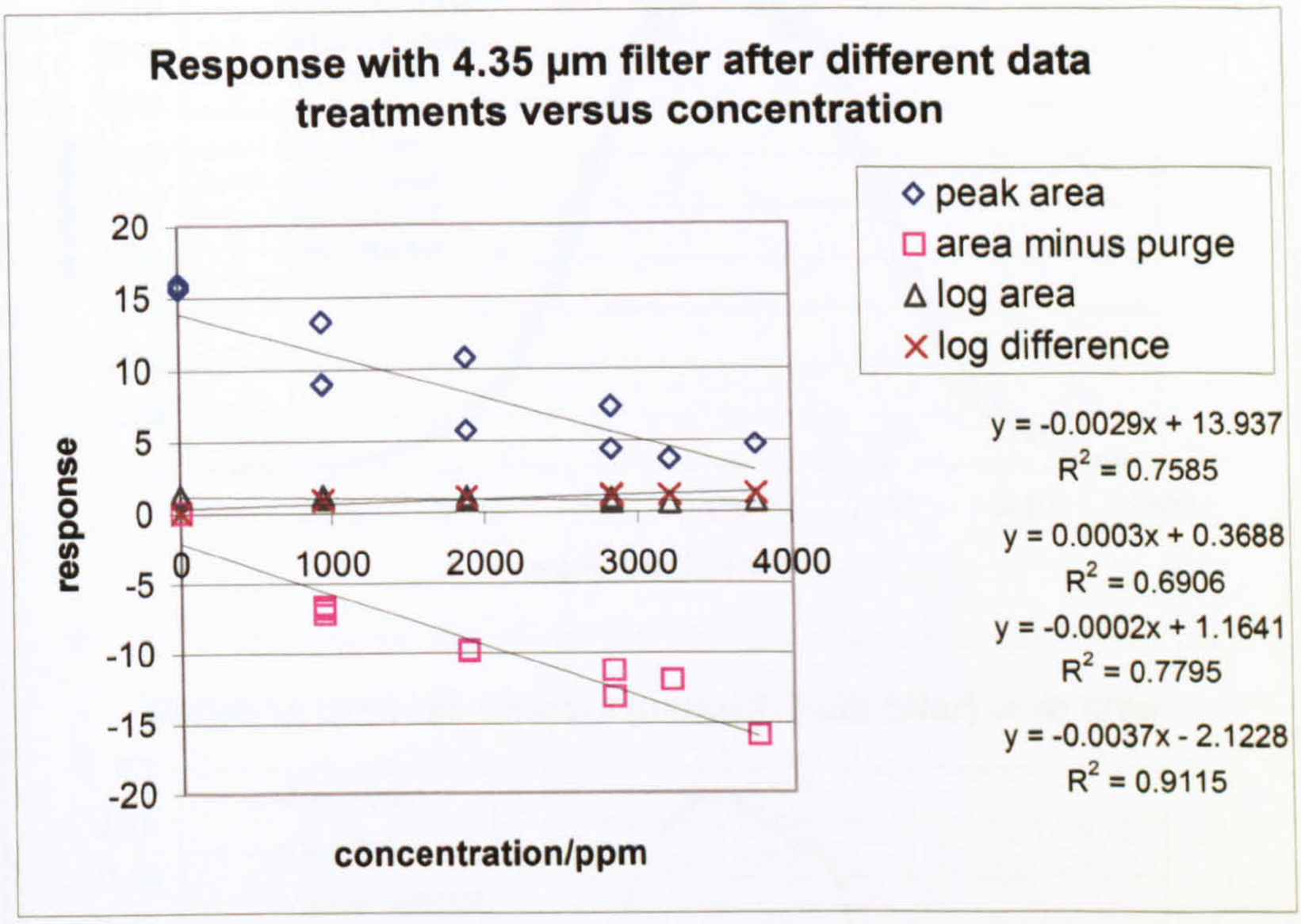


Figure 3.4.8 Calibration plots of response obtained with3.25 µm filter (with different data treatments applied) for different concentrations of standard gas mixture

Variation in response

Table 3.4.1 compares the deviations in response obtained with and without repositioning of the filter holder (the corresponding spectra obtained are shown in figure 3.4.9). The RSD rose by 0.046 % when the filter was realigned; switching the filter to a different holder had a more significant effect, increasing the RSD by 0.8 %.
Even allowing for the increased variation arising due to repositioning of filter and switching between filter holders (there were only two filter holders thus making measurements with all five filters required some switching between holders), the poor correlation in the calibration plots was not accounted for; however the IR source later developed problems therefore this possibly affected the results.

Variation between peak area for 2.9 µm filter			
Conditions	Average (N=3)	Standard deviation	% RSD
without moving filter	8.94	0.05	0.52
filter moved and realigned	9.01	0.05	0.57
switching filter to different holder	8.98	0.12	1.33
overall	8.98	0.07	0.83

Table 3.4.1 Deviations in response for 2.9 µm filter obtained with and without the repositioning of the filter

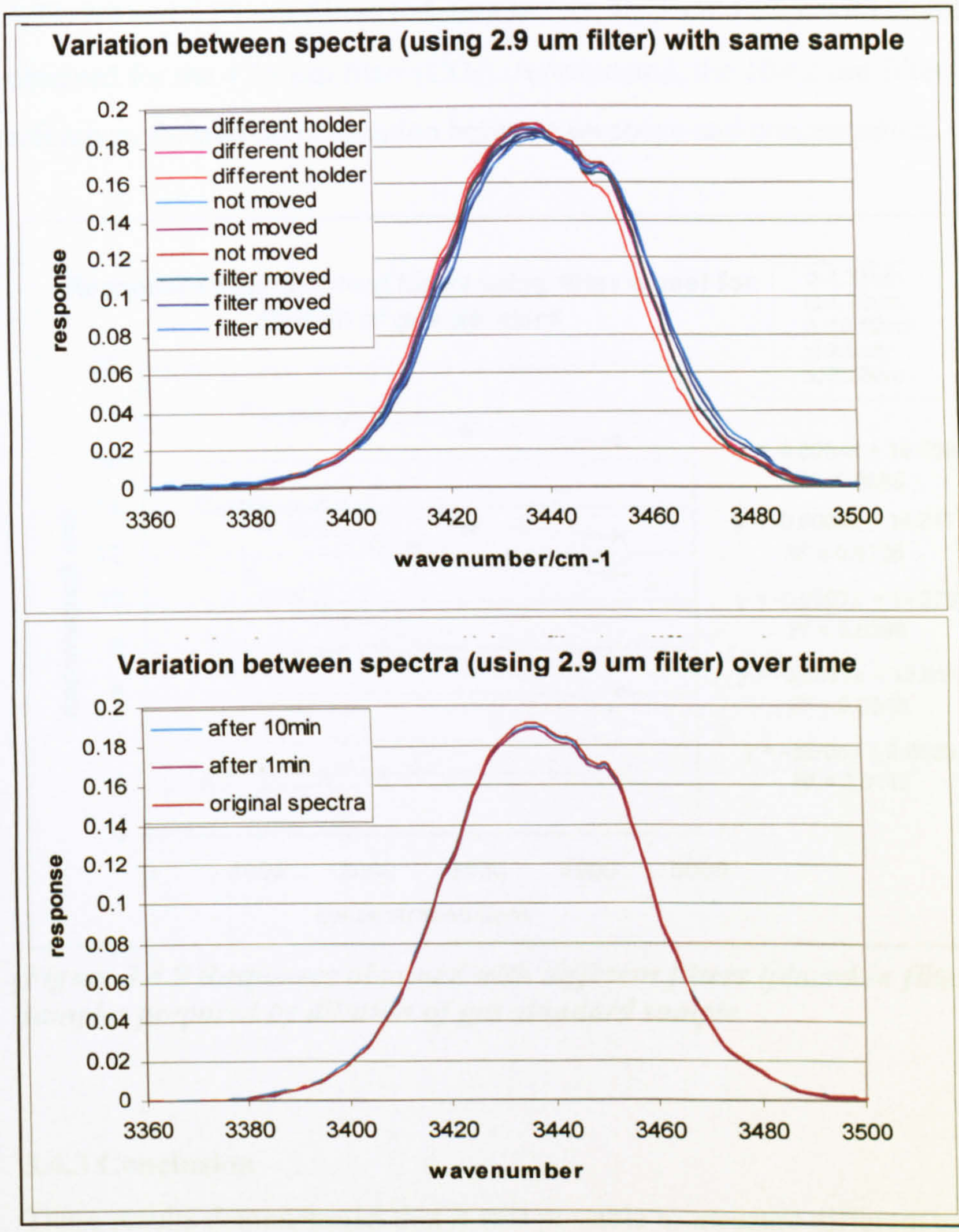


Figure 3.4.8 Transmission spectra obtained for 2.9 μm filter obtained with and without the repositioning of the filter

Automated Filter Wheel

To reduce the variations in response arising due to repositioning of the filters and to make it easier to switch between different filters, an automated filter wheel was employed. The wheel held up to six filters thus allowing the five filters to be placed in five of the positions and the remaining space left empty for full spectrum scans; the wheel could then be turned with either a manual switch button or via computer control. Figure 3.4.9 gives a plot of the responses obtained using the filter wheel to switch between different filters, for samples prepared by dilution of the 1 % gas standard sample. Good correlation between concentration and response were obtained for the

4.73, 2.9 and 3.25 μm filters (relating to CO, N_2O and CH_4), with slightly worse results obtained for the 4.35 μm filter (CO_2). As expected, the 10.62 μm filter, acting as reference, showed no correlation between response and concentration.

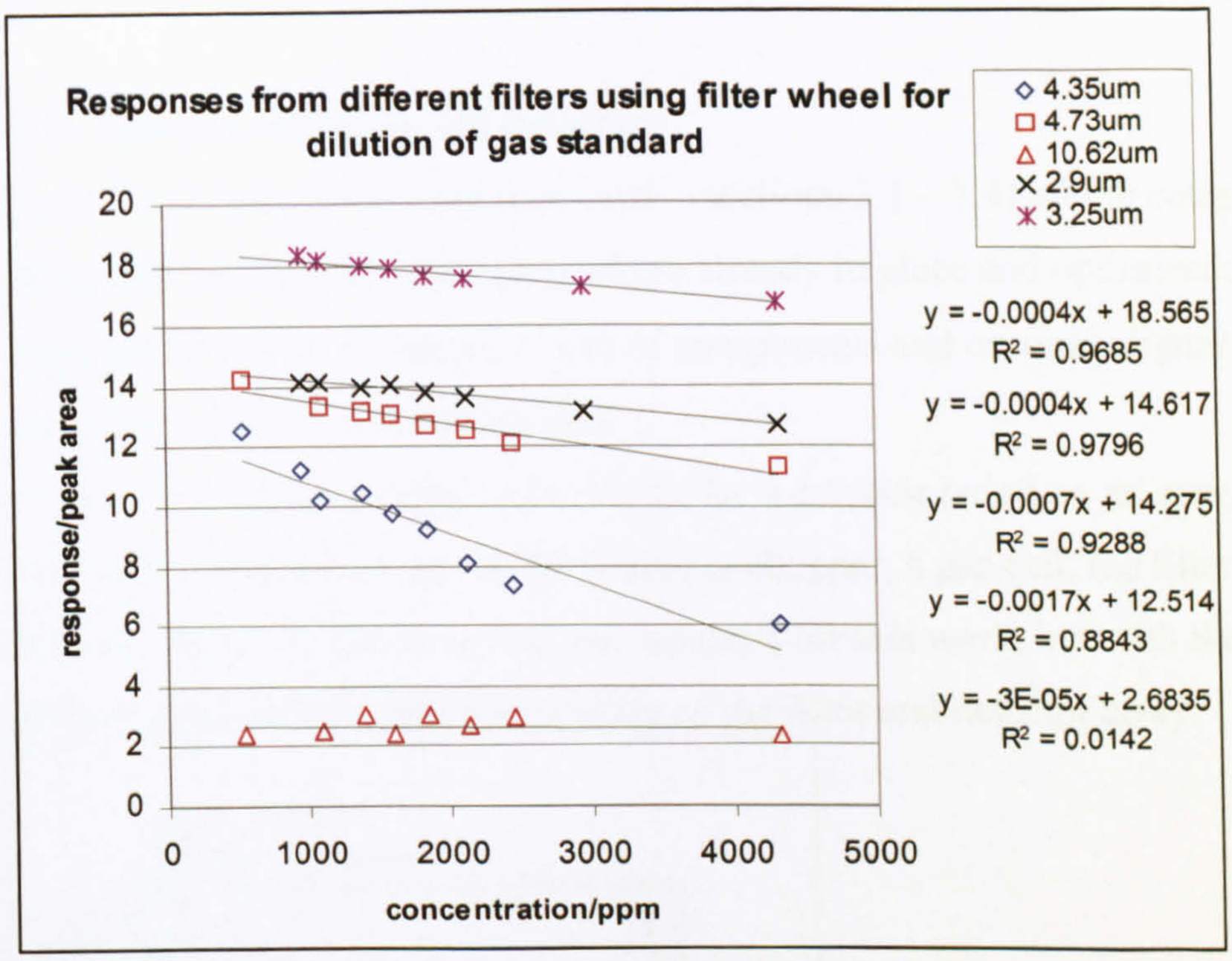


Figure 3.4.9 Responses obtained with different filters (placed in filter wheel) for samples prepared by dilution of gas standard sample

3.4.3 Conclusion

These results demonstrated that it was possible to measure different concentrations of CO, CO_2, CH_4 and N_2O using filters at the selected wavelengths of 3.73, 4.35, 3.25 and 2.9 μm, thus indicating it would be possible to build a detector array using detectors measuring at these wavelengths. These results were collected using a FTIR instrument, where all the optical components and signal processing were already optimised; the next stage of the work was to develop a simple prototype instrument including investigations into the design of the optical layout and means of signal processing and data recording.

3.5 Filter work independent of FTIR

This stage of the work focused on developing an infrared measuring system independent of the FTIR, thus demonstrating the feasibility of building a gas measuring system based on an array of detectors at wavelengths selected from the VS-MLR modelling.

3.5.1 General set-up for IR detectors

In the FTIR instrument used previously (sections 3.1 – 3.4) all the components, alignments and signal processing where already in place and optimised; for an independent set-up the arrangement of components and optimal alignment and conditions all had to be determined.

Figure 3.5.1[45] is an example of a design for a detector based on an array of IR detectors. The basic components are the IR source, a chopper, a gas cell, the filter array and detector array. A similar layout was adopted for this work, but with the automated filter and a general mid IR detector in place of the filter and detector array.

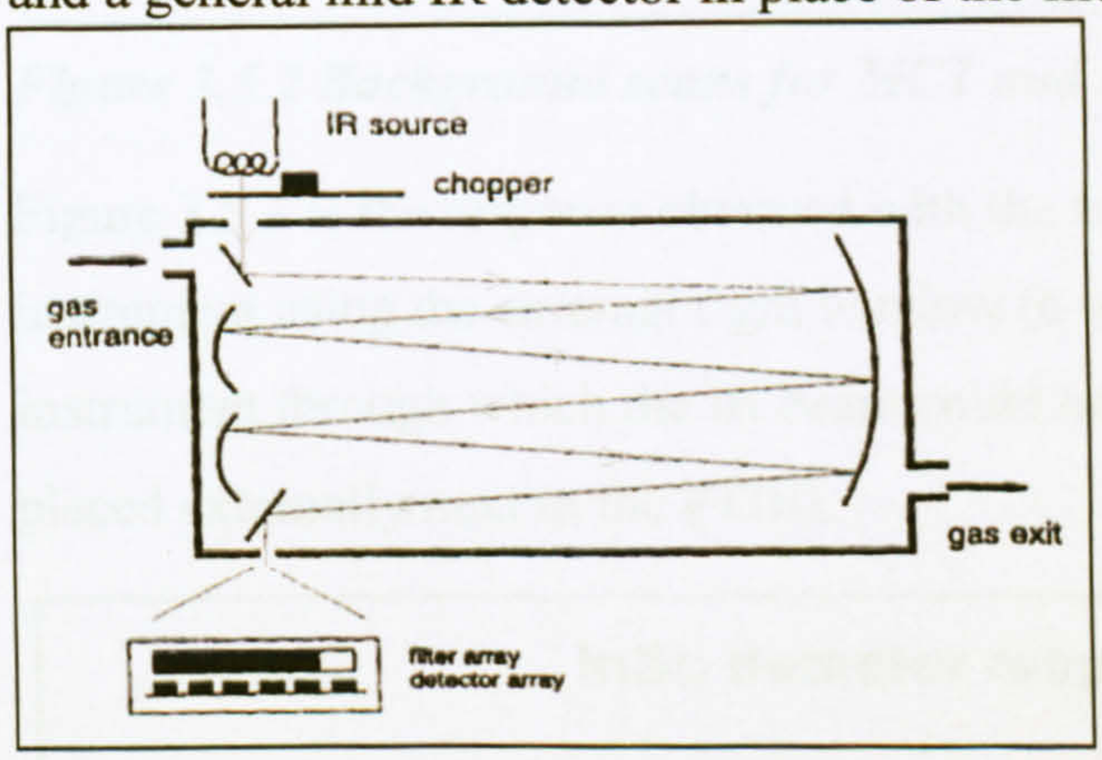


Figure 3.5.1 Set-up for IR detector[45]

3.5.2 Different Detectors

The available detectors included the MCT detector from BioRad and the InSb from Hamamatsu Photonics K.K. The internal detector of the FTIR instrument employed in all the previous work was a peltier-cooled DTGS detector. These detectors were tried with the different set-ups to determine which one worked best. Ultimately QWIP detectors will replace these detectors, thus the actual performance of the QWIPs may be different to the results obtained for this work, but the principles will be the same.

The responses obtained from the two BioRad detectors for background scans (no gas cell present) using the FTIR are shown in figure 3.5.2. The main differences between the responses are the increased response of the MCT detector for wavenumbers between

900 and 3500 cm^{-1} and different cut-off values (cut-off for the DTGS detector is 500 cm^{-1} whereas the cut-off for the MCT is around 800 cm^{-1}).

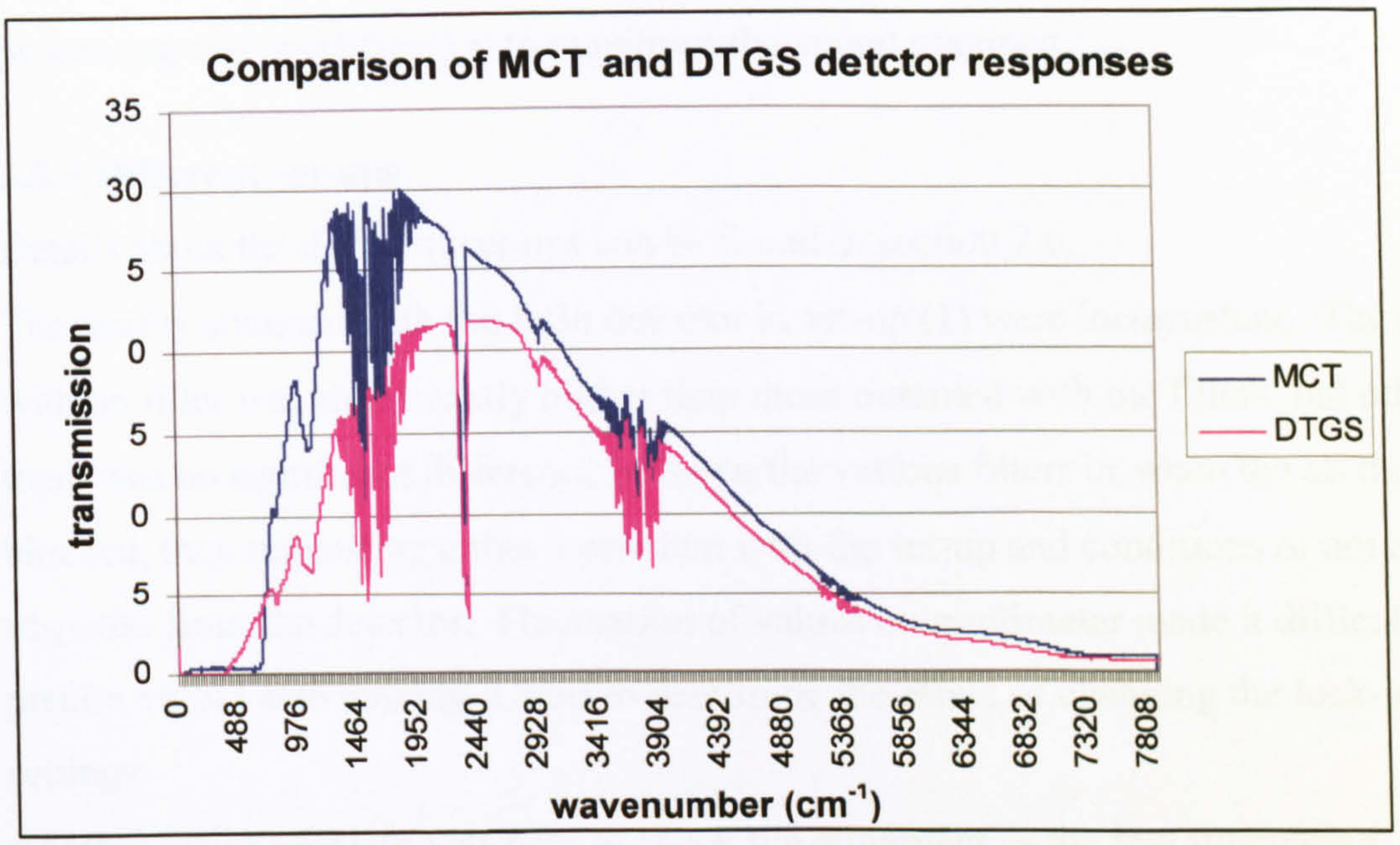


Figure 3.5.2 Background scans for MCT and DTGS detector using FTIR

Figure 3.5.3 is the response obtained with the InSb detector connected to Excaliber instrument using the external right window (a window on the right side of the instrument through which the IR beam could be directed so that the detector could be placed externally next to the FTIR).

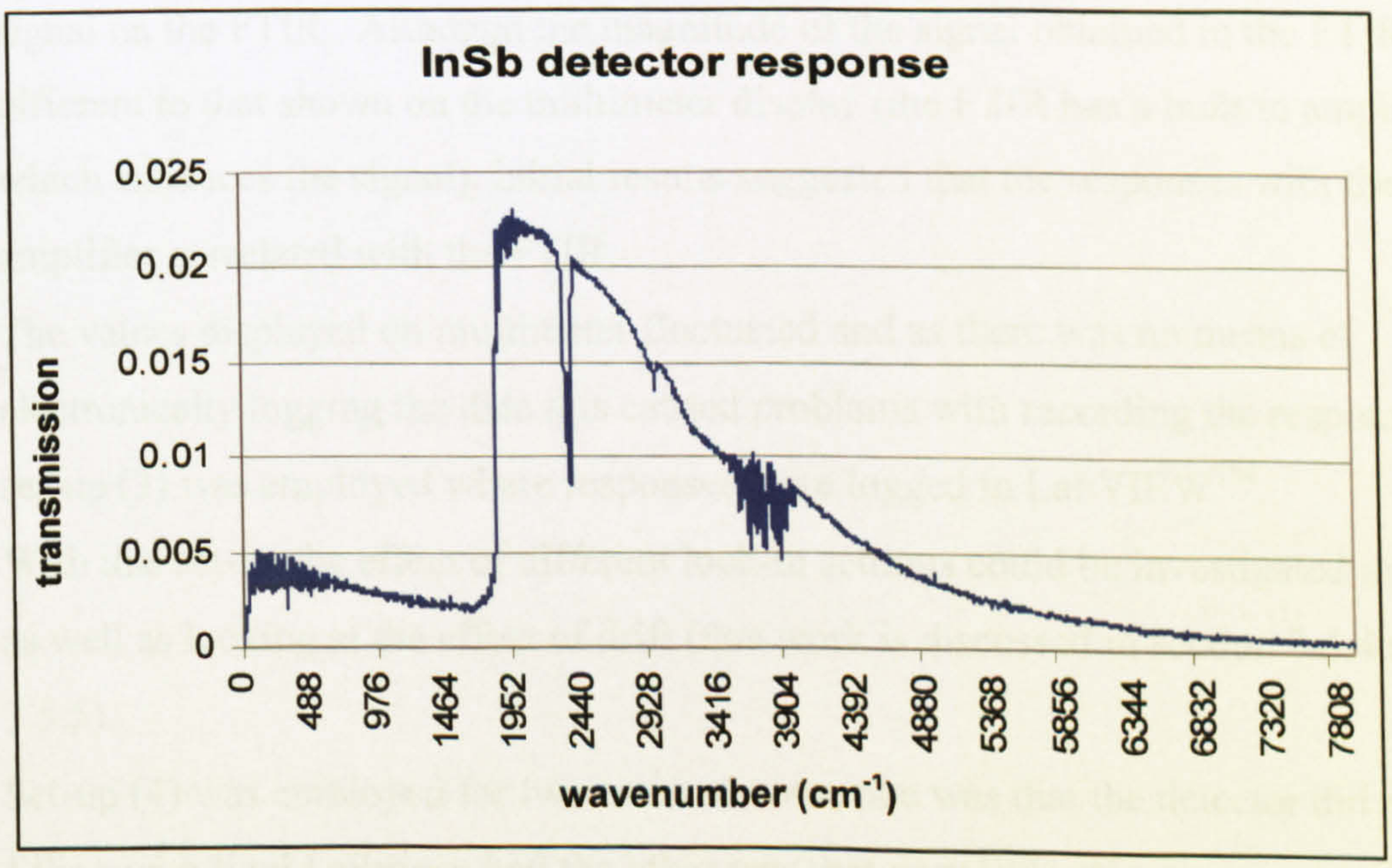


Figure 3.5.3 Background Scan obtained for InSb detector using FTIR

The InSb detector had a much higher cut-off point at 1800 cm^{-1}. The response obtained was low compared to the BioRad detectors this is assumed to be due partially to incompatibility with the BioRad hardware. The detectors have additional signal processing and amplification to maximise the signal obtained.

3.5.3 Different set-ups

Details about the different set-ups can be found in section 2.6.

The results obtained with the InSb detector in set-up (1) were inconsistant. The signal with no filter was significantly higher than those obtained with the filters, but otherwise there was no significant difference between the various filters or when the all the light blocked, thus suggesting either a problem with the set-up and conditions or not enough response from the detector. Fluctuation of values on multimeter made it difficult record precise values also making it hard to determine the effect of changing the lock-in settings.

Another major problem was how to check the alignment as the low fluctuating responses made it difficult to determine what the optimal response was. To counteract this problem, set-up (2) was adopted. The detector was switched from the InSb to the MCT as this gave a higher response with the FTIR. Set-up (2) allowed the signal obtained by the FTIR to be used as a means to optimise the alignment; this was achieved by adjusting the position of the components to maximise the interferrogram signal on the FTIR. Although the magnitude of the signal obtained in the FTIR was different to that shown on the multimeter display (the FTIR has a built in amplifier which enhances the signal), initial results suggested that the responses with the lock-in amplifier correlated with the FTIR.

The values displayed on multimeter fluctuated and as there was no means of electronically logging the data this caused problems with recording the response; hence set-up (3) was employed where responses were logged in LabVIEW™.

With this set-up the effect of different lock-in settings could be investigated more fully, as well as looking at the effect of drift (this work is discussed in section 3.5.4 and 3.5.5).

Set-up (4) was employed for two main reasons, one was that the detector did not require filling with liquid nitrogen and the other was that very little alignment was required. The source and detector were already optimised in the instrument and the gas cell could just be slotted into the sample holder with the filter wheel placed next to it, leaving just

the chopper to align. The disadvantage was the limited space in the sample chamber. Different amounts of gases were injected into the cell, but the responses with the different filters bore little relation to the gases injected.

One possible problem with the previous set-up could have been the effect of the interferometer, in the FTIR this resulted in the interferogram, which is Fourier transformed, whereas with the lock-in amplifier no such treatment was applied. To remove this effect (and make the set-up totally independent of the FTIR) the source was removed from the FTIR. Injecting different amounts of gases into the small gas cell did not have a significant effect on the responses, therefore it was replaced with the long path gas cell which had the disadvantage of a drop in the amount of light reaching the detector; however, the increased path length increases the proportion of absorption due to the gases injected. The following plot (figure 3.5.1) shows responses obtained with the long path gas cell. The expected results would be a change in response for the 2.9 μm filter upon the addition of N_2O, similarly changes in response for 4.73 μm with CO and 4.35 μm with CO_2 would also be expected.

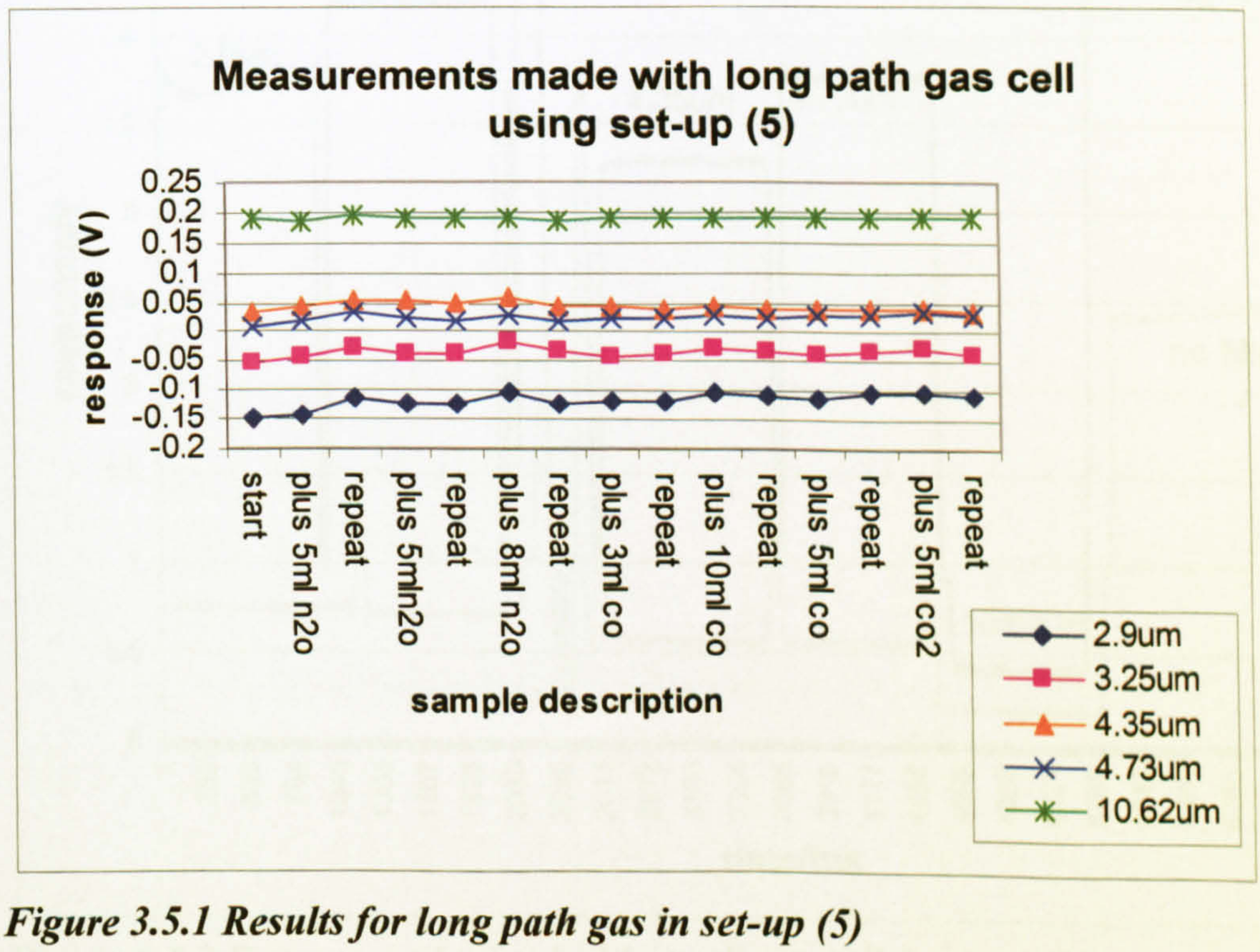


Figure 3.5.1 Results for long path gas in set-up (5)

A lens was incorporated into the set-up to focus the light onto the detector.

The small gas cell was used in the set-up because this was easier to align (with the long path gas cell, light is reflected back and forth between mirrors resulting in a significant loss in the amount of light finally reaching the detector, and small deviations in the

positioning of the cell are more significant). One disadvantage with this was that there was only a relatively small volume of gas (100 ml) with only a short path length (10 cm) hence this set-up is less effective for low concentrations of gases.

Figure 3.5.2 shows typical responses with the small gas cell, obtained from set-up (5) with the lens incorporated. 'Light blocked' is the response obtained when light from the source is blocked. (NB. Blocking the light entirely was difficult, as the set-up was not fully enclosed resulting in stray light reaching the detector, thus there was still a high response obtained with no filter present when the light was blocked)

The flat horizontal portions of the plot relate to the filters, with the fluctuating peaks in between corresponding to the wheel rotating to the next position.

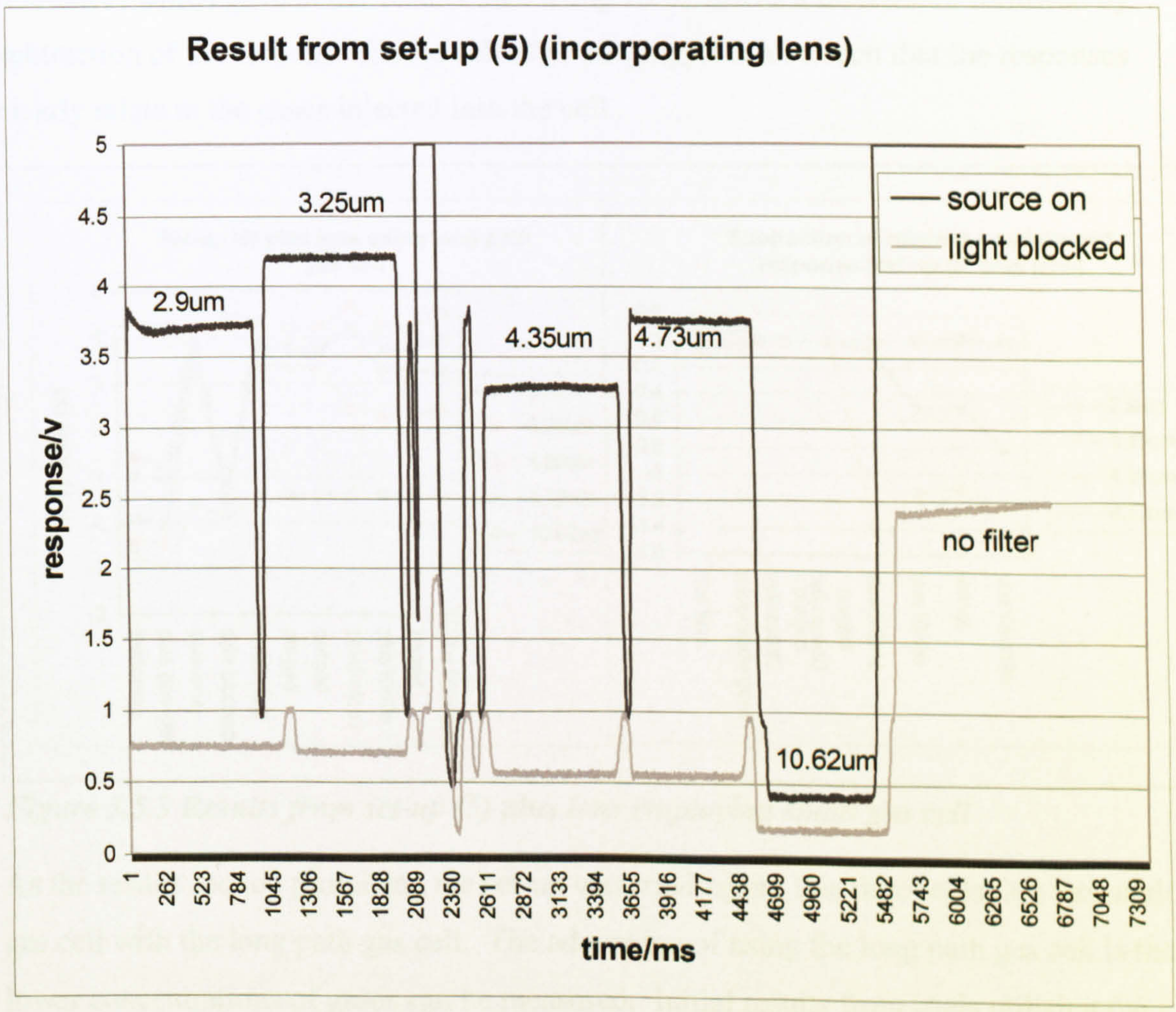


Figure 3.5.2 Responses obtained with small gas cell in set-up (5) incorporating lens

Figure 3.5.3 displays plots of the average response obtained for each filter under different conditions. The improvement in response after the addition of a lens into the

set-up is clearly seen by the higher response with the lens compared to the response obtained without the lens.

Upon the addition of CO_2, the response with the 4.35 μm filter dropped which is as expected. There was a slight increase in response with the other filters, but as there was no reason why this was the case it is assumed that this was most likely the result of drift).

The addition of N_2O resulted in the expected drop in response for 2.9 μm filter. In this case, the responses with the other filters also dropped; again this is attributed to drift.

By subtracting the reference response (in this instance the reference employed was the response for the 3.25 μm filter (as no CH_4 was being used this could be used as a reference) which gave better results than using 10.62 μm as a reference), followed by subtraction of the response for the cell after purging, it can be seen that the responses clearly relate to the gases injected into the cell.

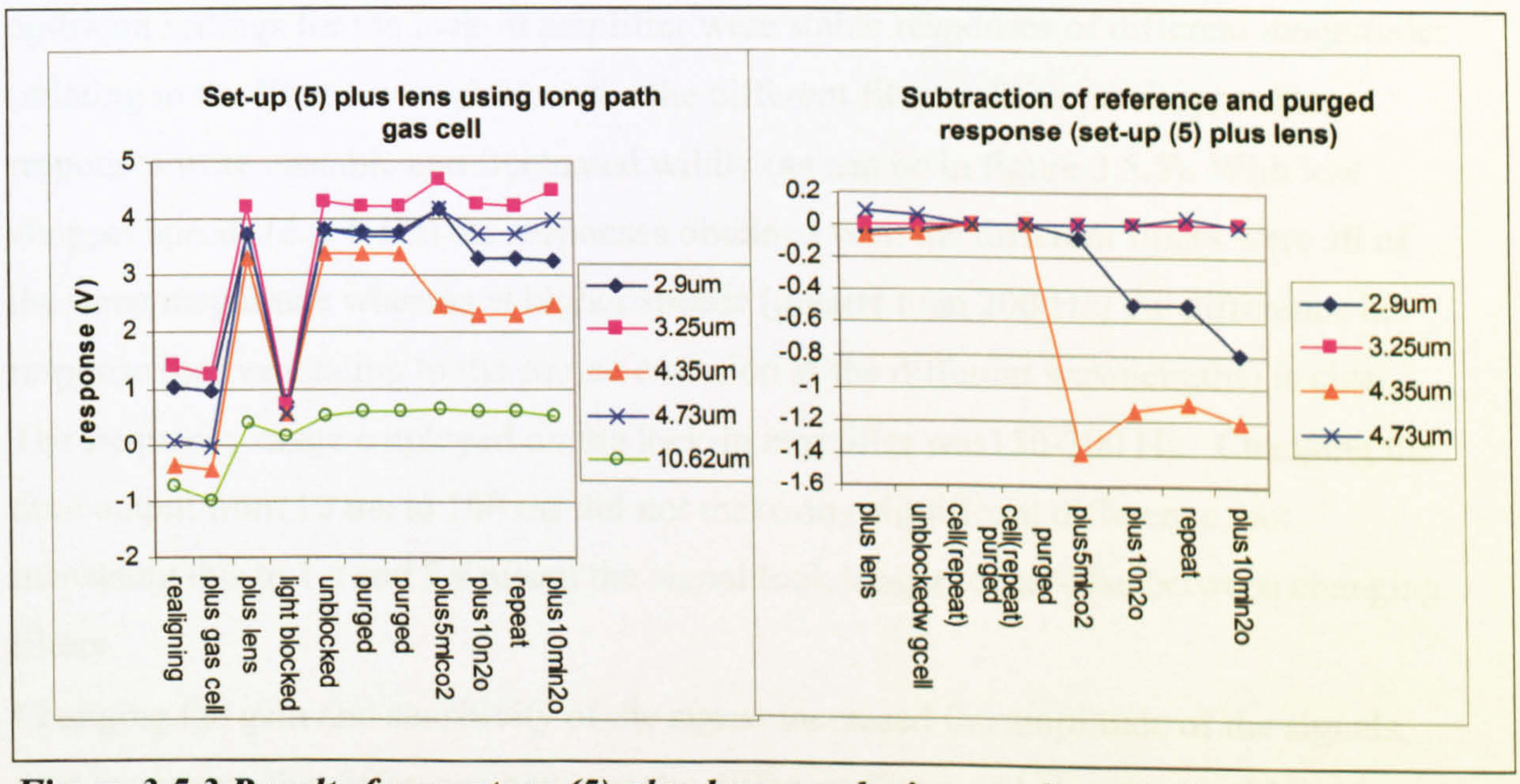


Figure 3.5.3 Results from set-up (5) plus lens employing small gas cell

As the results looked promising the set-up was tried again, this time replacing the small gas cell with the long path gas cell. The advantage of using the long path gas cell is that lower concentrations of gases can be measured. Initial results from trials utilising the long path gas cell are shown in figure 3.5.4.

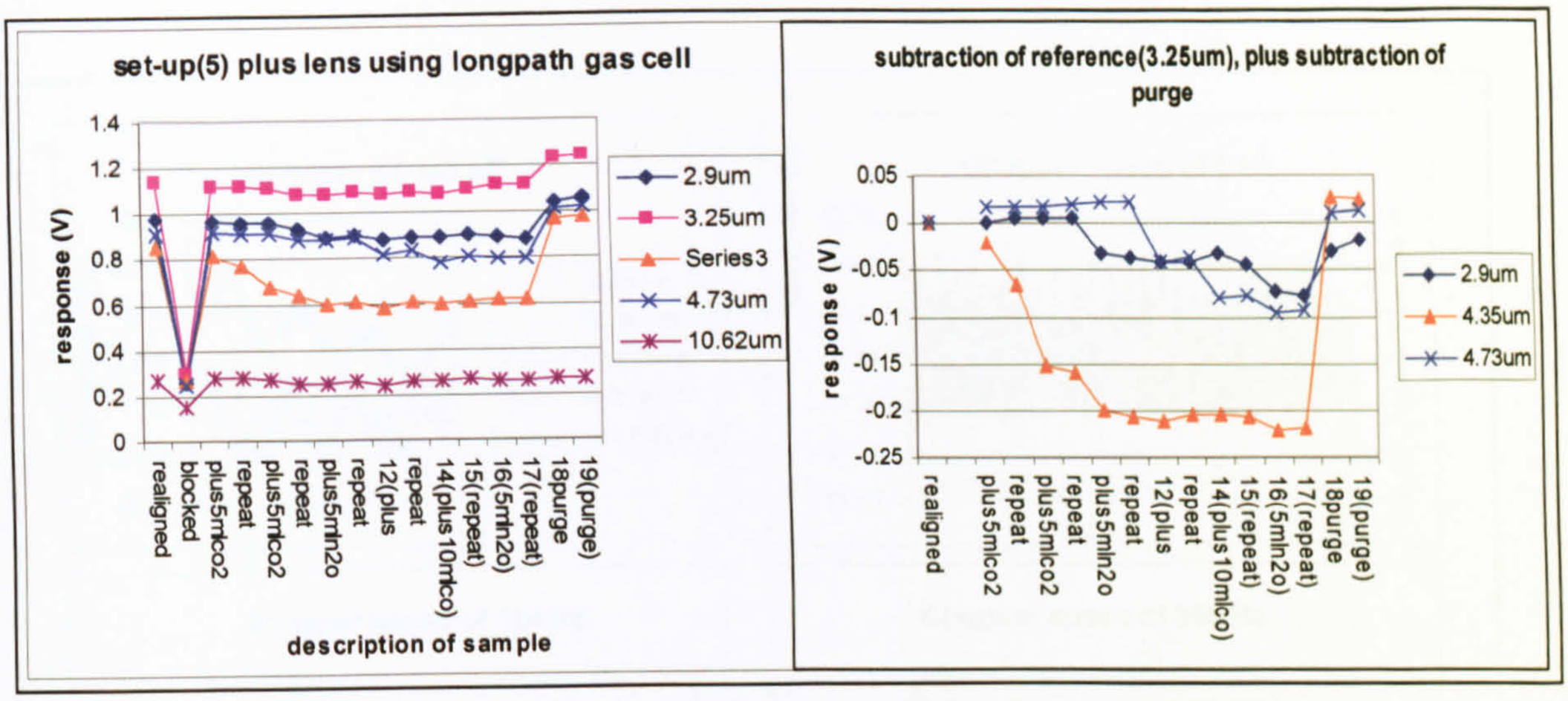


Figure 3.5.4 Results from set-up (5) plus lens using long path gas cell

3.5.4 Optimisation of the lock-in amplifier settings

As mentioned in the experimental section, the main criteria for determining the optimum settings for the lock-in amplifier were stable responses of different magnitudes (relating to the IR source emission) for the different filters. With no chopper the responses were unstable and fluctuated wildly (as can be in figure 3.5.5). With low chopper speeds (e.g. 5 Hz) the responses obtained with the different filters were all of the same magnitude whereas at higher speeds (greater than 200 Hz) the difference in response (corresponding to the source emission at the different wavelengths) is clear. The frequency range employed on the lock-in amplifier was150-300 Hz. Changing the time output from 10 ms to 100 ms did not make any significant difference, but increasing this to 1 s and 3 s meant the signal took longer to stabilise between changing filters.

Changing the gain and sensitivity of the signal increased the amplitude of the signals, thus increasing the difference between the different filters and also the sensitivity to changes in response, making detection of changes due to absorbing gases more noticeable. The DAQ card had a maximum output/input of ±5 V therefore increasing the sensitivity too much resulted in signal overload.

The response obtained with no filter was significantly higher than with the filters and tended to exceed the ±5 V cut-off, thus this response was ignored and only used for alignment purposes.

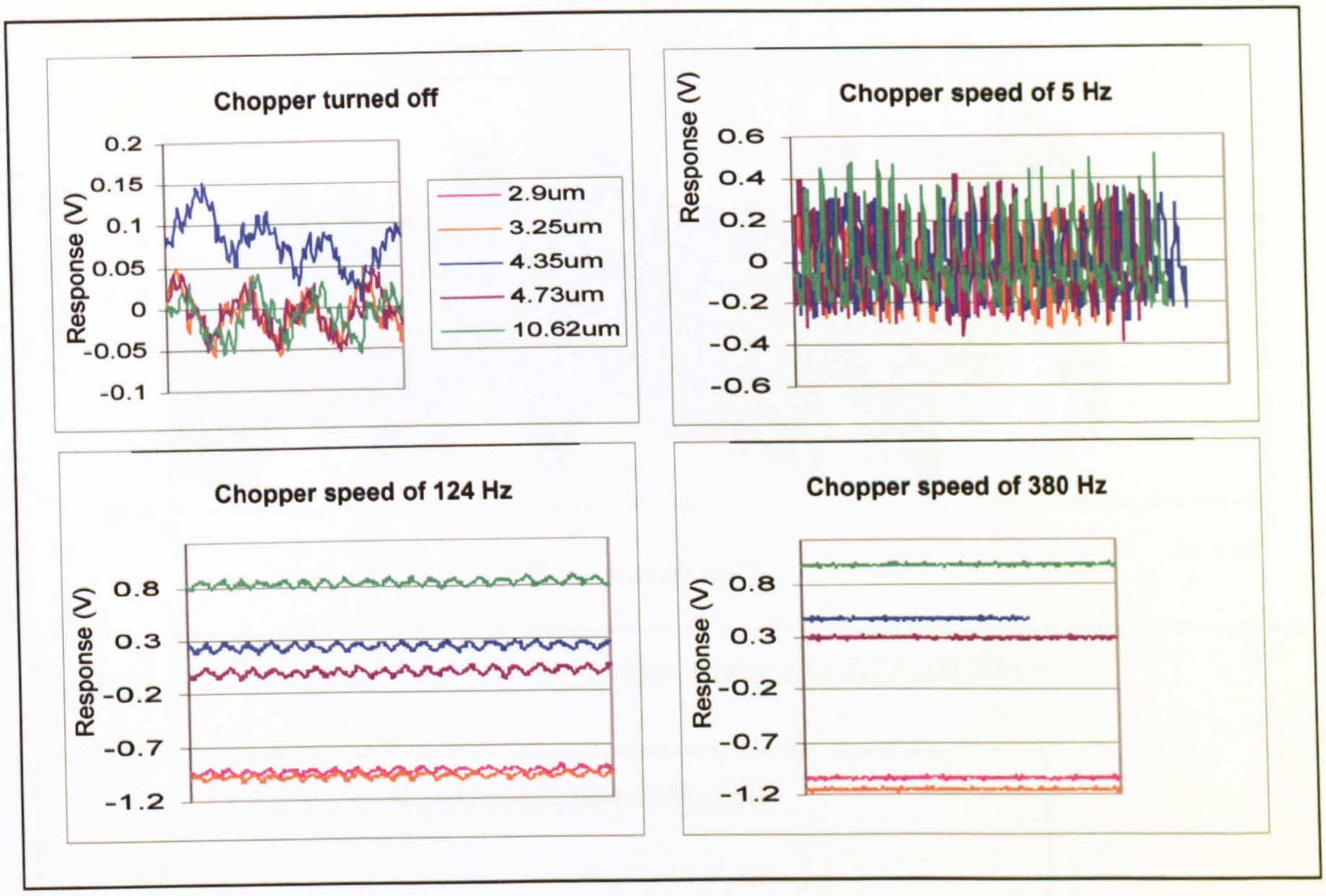


Figure 3.5.5 Effect of changing the chopper speed

3.5.5 Signal drift

A typical plot of signal drift is shown in figure 3.5.6; changing the lock-in conditions did little to rectify this drift. The most likely causes for the drift are variations in the blackbody source emission (due to small deviations in temperature of the source resulting in variations in emission), or variation in the detector response (later work indicated the problem lay mainly with the detector). Figure 3.5.7 shows the changes in response for each filter over time. In this case, the response shown at a given time for a specific filter is an average (calculated in Microsoft Excel) response over 10 seconds (LabVIEW™ records 100 measurements per second, thus 10 seconds corresponded to 1000 measurements). The different filter responses all follow a similar trend so it should be possible to compensate for this variation using the reference filter (10.6 μm response).

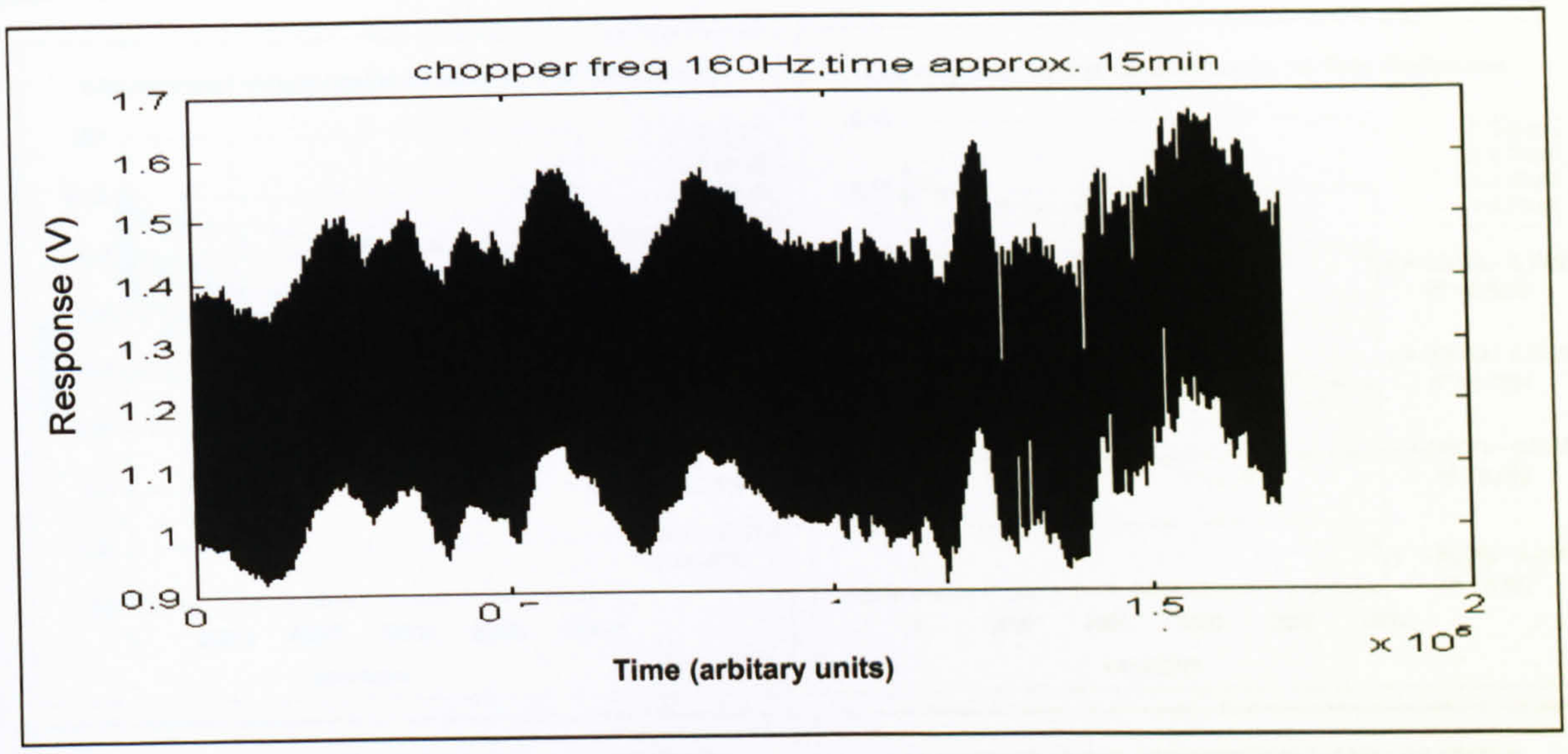


Figure 3.5.6 Change of response over time using the 3.25μm filter

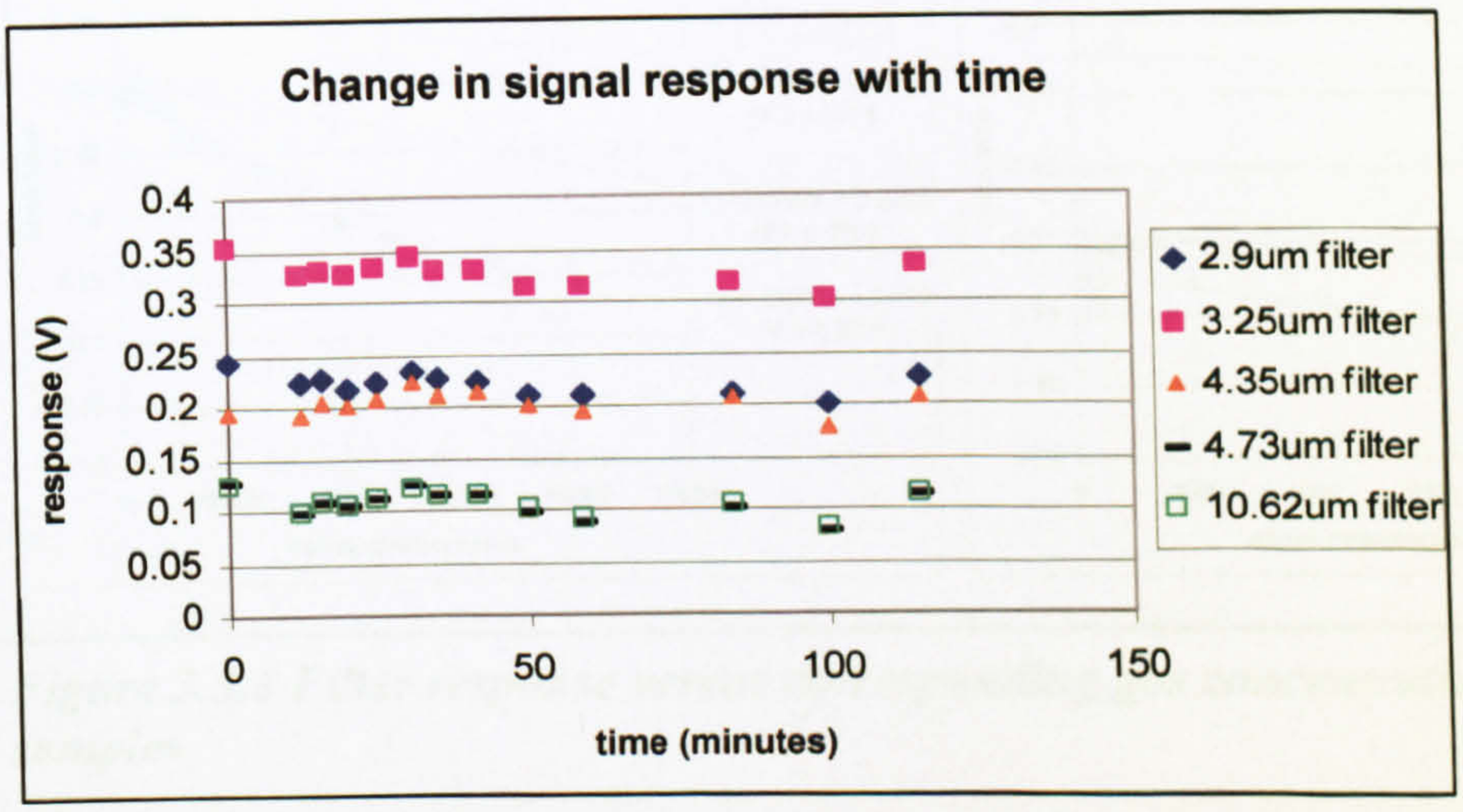


Figure 3.5.7 Change of response over time for different filters

3.5.6 Responses obtained for dilution runs

The results obtained from dilution runs are shown in figure 3.5.8. These are plots of filter response versus the concentration of gas that the filter relates to, i.e. 2.9 μm filter response versus N_2O concentration, 3.25 μm filter response versus CH_4 concentration, 4.35 μm filter response versus CO_2 concentration and 4.73 μm filter versus CO concentration. From these plots it can be seen that there are changes in response to changes gas concentration; however, there is poor sensitivity to small changes in gas concentrations (i.e. less than 100 ppm changes) as demonstrated by the poor correlations for concentrations of less than 10,000 ppm (N.B. The different magnitude in response between the two runs is due to changes in alignment in the optical set-up).

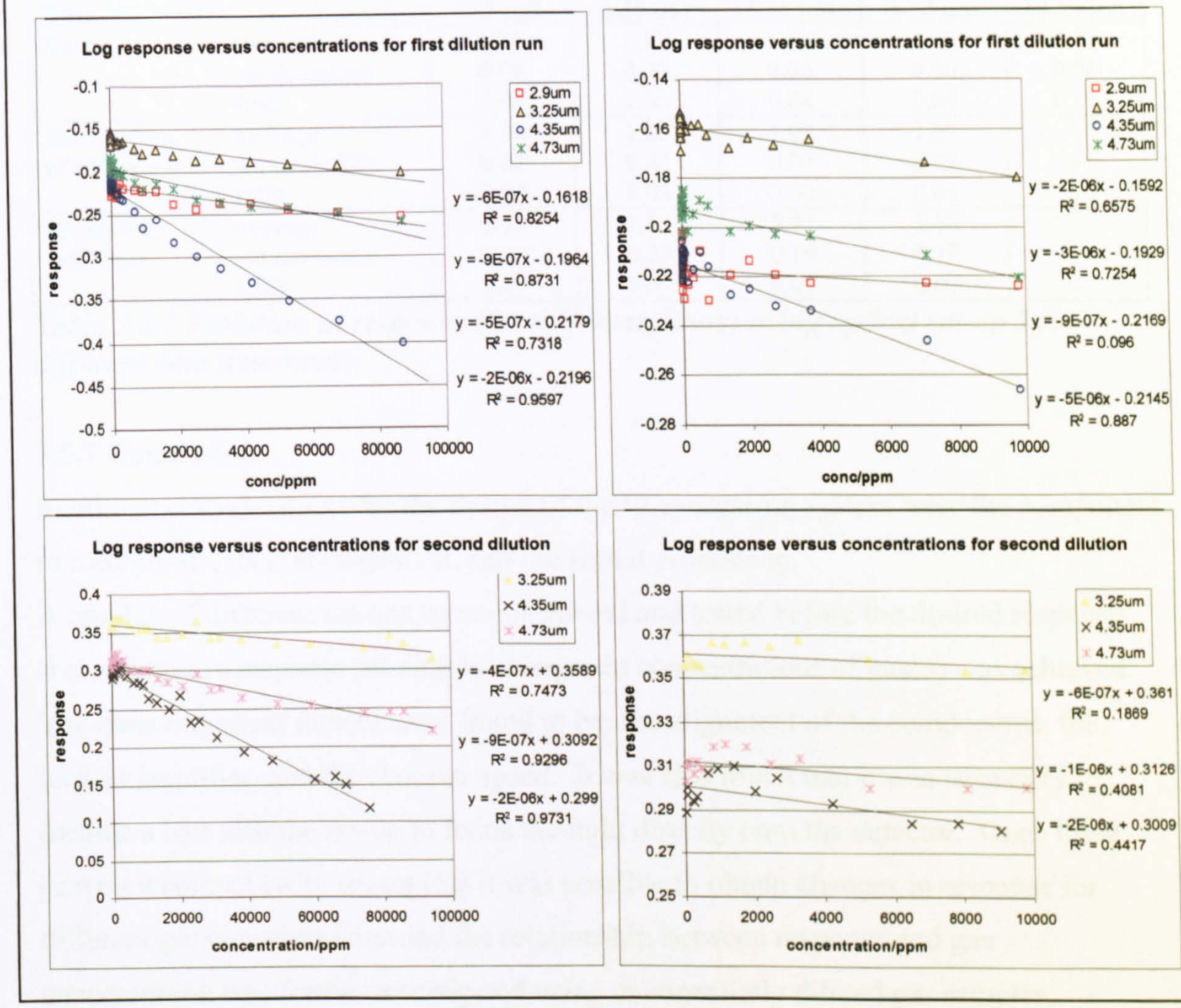


Figure 3.5.8 Filter response versus corresponding gas concentration for dilution of samples

Comparing these results with those obtained when using the FTIR it is clear that there was significant room for improvement in the optical set-up.

One of the main problems causing the lack of sensitivity was the drift in response (as shown earlier in figures 3.5.2 and 3.5.3). Repeat measurements were made over a 3-hour interval, measuring the same gas sample without changing any of the optical layout or amplifier settings so that the RSDs for different data treatments could be compared (table 3.5.1). This showed subtraction of the reference filter (10.62μm) response was more successful than dividing by this response.

Data treatment		2.9 μm	3.25 μm	4.35 μm	4.73 μm	10.62 μm
Raw	average	1.76	1.88	1.63	1.67	0.61
	std deviation	0.06	0.07	0.06	0.06	0.06
	RSD	0.04	0.04	0.04	0.04	0.11
Subtraction of reference	average	1.16	1.28	1.02	1.07	-
	std deviation	0.02	0.04	0.02	0.01	-
	RSD	0.02	0.03	0.02	0.01	-
Divided by reference	average	2.93	3.13	2.71	2.78	-
	std deviation	0.21	0.23	0.19	0.19	-
	RSD	0.07	0.07	0.07	0.07	-

Table 3.5.1 Variation in response (for different filters) using optical set-up 5 for different data treatments

3.5.7 Conclusion

Important considerations for the design of the IR measuring system were the component to incorporate, their arrangement, and the signal processing.

A number of different set-ups were considered and tested before the desired response (i.e. changes in response relating to changes in concentrations of gases) was achieved. The most important factors were found to be; the alignment of the components, the lock-in amplifier, and the chopper speed. It was also found that it was necessary to include a lens into the set-up to focus the light directly onto the detector. Once these criteria were met (with set-up (5)) it was possible to obtain changes in response for different gas concentrations and the relationship between response and gas concentration was further investigated using exponentially diluted gas samples.

The results obtained for the dilution runs gave linear changes in response for 10,000 ppm changes in response; however, there was poor sensitivity to smaller changes of less than 100 ppm in concentrations. In the previous work (section 3.4) where similar sample dilutions were prepared and the response for different filters employed in the FTIR were collected, a higher sensitivity was achieved for lower concentrations.

Since the previous work utilised the same main components (with the exception of the lock-in amplifier) i.e. the same light source, detector, gas cell and narrowband filters, it could be assumed that there was considerable room for improvement either with the optical set-up or in the signal processing.

One of the main sources of potential problems was the effect of ambient air and temperature on the optical set-up. The set-up was unenclosed therefore ambient air and light could effect the measurements. Another problem was the non-permanent fixture of the components; this was to facilitate the easy rearrangement of components to investigate the effect of different arrangements, but had the disadvantage that the

components required realignment for each session, thus affecting the repeatability of results collected over different sessions. To counteract these problems the next stage of the work was to design and build a sealed box (that could be purged) to house the optical set-up with a metal plate to which all the components to be fixed in place.

3.6 The purgable box

To remove the effects of ambient air, in particular CO_2, and to also reduce the amount of stray light reaching the detector, it was decided to build a black sealed box that could be purged. This work looks at the testing of this box, which is designed to be a prototype for the IR analyser that will eventually be built with QWIPs.

Description of box

A rectangular box was made of black vinyl material, although this resulted in a lot of empty space thus increasing the area to purge (compared to a T-shaped box which would be the minimum geometry required to enclose the source, detector and gas cell), it was the simplest design to construct.

The optical alignment of the components was based on set-up 5 from the previous work (section 3.5). The components were mounted on a metal plate to keep them fixed in place thus eliminating the need to constantly check the alignment.

3.6.1 Replacement of MCT detector with the DTGS detector

Originally the MCT detector was employed, but problems with large drops in response over short time intervals occurred due to loss of the vacuum in the dewer flask. While it was being fixed, the DTGS detector was utilised, this had not been used in the previous set-ups due to the physical restraints of the apparatus; however, with the new box, connectors and independent power supply (previously the IR source was still powered from the FTIR so could only be placed in limited positions) this was now possible.

The sensitivity of DTGS detectors was lower than for the MCT detector and it was found that the sensitivity setting on the lock-in amplifier needed to be increased in order to obtain responses of a similar magnitude to those obtained by the MCT.

Initial results obtained with the DTGS detector were found to be more consistent and less prone to long-term drift than the MCT. Figure 3.6.1 compares the drift obtained over a two-minute interval using the 2.9 μm filter with the DTGS detector.

3.6.2 Effect of purging

Purging the box resulted in a drop in signal which eventually levelled off. Increasing the flow rate resulted in a larger drop in signal. Possibly this was due to a cooling effect of the helium flow on the source or some kind of pressure effect. An example of this is

shown in figure 3.6.2. Once the responses had stabilised, a 20 cm^3min^{-1} flow rate of helium was passed through the box (from the reference column outlet on the GC), this was to ensure the box was constantly purged.

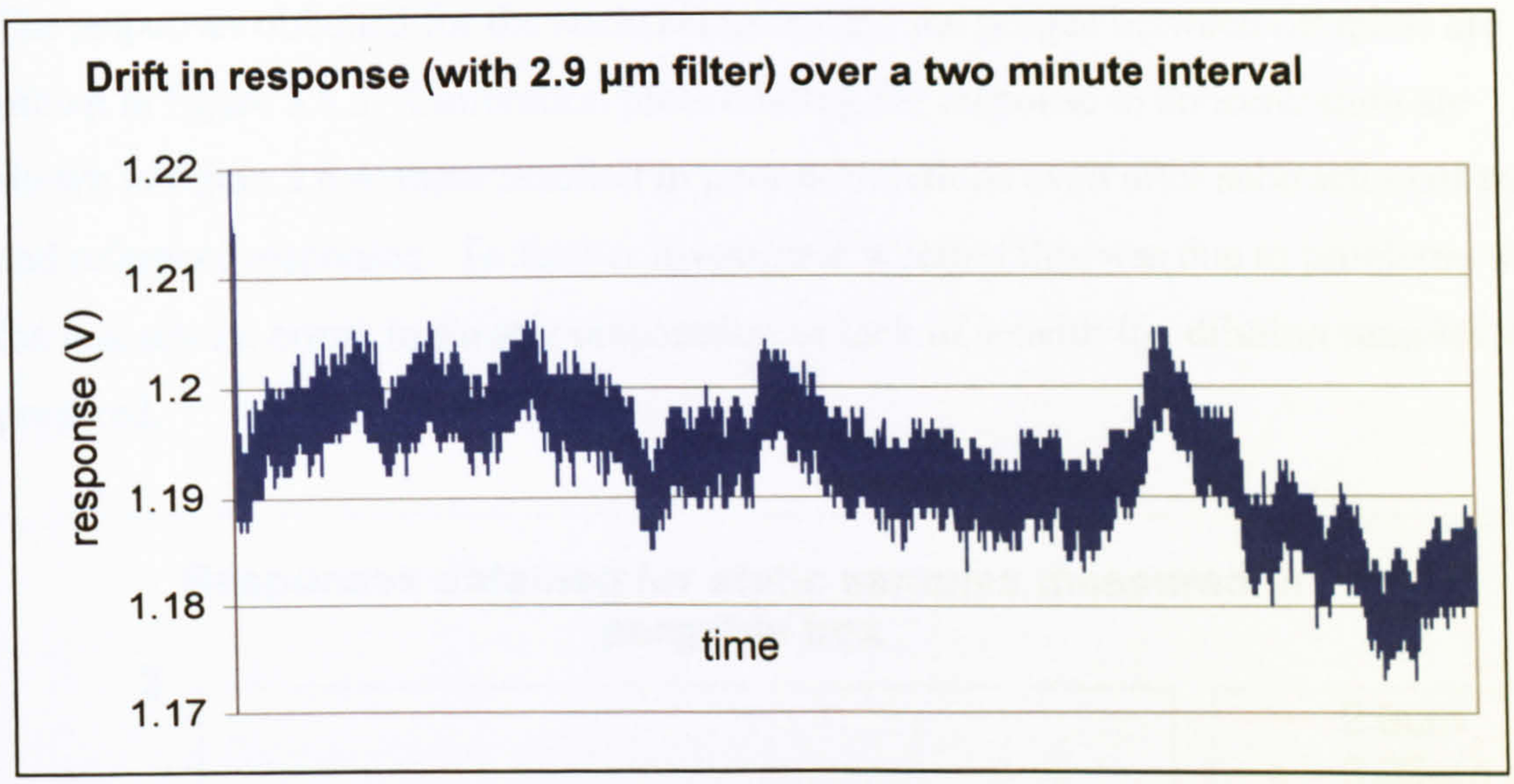


Figure 3.6.1 Drift in response from DTGS using 2.9µm filter (in purgable box set-up)

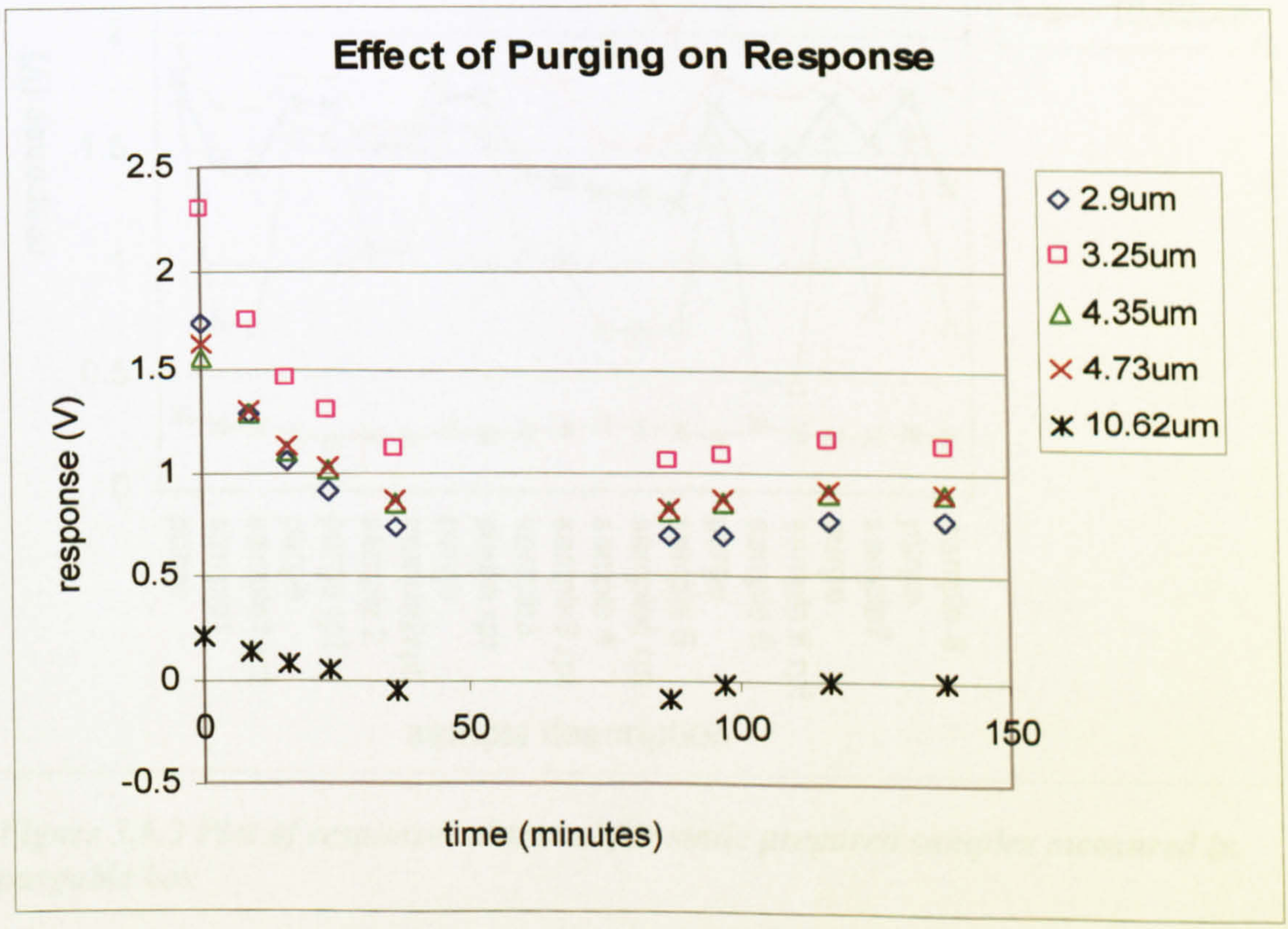


Figure 3.6.2 The effect of purging (60 cm^3min^{-1} flow rate) on the responses obtained with different filters

3.6.3 Testing the box

3.6.3.1 Static samples

The responses obtained for the static samples (and the purges between samples) are shown in figure 3.6.3. Calibration plots relating the response to concentration are shown in figure 3.6.4, these resulted in poor correlations even after subtraction of purge and reference responses. To further investigate whether this was due to problems with the box set-up, errors in sample preparation or lack of sensitivity, dilution samples were prepared.

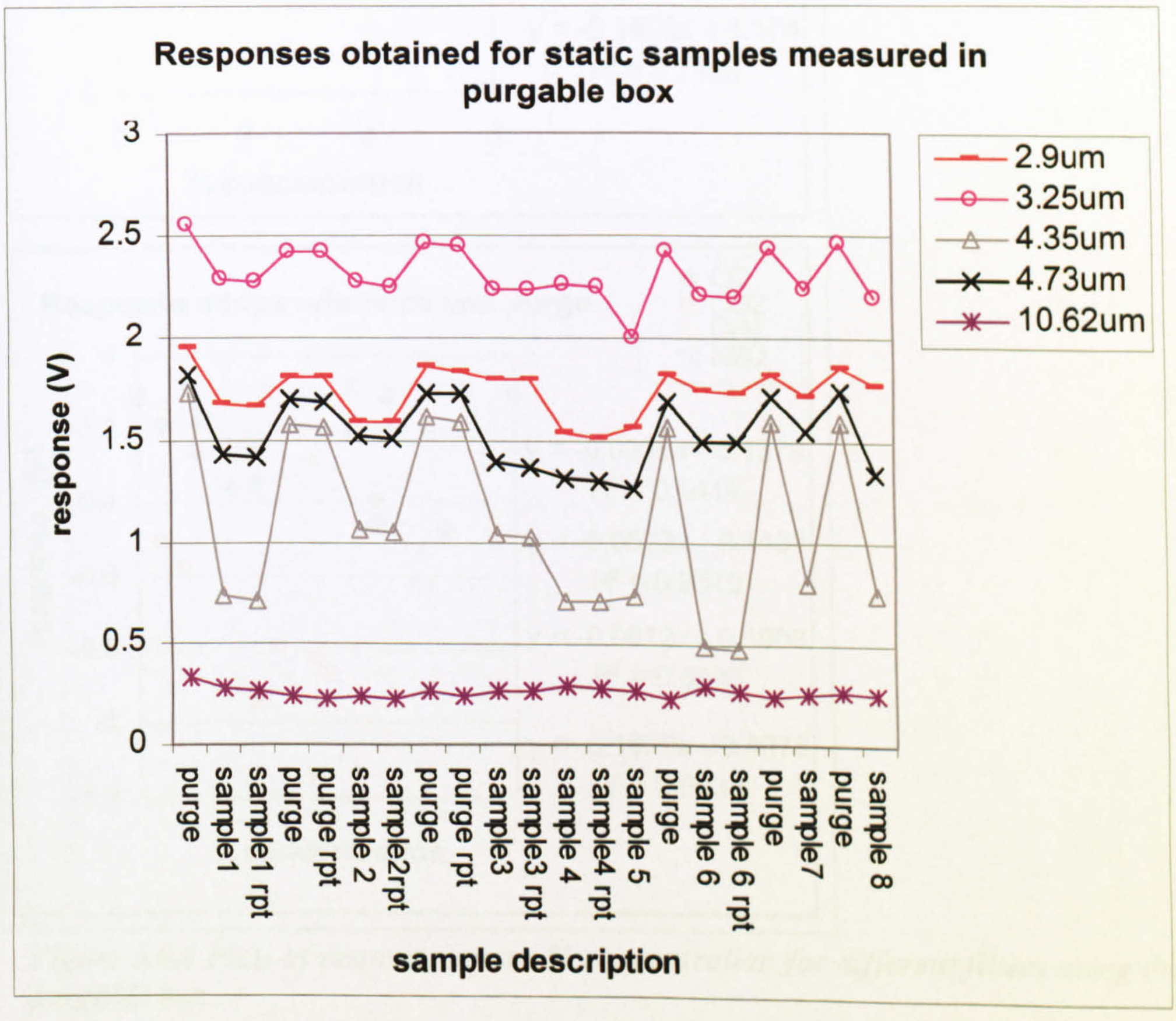
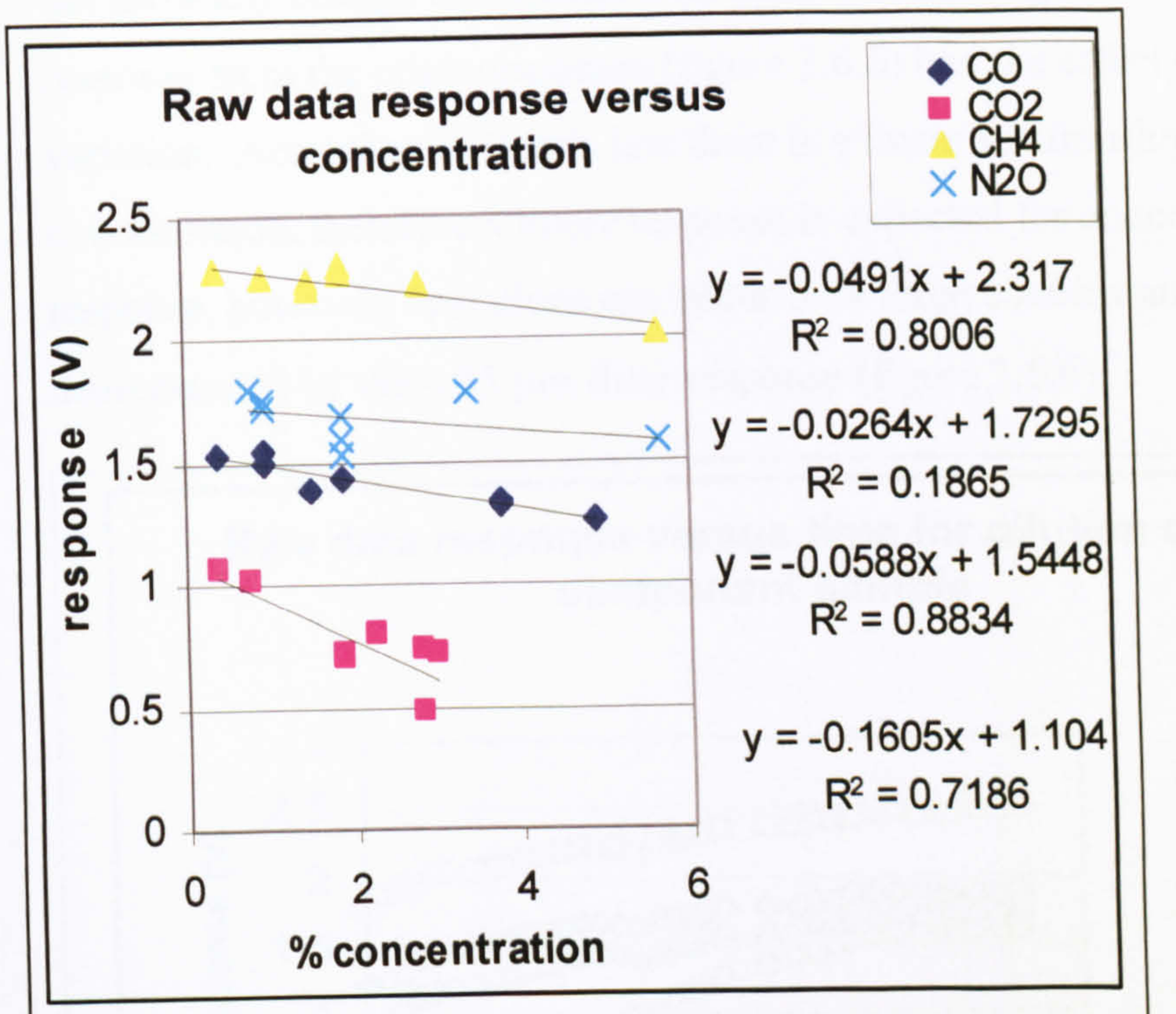


Figure 3.6.3 Plot of responses obtained for static prepared samples measured in purgable box



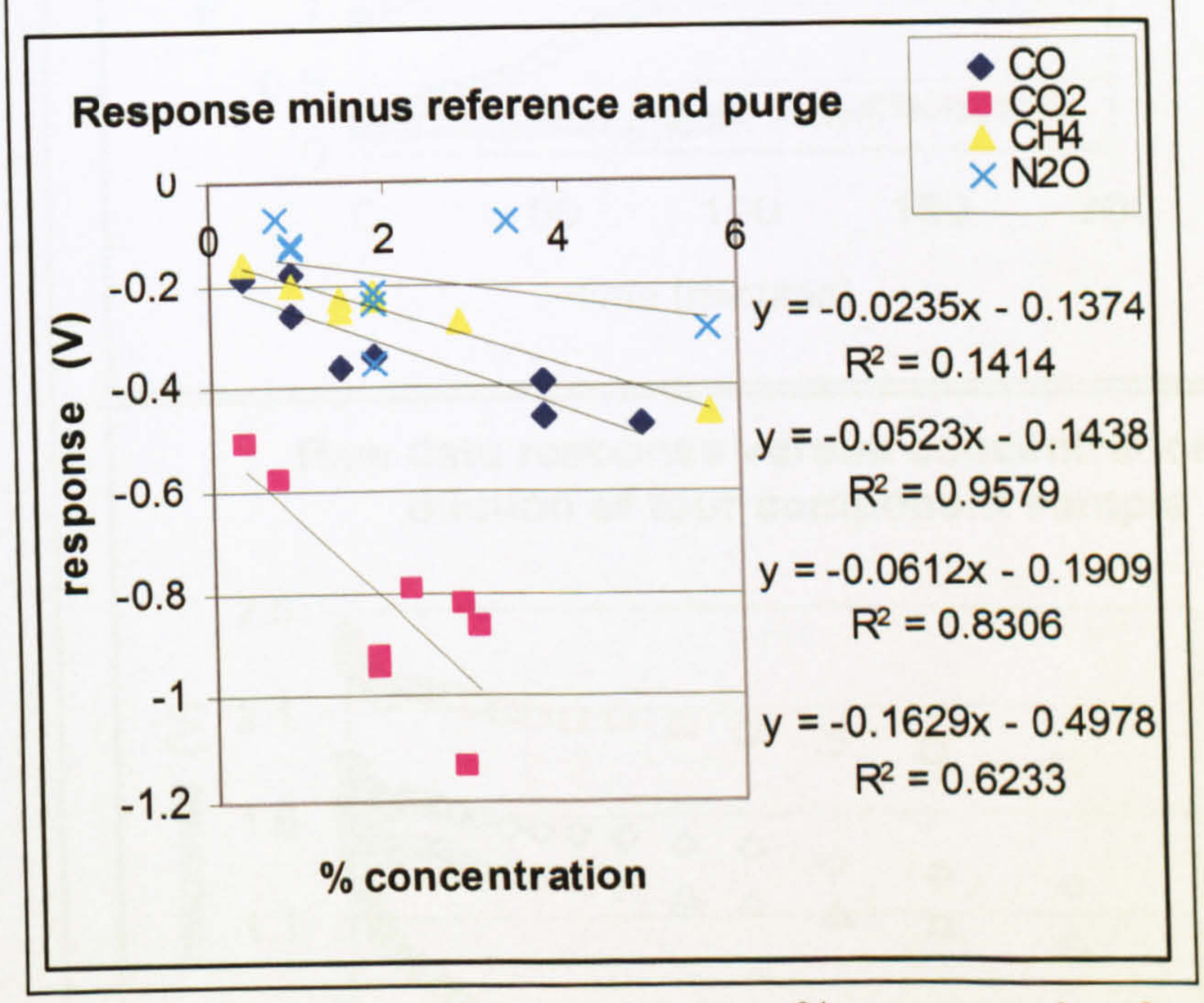


Figure 3.6.4 Plots of response versus % concentration for different filters using the purgable box

3.6.3.2 Dilution samples

Figure 3.6.5 displays the results for the dilution of a mixture of the four gases (sample 1). Looking at the raw data, there are clear 'kinks' in the response that are particularly noticeable when the time measurements are converted to corresponding concentration levels; these are reflected in the responses for the 10.62 μm filter, which as there is no SF_6 (or any other species that absorbs at this wavelength) present in the sample, should

not show any change in response over time. Subtracting the 10.62 μm reference response from the other responses (figure 3.6.6) had the effect of removing most of this variation. According to Beer's law there is a linear relationship between absorption and concentration, therefore a linear response is expected for concentration versus log response; however, deviations can occur over large concentration ranges as demonstrated by the 4.35 μm filter response (figure 3.6.6).

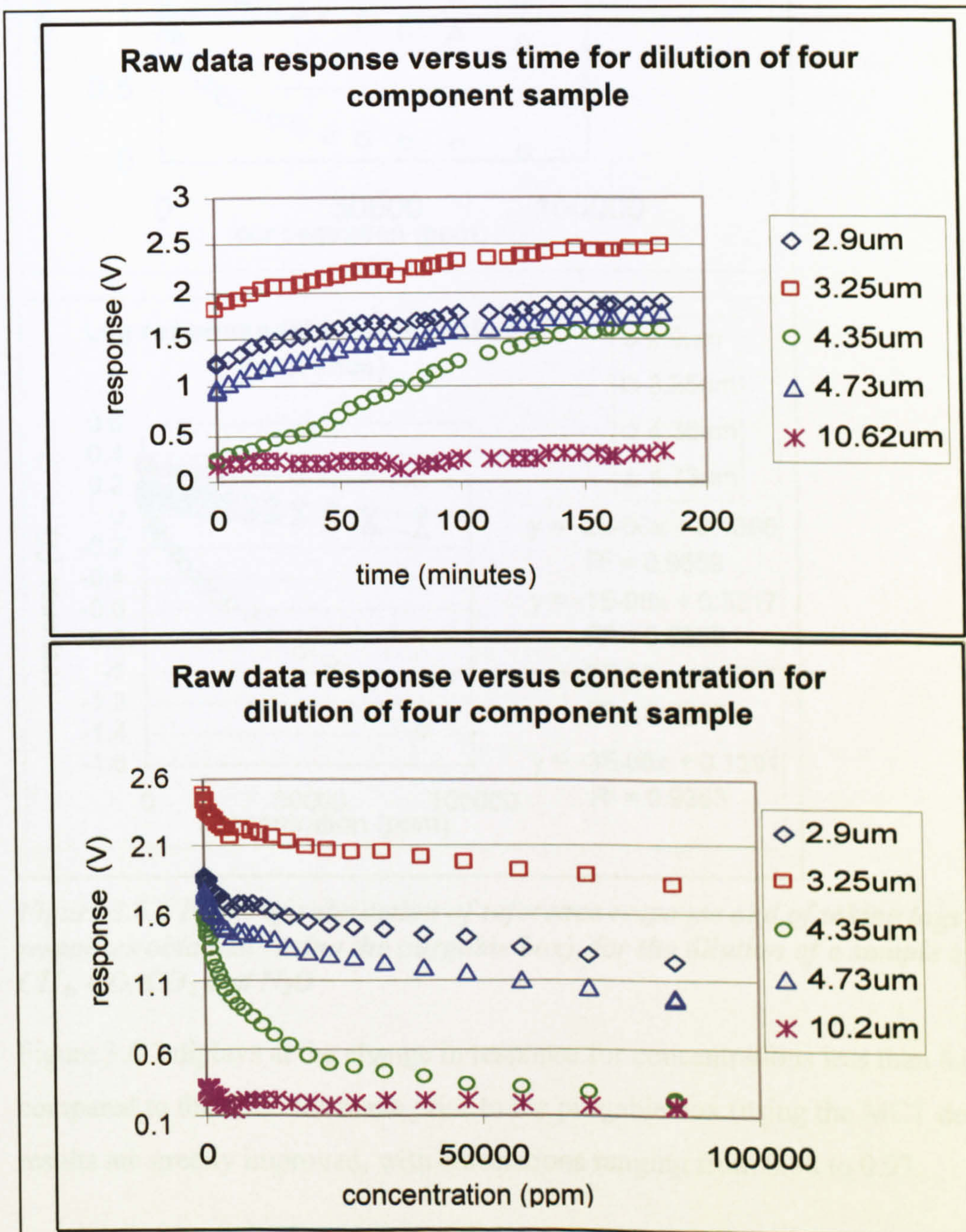


Figure 3.6.5 Raw responses obtained using the purgable box, for the dilution of a sample mixture of CH_4, CO, CO_2 and N_2O using a flow rate of 20 cm^3min^{-1}. (N.B. The concentrations are calculated according to equation 1.4 shown in section 1.2.2)

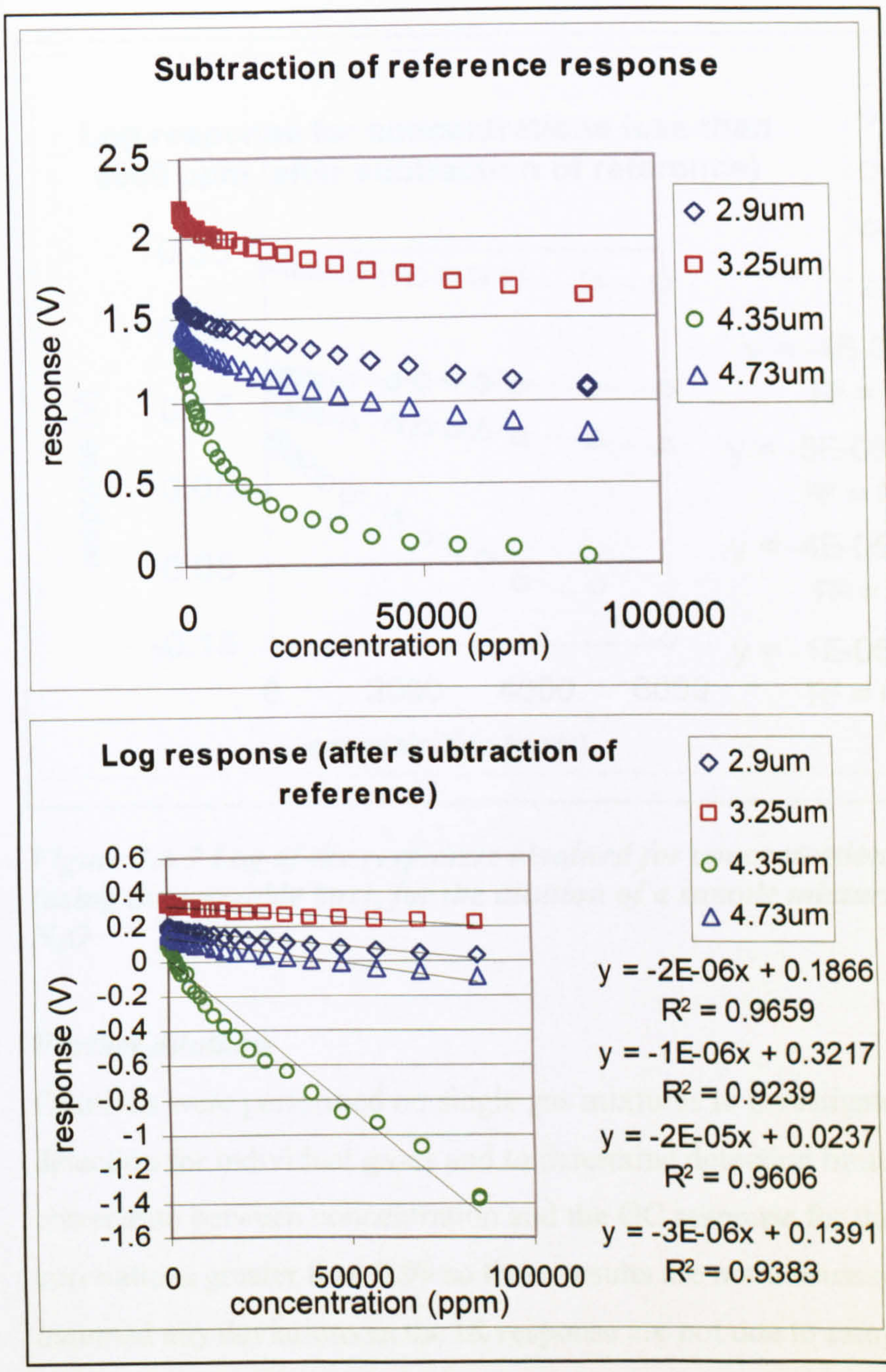


Figure 3.6.6 Effect of subtraction of reference response and of taking logs on the responses obtained (using the purgable box), for the dilution of a sample mixture of CH_4, CO, CO_2 and N_2O

Figure 3.6.7 diplays at the change in response for concentrations less than 6,000 ppm, compared to the dilutions made prior to the purgable box (using the MCT detector) the results are greatly improved, with correlations ranging from 0.88 to 0.97.

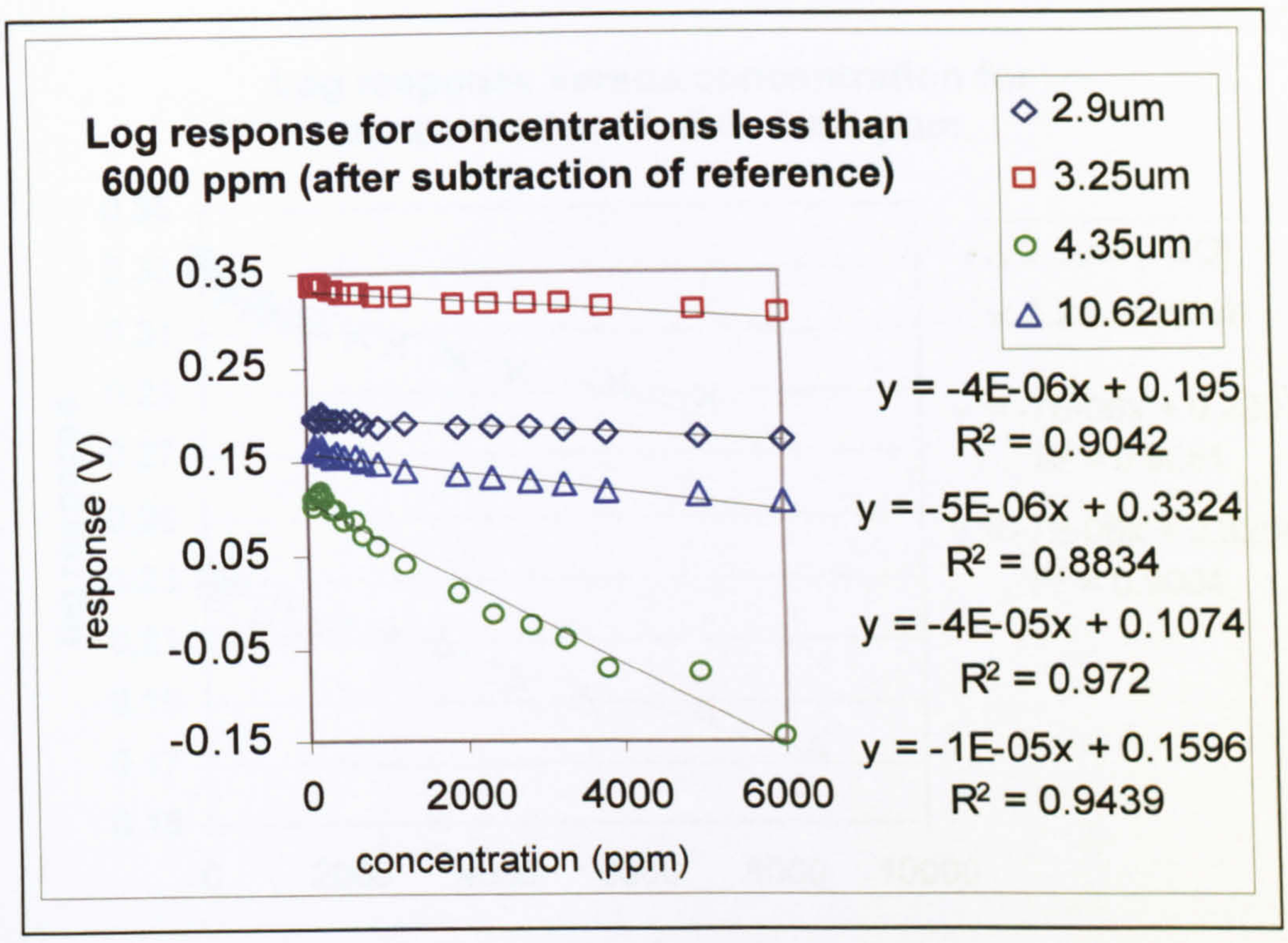


Figure 3.6.7 Log of the responses obtained for concentrations of less than 6,000 ppm (using the purgable box), for the dilution of a sample mixture of CH_4, CO, CO_2 and N_2O

Further dilutions

Dilutions were performed on single gas mixtures to investigate the sensitivity of detection for individual gases and to determine detection limits. In all cases the correlation between concentration and the GC response for the gases of interest gave correlations greater than 0.99 so these results are not discussed further, and it can be assumed any deviations in the IR response are not due to sample variations.

The correlations for log response versus concentration, over the concentration range 1,000 to 60,000 ppm for N_2O and CH_4 were both 0.97.

For concentrations less than 10,000 ppm the R^2 values obtained for N_2O and CH_4 were 0.956 and 0.900 respectively. Responses for concentrations less than 1000 ppm were varied indicating the measuring system was not sensitive enough for these concentrations. Removing responses below 1000 ppm resulted in R^2 values of 0.947 and 0.954 for N_2O and CH_4 (refer to figure 3.6.8).

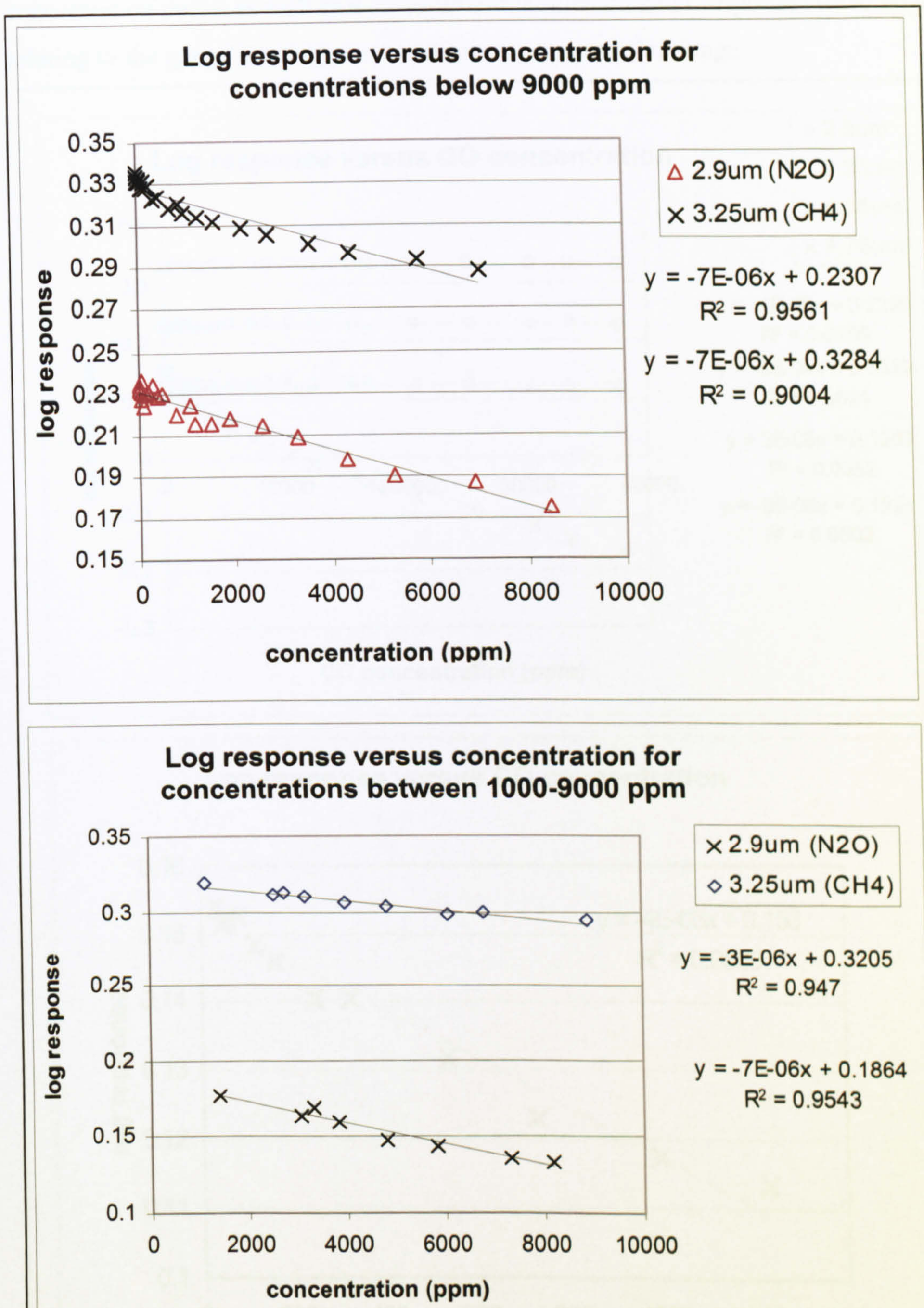
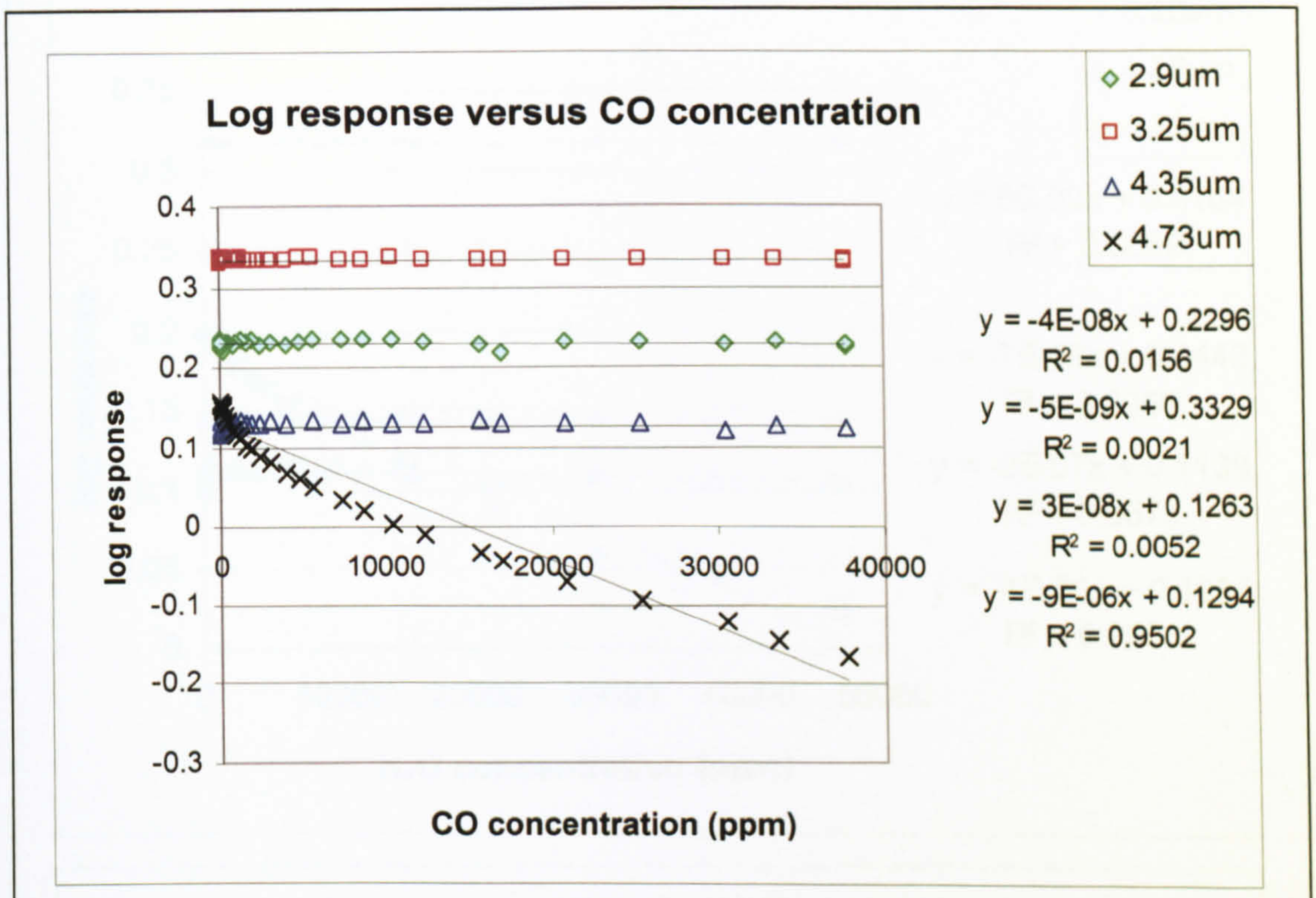


Figure 3.6.8 Log of the responses obtained for concentrations of less than 10,000 ppm, using the purgable box, for the dilution of a sample mixture of CH_4 and N_2O

Correlation plots obtained for the dilutions of individual gas samples are shown in figures 3.6.9 to 3.6.12. These display the different filter responses over the entire

concentration range, as well as a close-up of the relevant filter response (wavelength relating to the gas of interest) over a narrower concentration range.



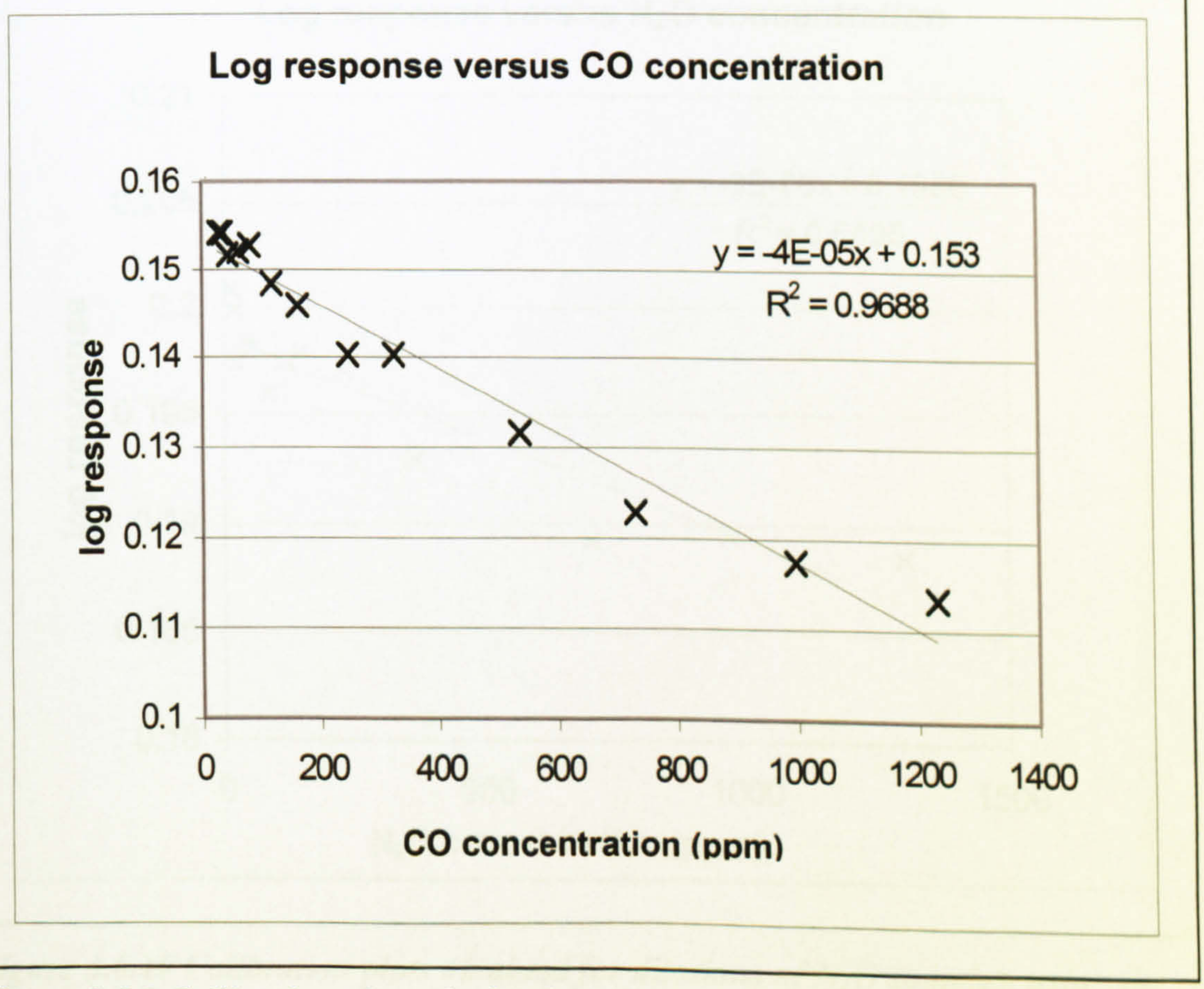
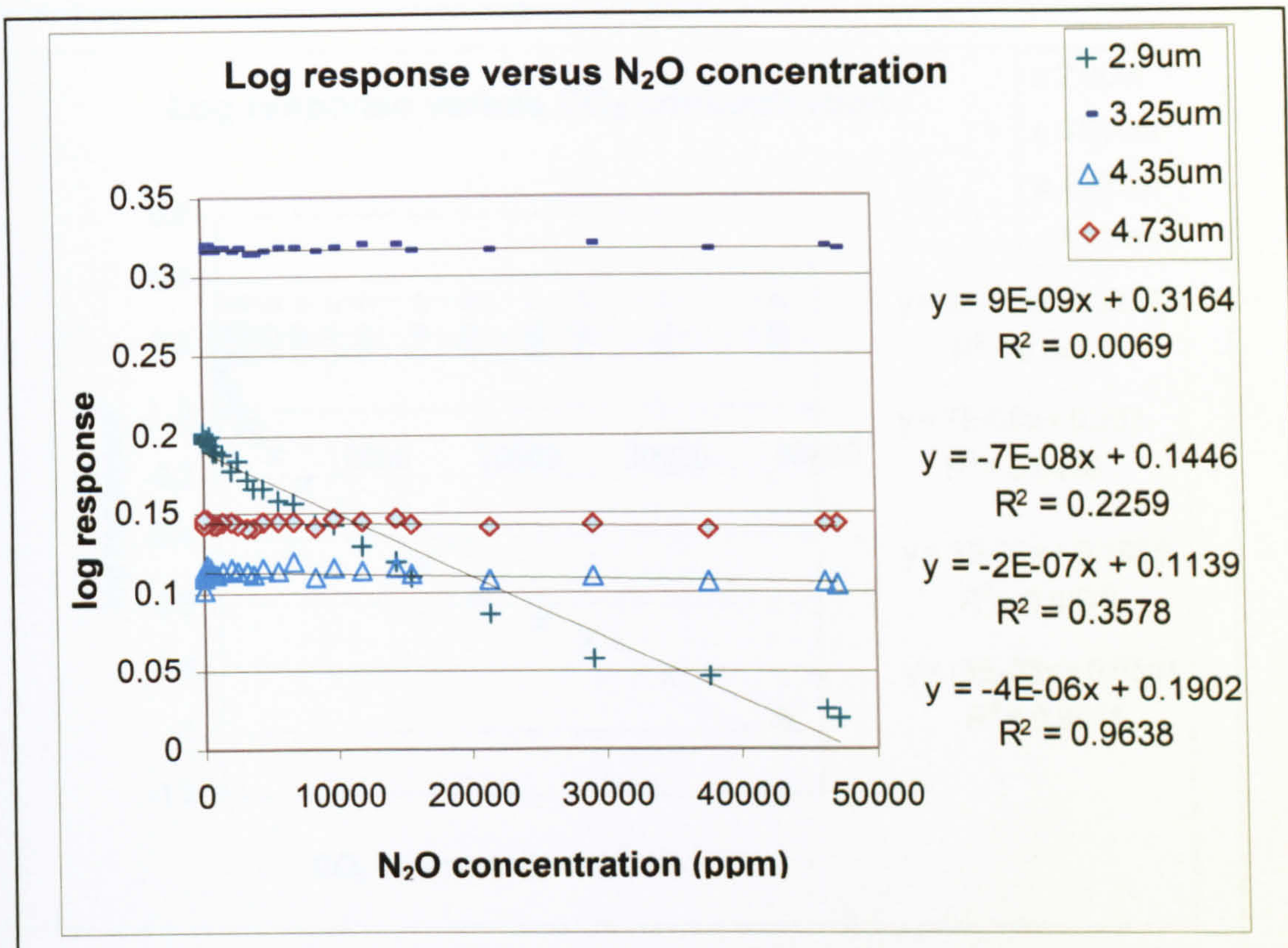


Figure 3.6.9 Calibration plots obtained for dilutions of CO samples, using the purgable box



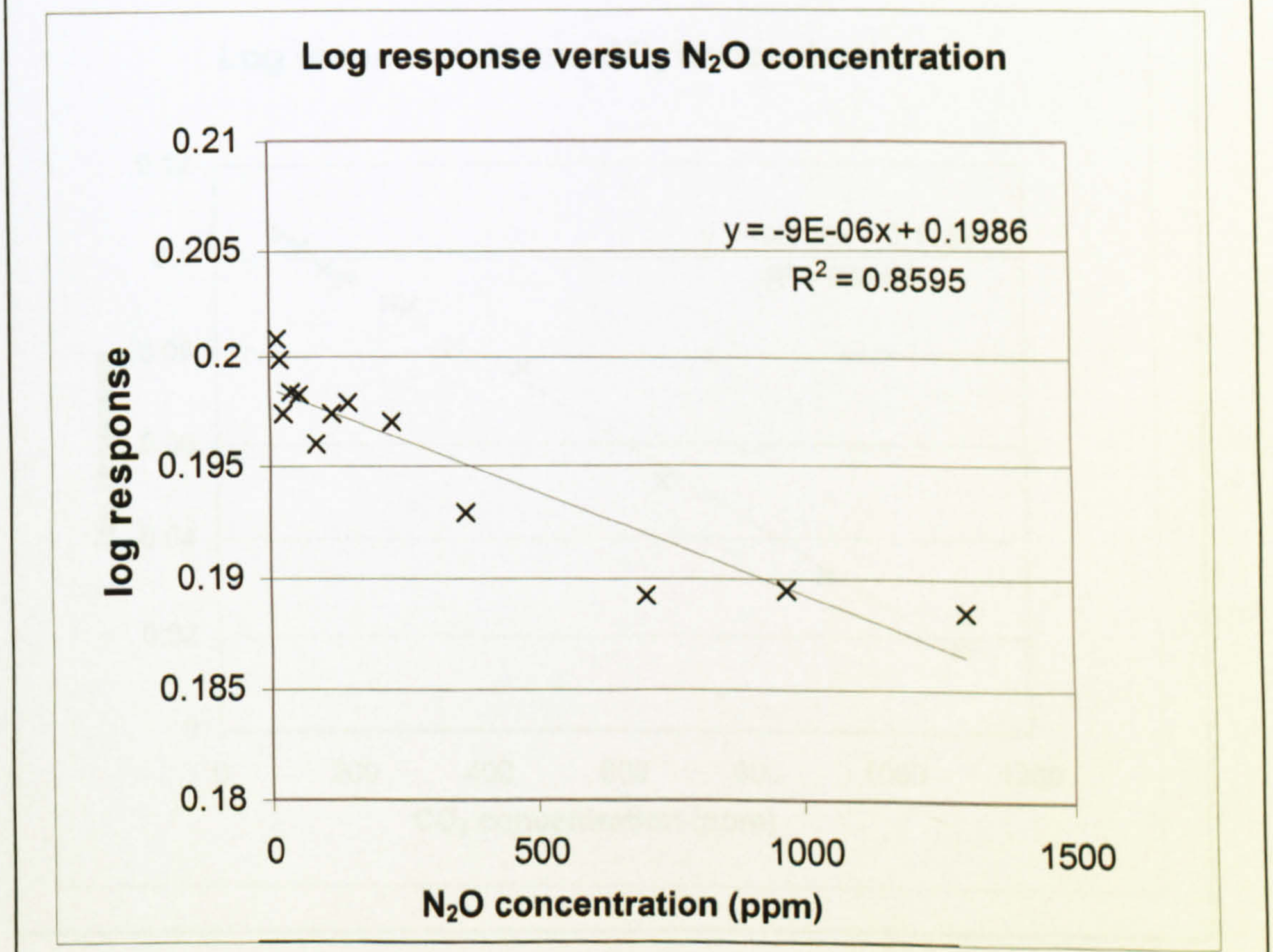
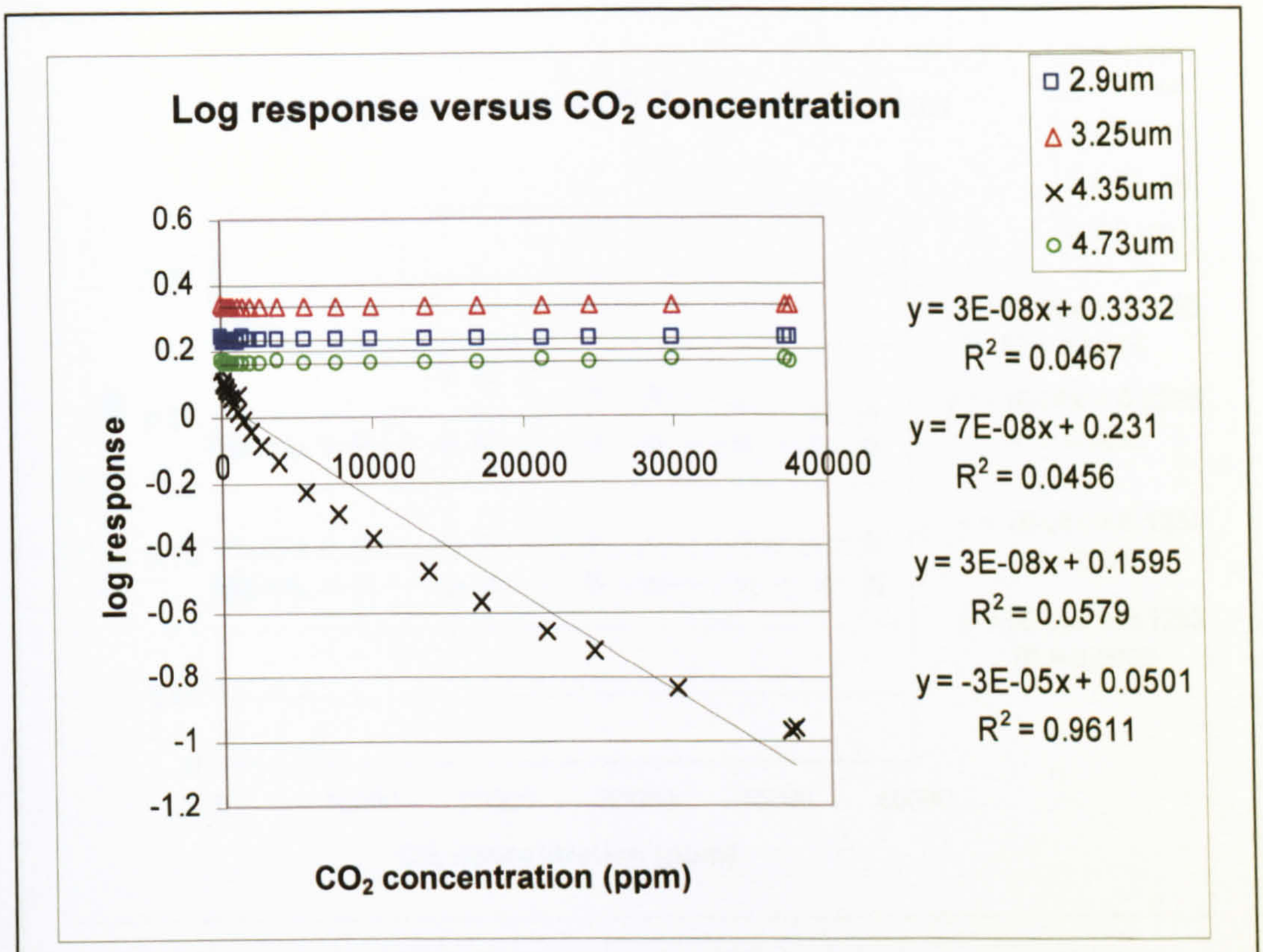


Figure 3.6.10 Calibration plots obtained for dilutions of N_2O samples, using the purgable box



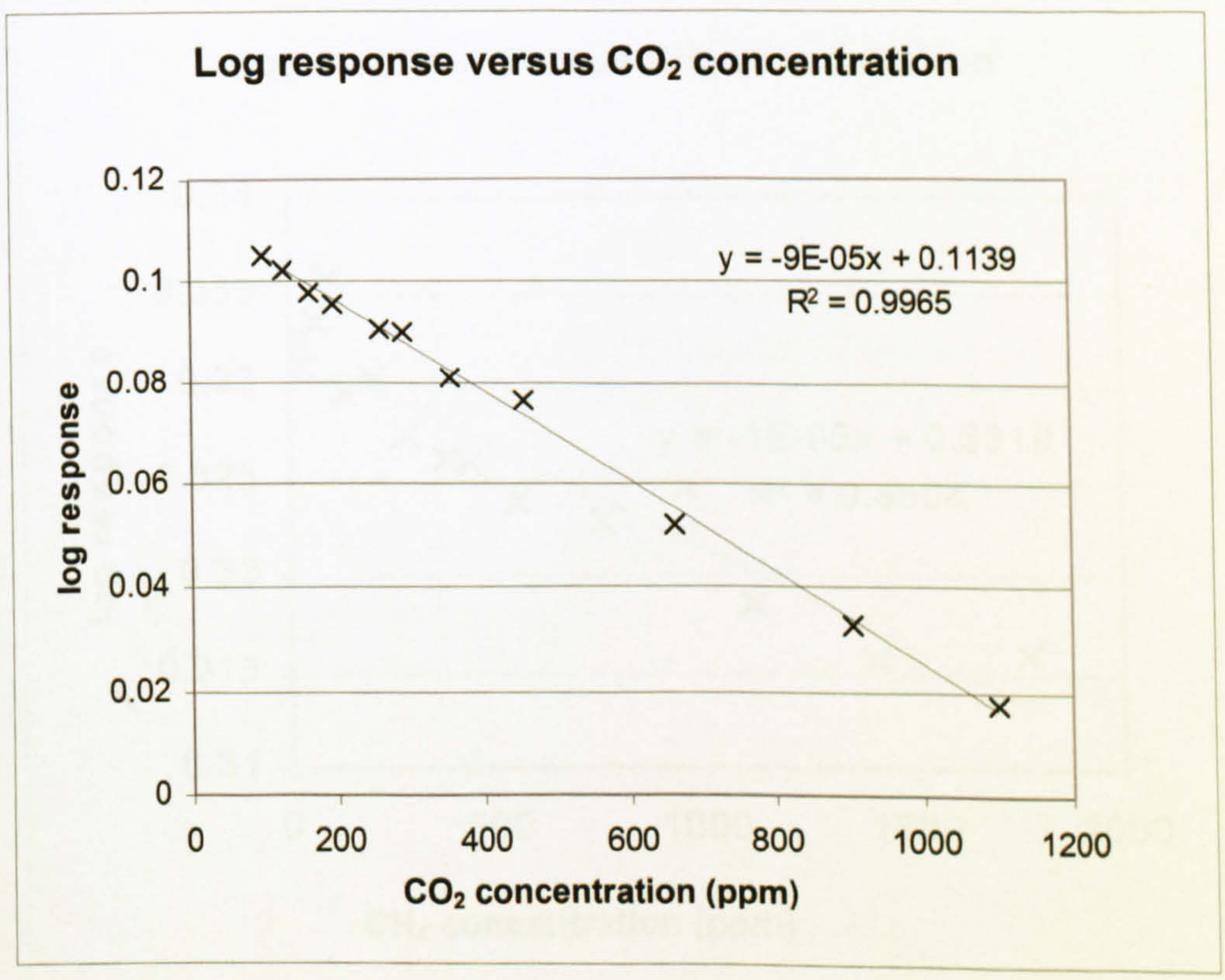
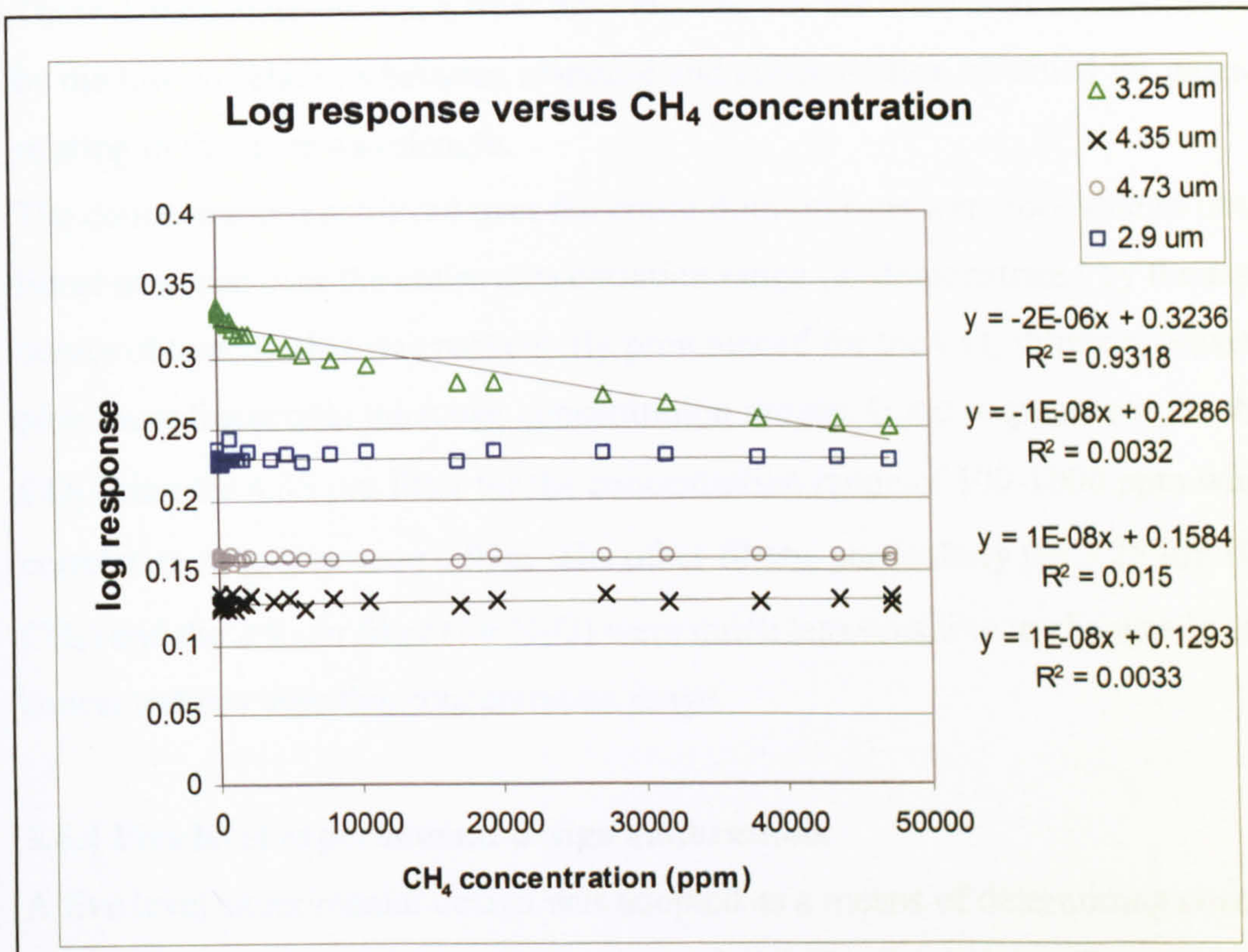


Figure 3.6.11 Calibration plots obtained for dilutions of CO_2 samples, using the purgable box



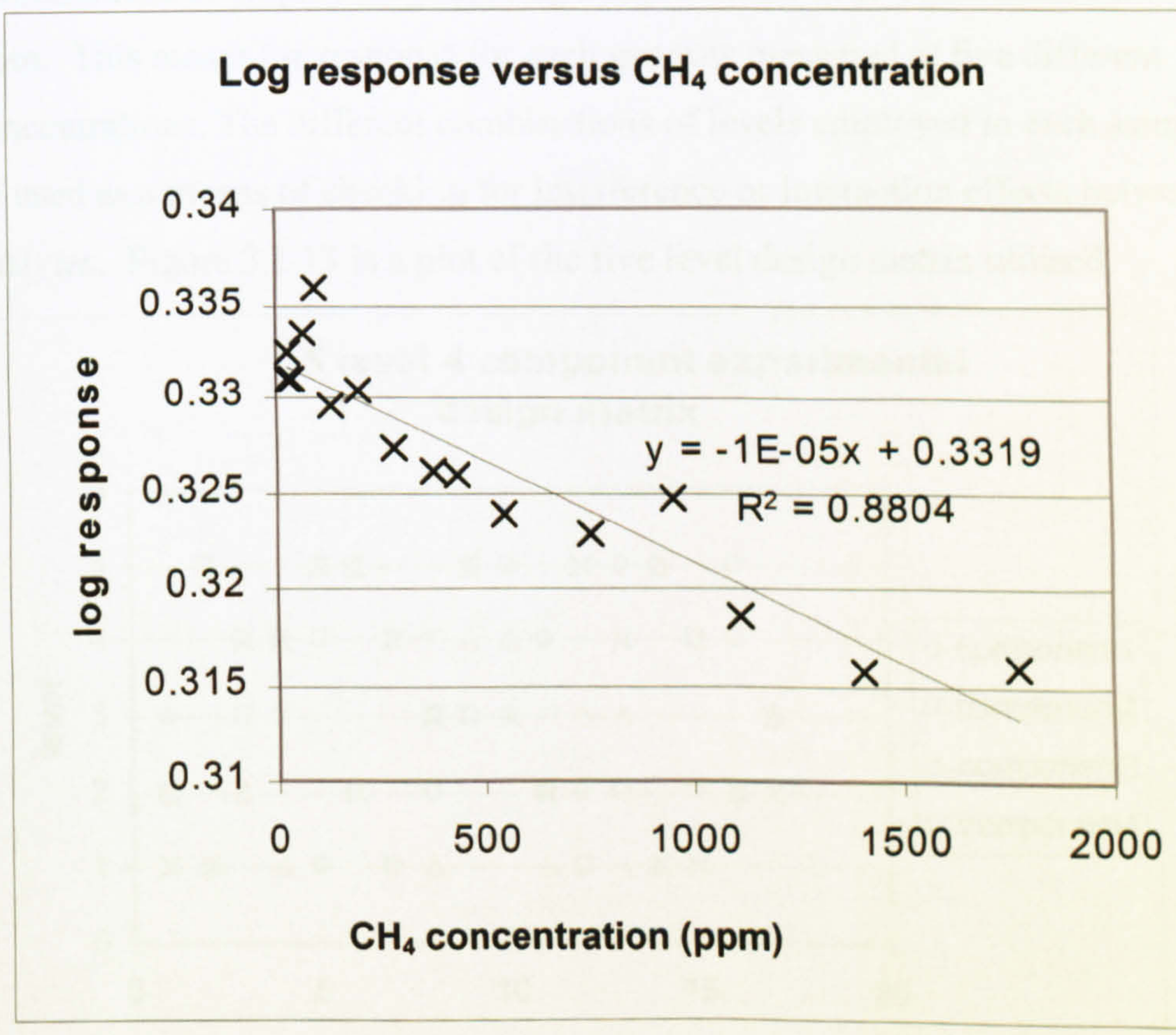


Figure 3.6.12 Calibration plots obtained for dilutions of CH_4 samples, using the purgable box

These demonstrate that each filter only responds to the analyte of interest, as indicated by the low correlations between response and concentration obtained for gases not relating to the filter wavelength.

The concentrations achieved over the entire dilution runs were too wide to obtain a linear response over the entire concentration range (as demonstrated by the response curves obtained); this was particularly pronounced for the CO_2 dilution; however, the plots were linear over narrower concentration ranges. Good responses were obtained for CO_2 using the 4.35 μm filter for the concentration range of 100-1000 ppm which gave a correlation of greater than 0.996. The other filters, particularly the 3.25 μm filter (for CH_4) and the 2.9 μm filter (for N_2O) were much less sensitive to changes in analyte concentrations over this concentration range.

3.6.4 Five level experimental design calibrations

A five level experimental design was adopted as a means of determining concentration combinations to employ for preparing sample mixtures to build calibration models from. This meant the response for each gas was measured at five different concentrations. The different combinations of levels employed in each sample can also be used as a means of checking for interference or interaction effects between different analytes. Figure 3.1.13 is a plot of the five level design matrix utilised.

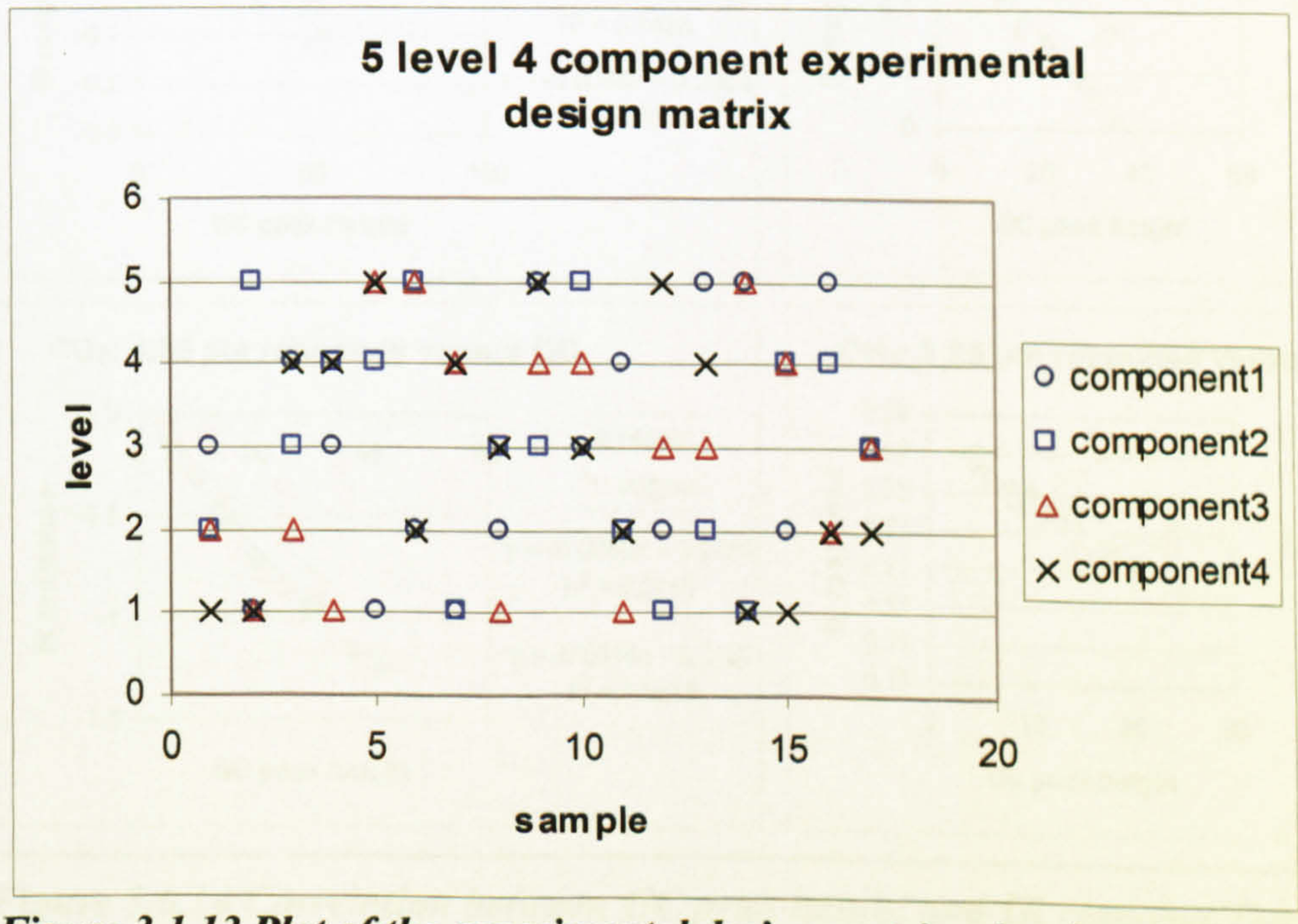


Figure 3.1.13 Plot of the experimental design matrix, showing the levels of each component employed for each sample

Results from the initial samples gave very poor correlations between level (gas concentration) and responses obtained by GC and IR, therefore the work was repeated to check whether this was due to equipment or analyst errors. Previous work relating the FTIR response obtained with the narrowband filters to GC peak heights gave good correlations for CH_4, CO_2, and N_2O, thus correlation plots between the IR and GC response were utilised to indicate whether the poor correlation between responses and concentration levels were due to errors in sample preparation or with the measurement system. Figure 3.6.14 shows correlation plots between GC and IR responses (log of response after subtraction of reference) for the original data and the repeated data. The original data gave poor correlations, so this was assumed to be due to instrumental or sample preparation error. The data collected after repeating the work gave relatively good correlations between the GC and IR responses, therefore it can be assumed the previous data was erroneous.

Figure 3.6.15 displays plots of IR response versus concentration level (for the repeated data), ideally there should be distinct differences in response for each of the different concentration levels.

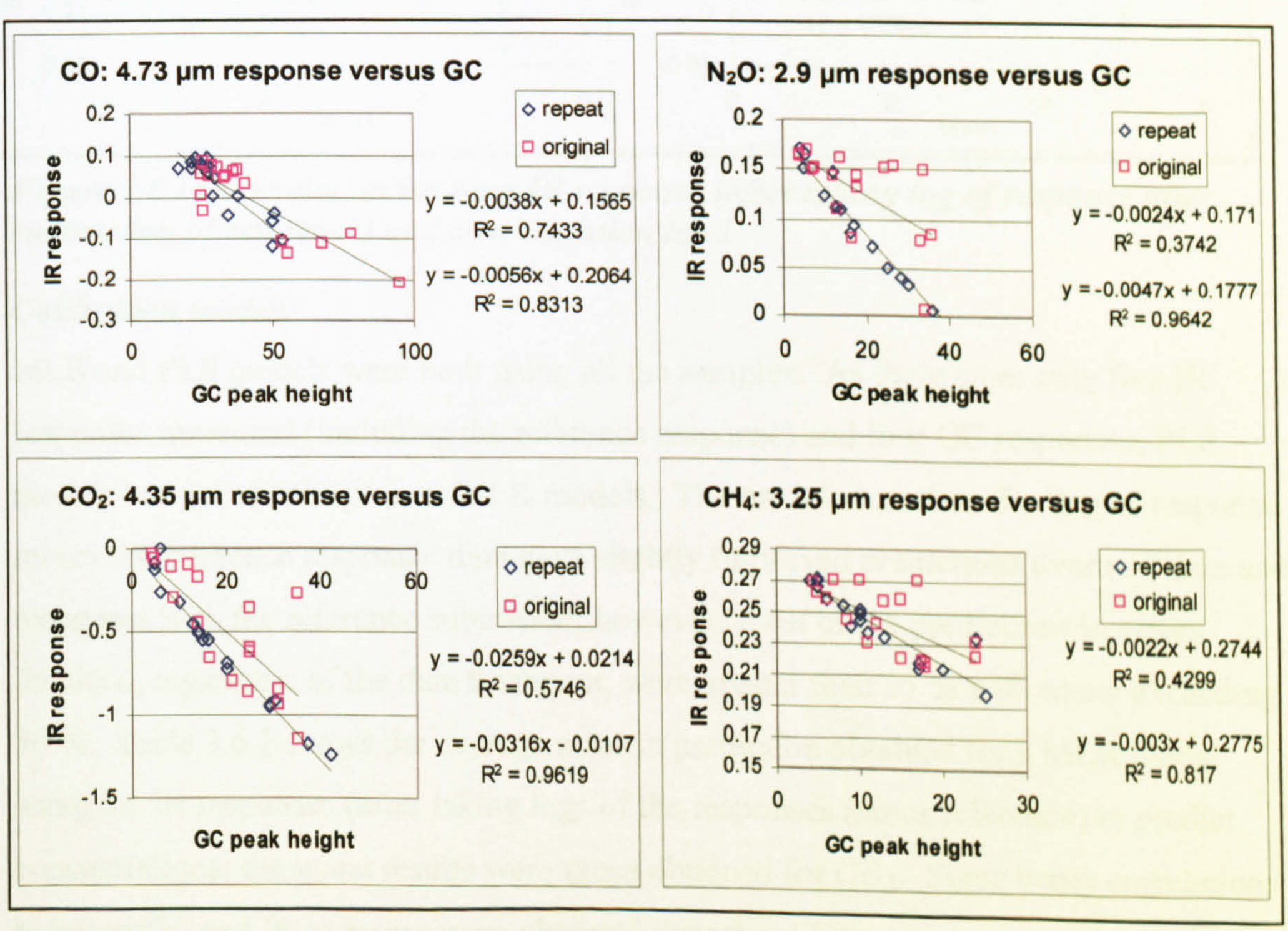


Figure 3.6.14 Correlation between GC peak height and IR response (log of response after subtraction of reference) for sample concentrations prepared from a five level experimental design

From these plots it can be seen there is very poor correlation for the CO and CH_4 and there is a wide deviation of response. There is clearly some overlap between the responses obtained for different concentration levels, indicating a problem with sensitivity.

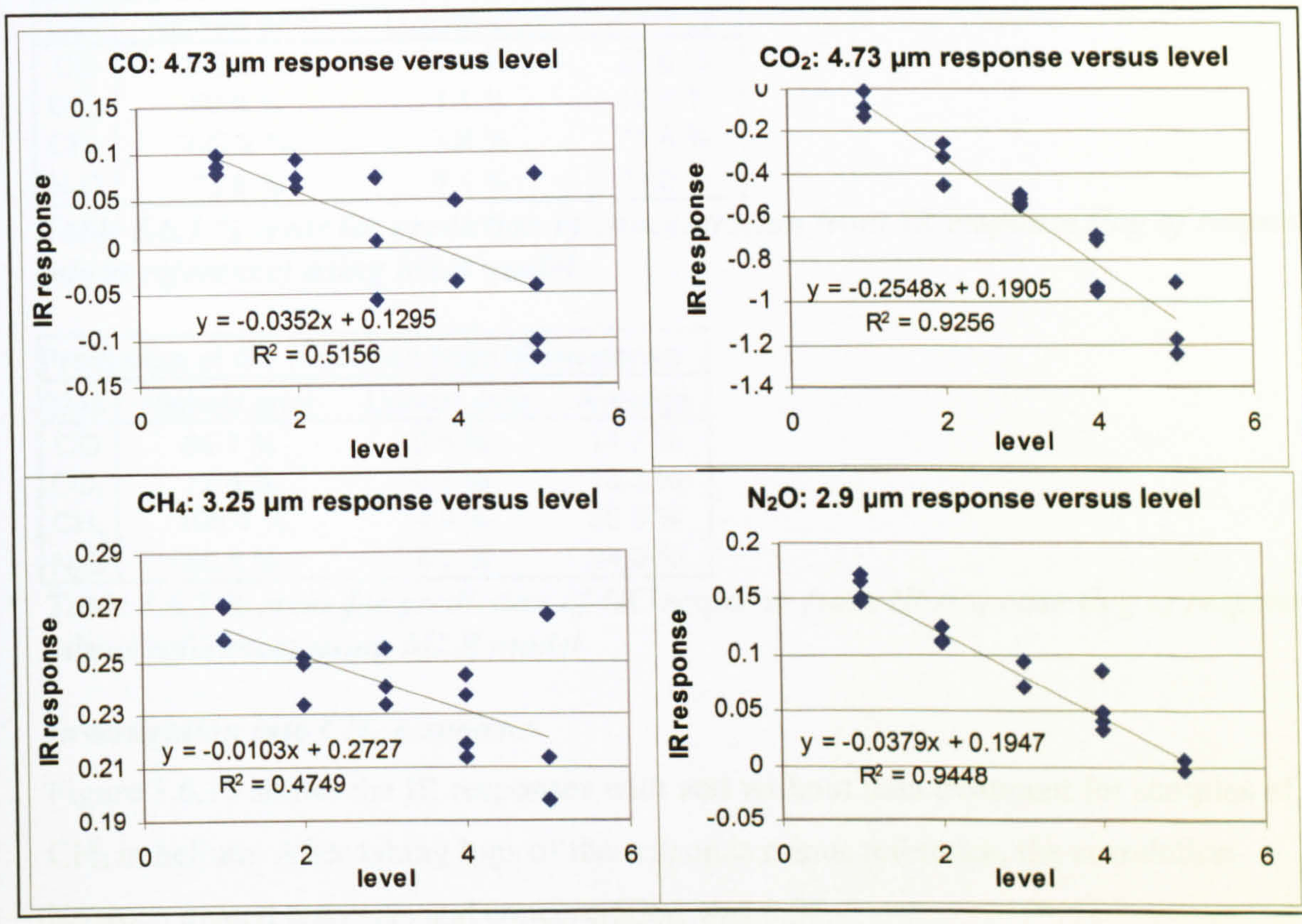


Figure 3.6.15 Correlation between IR response (after taking log of response after subtraction of reference) and concentration level

Calibration models

MLR and PLS models were built using all the samples. As there were only five IR responses measured (including the reference response) and four GC responses, PLS models had no advantage over MLR models. The models based on the 'log of response minus the reference response' data gave slightly improved predictions over raw data and responses with the reference subtracted; however, most of the predictions in errors obtained, regardless of the data treatment, were greater than 25 % with many exceeding 50 %. Table 3.6.1 shows the average error in prediction obtained for a MLR model using the IR responses (after taking logs of the responses minus reference) to predict concentrations; the worst results were those obtained for CH_4. Since better correlations between GC and IR response were obtained than those between concentration and IR response, models were also built to predict the GC response from the IR data. Figure 3.6.2 shows the average error in prediction obtained for a MLR model using the IR

responses (after taking log of response minus reference) to predict GC responses; the average % error in prediction was improved for CO and CH_4, but worse for CO_2 and N_2O, whilst the error in prediction for CH_4 was still significantly higher than the others.

Prediction of concentration from IR response			
Gas	Highest error	Lowest error	Average
CO	102.6 %	0.8 %	22.6 %
CO_2	50.9 %	1.1 %	12.6 %
CH_4	222.2 %	7.8 %	71.4 %
N_2O	73.6 %	1.1 %	21.3 %

Table 3.6.1 % error for prediction of concentration from IR response (log of response minus reference) using MLR model

Prediction of GC response from IR response			
Gas	Highest error	Lowest error	Average
CO	35.1 %	0.5 %	14.7 %
CO_2	72.4 %	0.3 %	13.3 %
CH_4	196.0 %	4.9 %	56.9 %
N_2O	88.8 %	4.2 %	24.9 %

Table 3.6.2 % error for prediction of GC response from IR response (log of response minus reference) using MLR model

Investigation into CH_4 responses

Figure 3.6.15 shows the IR responses with and without data treatment for samples of CH_4 in helium. After taking logs of the response minus reference, the correlation between treated response and concentration was 0.95.

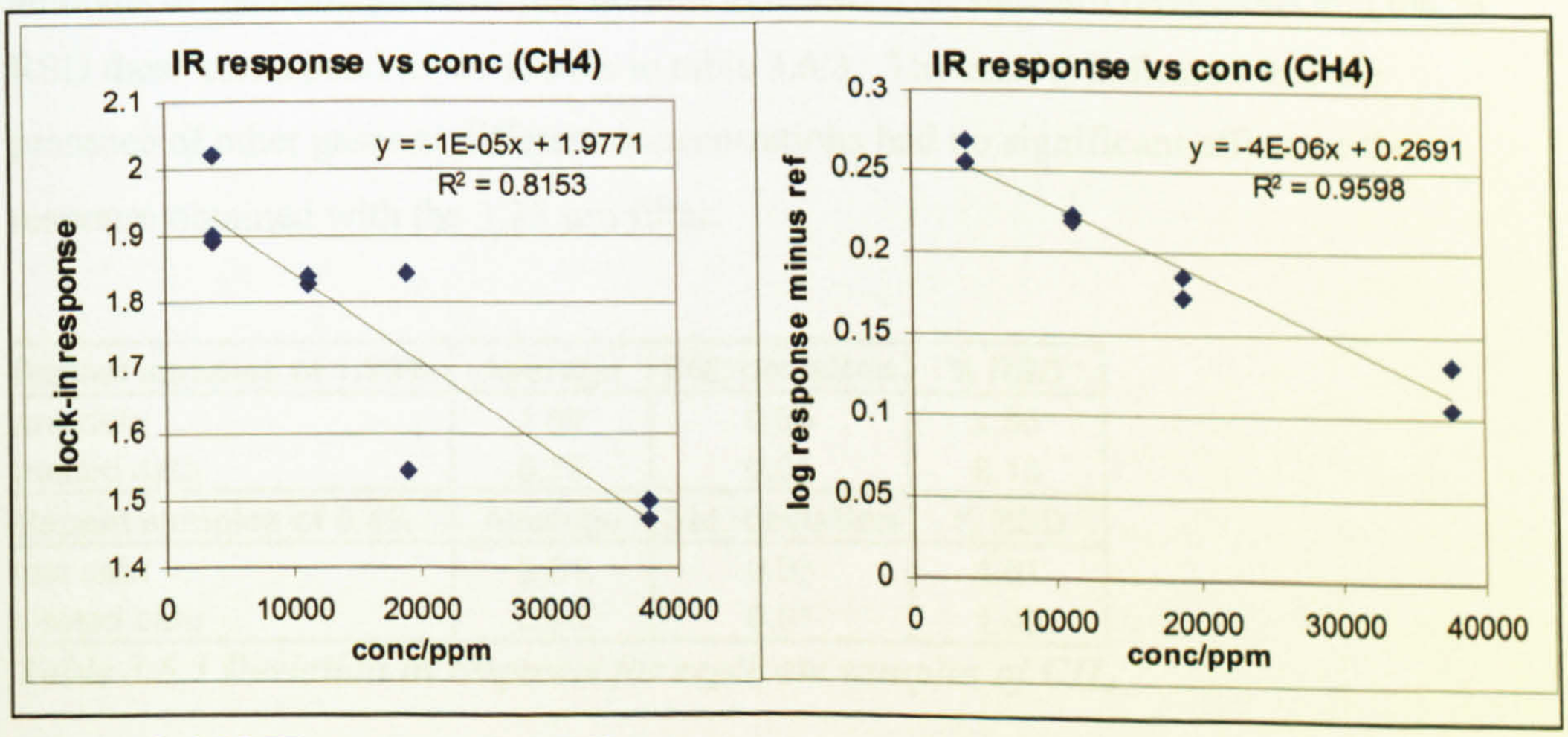


Figure 3.6.15 IR response for different concentrations of CH_4

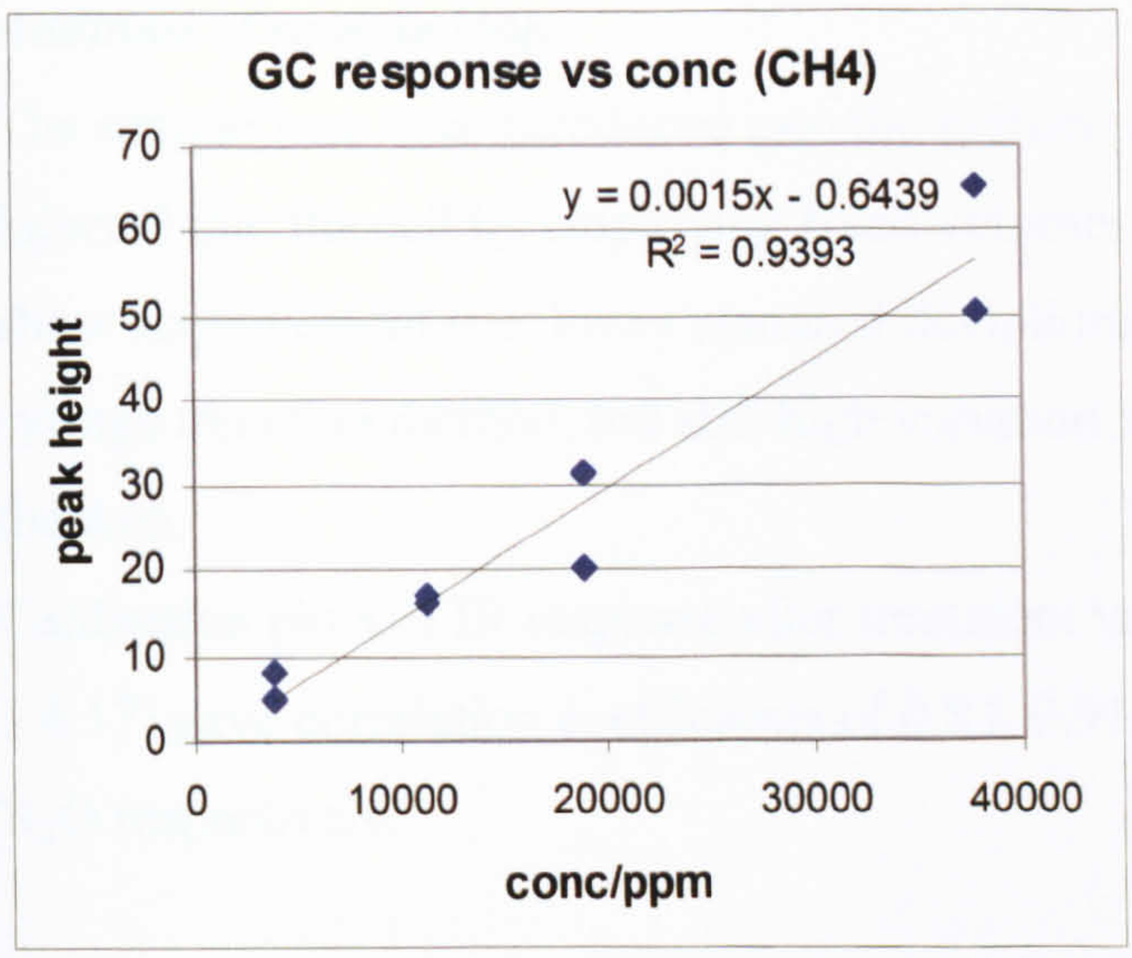


Figure 3.6.16 GC responses for different concentrations of CH_4

The correlation between the data shown in figure 3.6.15 (GC versus IR response for different concentrations of CH_4) is vastly improved over the previous IR response versus concentration of CH_4 in different sample mixtures (figure 3.4.14). This suggests that in previous work (section 3.6.4) there was either a problem with interference from the other gases or experimental error.

Reproducibility of response

To further check reproducibility of CH_4 responses, repeat sample of 1.89 % (used for level 5) and 0.4 % (employed for level 1) concentrations were prepared with varying amounts of the other gases or just helium injected. The standard deviations and the % RSD these correspond to are shown in table 3.6.3. The results indicated that the presence of other gases at different concentrations had no significant effect on the response obtained with the 3.25 μm filter.

Repeat samples of 1.89%	**Average**	**Std. deviation**	**% RSD**
raw data	1.69	0.05	2.86
treated data	0.18	0.01	8.13
Repeat samples of 0.4%	**Average**	**Std. deviation**	**% RSD**
raw data	2.01	0.03	1.61
treated data	0.26	0.01	1.99

Table 3.6.3 Deviation in response for replicate samples of CH_4

Addition of sample loop

The sample loop was introduced into the system to reduce variation in the amounts injected into the cell by employing fixed volumes (as opposed to a syringe). The results show improvement (i.e. lower standard deviations between repeat injections) over the syringe injection method, but still high variation in the model due to inherent error in the data.

Calibration plots of IR response after treatment versus concentration levels (figure 3.6.17) gave correlation coefficients of 0.83, 0.91, 0.91 and 0.92 for CO_2, CO, CH_4 and N_2O respectively.

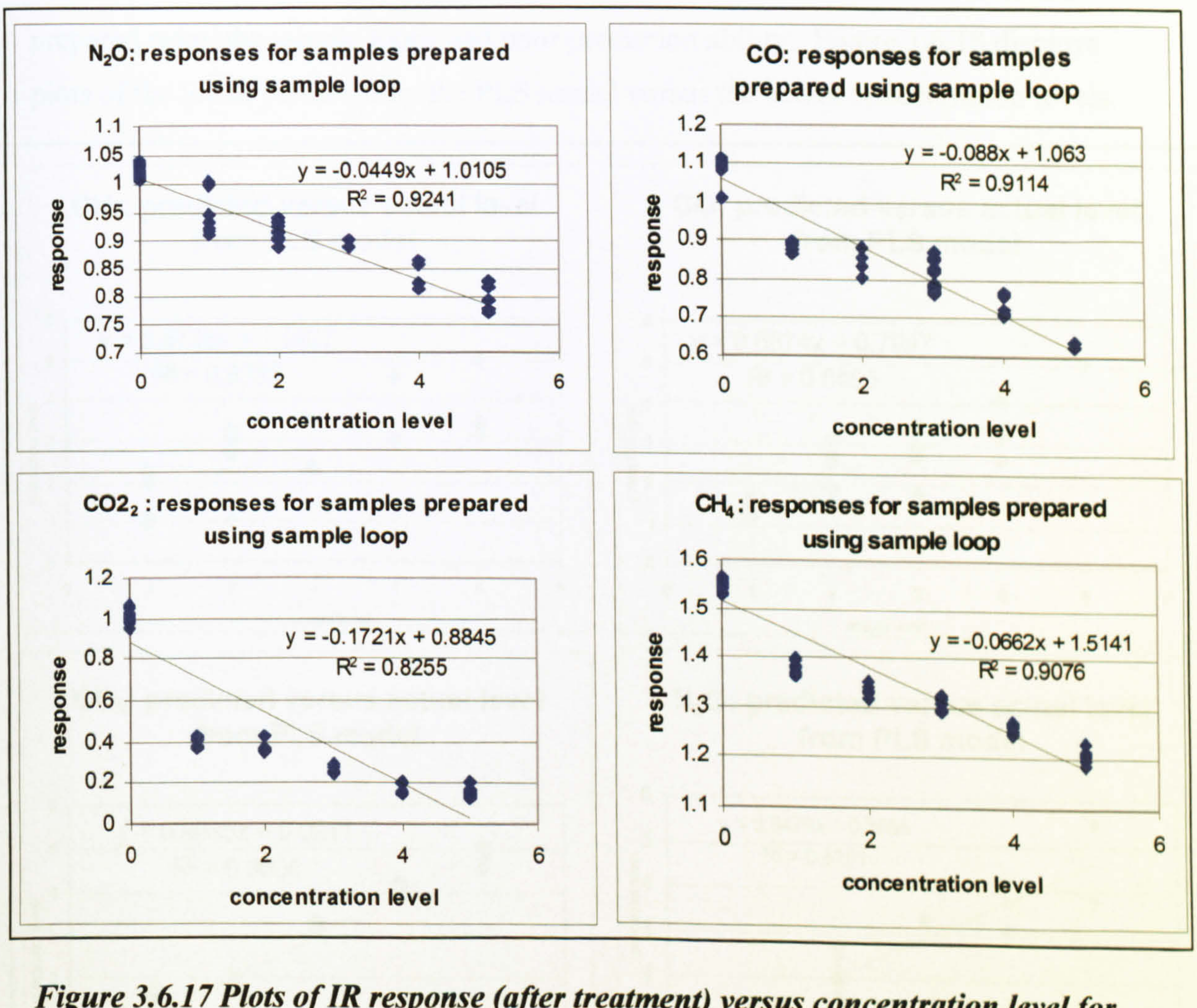


Figure 3.6.17 Plots of IR response (after treatment) versus concentration level for samples prepared using sample loop

Table 3.6.4 lists the average response, giving the standard deviation, and % RSD obtained for each concentration level.

	Average response				Standard deviation				% RSD			
level	CO_2	CO	CH_4	N_2O	CO_2	CO	CH_4	N_2O	CO_2	CO	CH_4	N_2O
1	0.41	0.88	1.38	0.98	0.03	0.02	0.01	0.04	6.99	1.69	1.04	4.49
2	0.36	0.85	1.34	0.91	0.01	0.03	0.01	0.02	3.26	3.88	0.79	1.72
3	0.26	0.79	1.31	0.90	0.02	0.03	0.01	-	8.46	3.52	1.03	-
4	0.18	0.72	1.26	0.83	0.04	0.01	0.01	0.03	20.47	0.88	0.65	3.14
5	0.16	0.63	1.21	0.78	0.01	0.01	0.02	0.02	4.26	1.89	1.64	2.09

Table 3.6.4 Average response, standard deviation and % RSD for filter responses (2.9 μm filter for N_2O, 3.25 μm filter for CH_4, 3.45 μm filter for CO and 4.73 μm filter for CO) after data treatment (log of response minus reference)

PLS modelling

With the exception of CO_2, the PLS models built from the data collected for samples prepared using the sample loops had poor prediction ability. Figure 3.6.18 displays plots of the levels predicted by the PLS model versus the actual concentration levels.

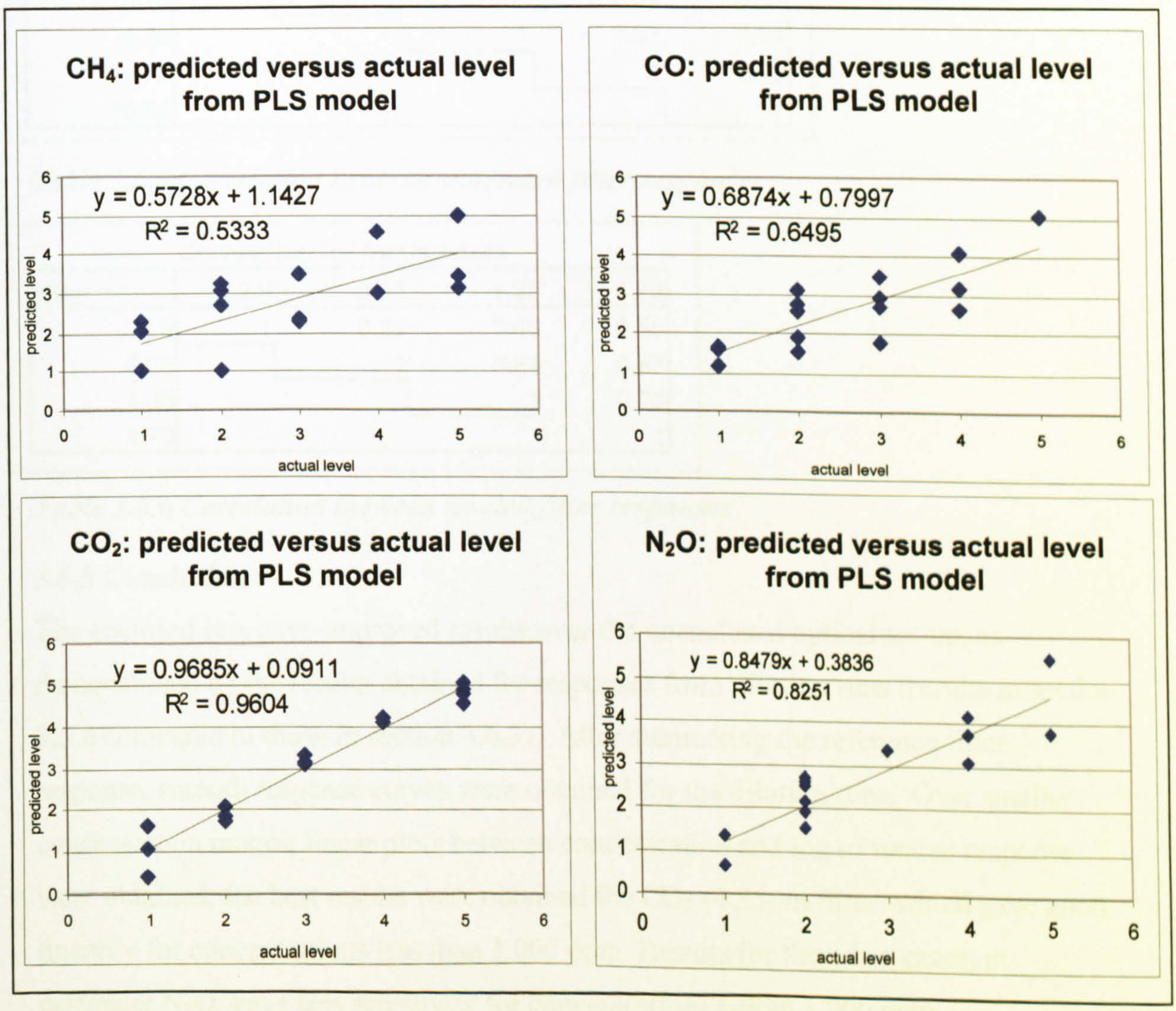


Figure 3.6.18 Predicted versus actual levels for PLS models

The main reason for the poor predictions is due to the overlap in magnitude in response between the responses for different concentration levels resulting in a lack of sensitivity to the changes in concentration.

Correlation between filter responses

Tables 3.6.5 and 3.6.6 show the correlation between different filter responses. These show a higher than expected correlation between different filter responses even after subtraction of the reference response, thus implying there are still problems with the detection system, most likely drift in the response that is not fully compensated for by the data treatment.

Correlation for untreated data

filter	2.9	3.25	4.35	4.73	10.62
2.9	1	0.74	0.76	0.75	0.69
3.25		1	0.98	0.99	0.97
4.35			1	0.97	0.94
4.73				1	0.99
10.62					1

Table 3.6.5 Correlation between untreated filter responses

Correlation for treated data

filter	2.9	3.25	4.35	4.73
2.9	1	0.39	0.46	0.55
3.25		1	0.84	0.87
4.35			1	0.73
4.73				1

Table 3.6.6 Correlation between treated filter responses

3.6.5 Conclusion

The enclosed box gave improved results over the unenclosed optical set-up, as demonstrated by the results obtained for responses from dilution runs (results in section 3.5.6 compared to those in section 3.6.3). After subtracting the reference filter response, smooth response curves were obtained for the dilution runs. Over smaller concentration ranges, linear plots between concentration and log of treated response were obtained; the best results were obtained for CO_2 (4.35μm filter) which gave good linearity for concentrations less than 1,000 ppm. Results for the other gases, in particular N_2O, gave less sensitivity for concentrations below 1,000 ppm.

The next step of collecting responses from mixtures of different concentration combinations gave very poor correlations between concentrations and responses. The

poor correlation between GC and IR responses indicated this was caused by equipment problems as opposed to errors in sample preparation. Data collected from a repeat of the work gave improved correlation between the GC and IR responses, hence further analysis was carried out the data obtained from the repeat data.

From analysis on these data it was clear, by examining plots of the data, that in some cases there was a wide deviation in response for gases of the same concentration, for example the level 5 concentration responses for CO and CH_4 (figure 3.6.15) obtained with filters 2.9 and 3.25 μm respectively.

To address this, further investigations were performed on the reproducibility of CH_4 responses. The correlation obtained between concentration and response from these data was lower than that obtained for the previous data. This suggested either interference from the other gases or experimental error. Further work assessing the reproducibility of CH_4 responses showed that the addition of other gases to the sample made no difference to the response, thus ruling out the possibility of interference from the other gases.

To reduce error in the sample preparation technique a sampling valve was introduced into the system to ensure only fixed volumes of gas were introduced into the system (as opposed to varying amount from a gas syringe). This improved the correlation between concentration level and response, with R^2 values of 0.94, 0.91, 0.95 and 0.87 for CO_2, CO, CH_4 and N_2O respectively; however, responses for different concentration levels still lacked sensitivity to the different concentrations. Deviations in response at the same concentration levels resulted in overlap in the magnitude of response between different concentration levels. As a result, the PLS model built from this data still had high prediction errors.

A study of correlations between filter responses was conducted to determine if the variation in filter responses were independent of each other or linked. Without the data treatment, high correlations were obtained between all the filters due to the effect of detector drift and source variations i.e. a drop in detector response causing a drop in all the filter responses. With data treatment the effect of these variations should be removed leaving just random error and there should be little or no correlation between the different filters. Correlations between filter responses after data treatment were higher than expected between 4.35μm and 3.25μm, and between 4.73μm and 3.25μm which implied the data treatment was not fully compensating for drift in detector/ system response.

A main problem with the system was that most of the equipment was borrowed from other instruments therefore was not specifically tailored for this set-up. Since the equipment was borrowed it had to be shared with other research projects, thus resulting in the chopper and occasionally the detector being removed in between measurement sessions, which could have also attributed to the reproducibility problems.
The mid IR and narrowband filters will eventually be replaced with QWIP detectors, which should simplify the system (less components to align), and therefore reduce errors in the system. Without the filters there should be a higher % of light reaching the detector which coupled with the expected higher sensitivity of the QWIP detectors should improve the detection limits. There is also the possibility that the lock-in amplier and chopper may be removed from the set-up which would further simplify the system.

3.7 Results and Discussion for Vapour Analysis

The aim of this stage of the work was to look at the feasibility of developing a QWIP IR detector array for the qualification and possible quantification of different organic vapour mixtures. This was investigated by first collecting spectra for different organic compounds and investigating the similarity by cluster analysis, before performing variable selection on a few simple mixtures.

3.7.1 Spectra collected by sampling vapour with syringe

In order to obtain spectra, of a suitable format, for a variety of compounds, samples were collected and scanned using the same (Excaliber) FTIR instrument. Samples were collected using a syringe to sample the air, along with any vapour present, from bottles containing the compounds of interest. Figure 3.7.1 displays some of the absorption spectra obtained from different compounds, 'noise' refers to samples of ambient air (mainly consisting of CO_2 and water vapour).

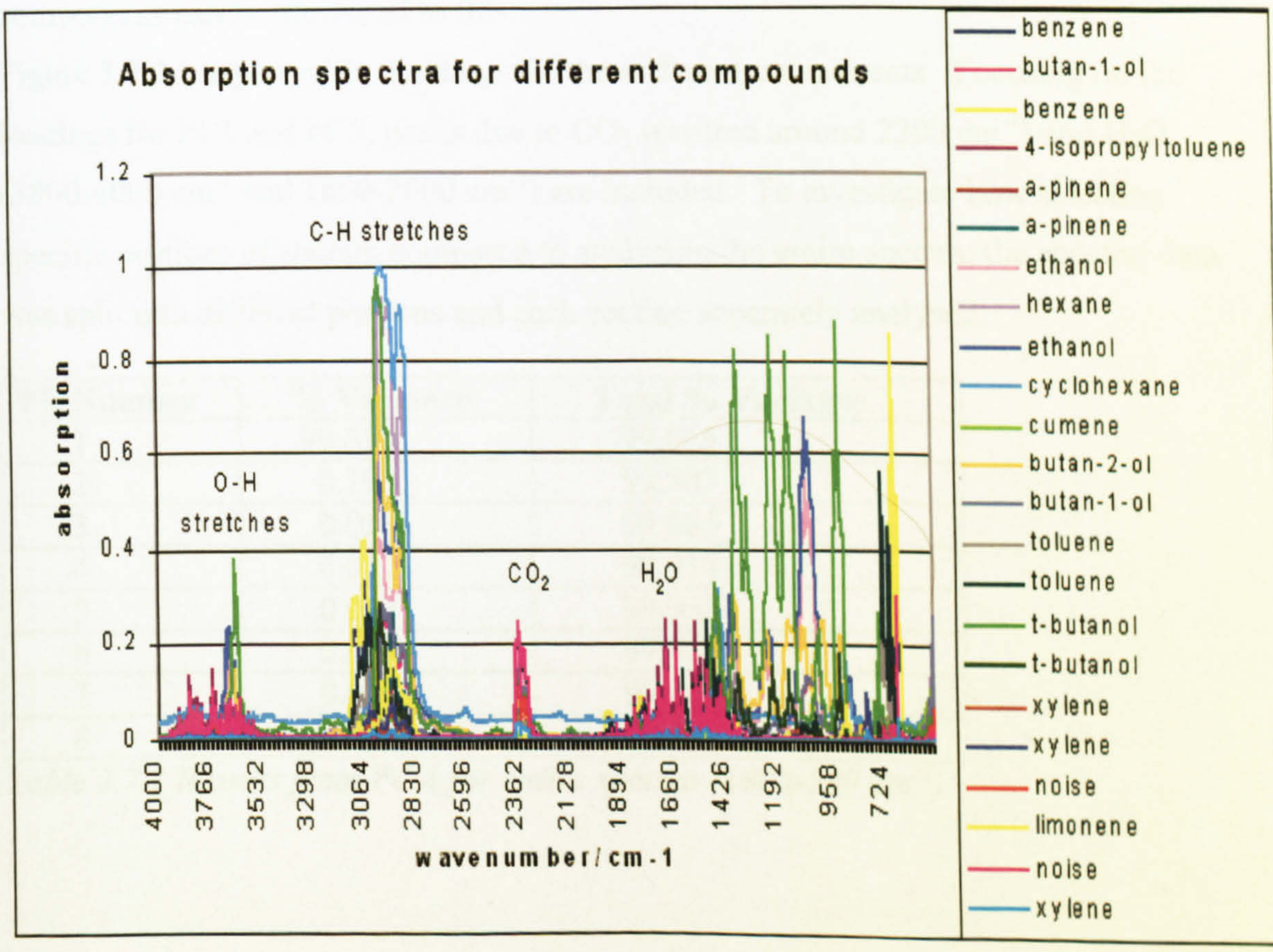


Figure 3.7.1 Absorption spectra obtained by vapour sampling with syringe

Absorption peaks due to CO_2 and H_2O were visible (to a varying extent) in all the spectra; the spectra obtained for each of the different compounds were compared to reference spectra (from NIST[64] and Aldrich[97]) thus confirming the remaining absorption peaks were due to the compounds sampled.

3.7.1.1 Principal Component Analysis (PCA)

PCA on entire spectra

PCA was performed on the entire spectra; table 3.7.1 lists the amount of variance described by each principal component. In this case over 99 % of the variance was described by the first principal component (PC1) and very little information can be abstracted from the remaining variance. Score plots of the principal components showed most of the samples clustered around the origin with one or two outliers. As mentioned earlier all the spectra contained absorption peaks, of varying amounts, due to water vapour and CO_2, hence some of the variance described by the principal components can be attributed to this.

Figure 3.7.2 is a plot of the loadings for the different components. Focusing on the loadings for PC1 and PC3, peaks due to CO_2 (centred around 2200 cm^{-1}) and H_2O (3800-4000 cm^{-1} and 1800-2000 cm^{-1}) are included. To investigate how selecting specific portions of spectra compared to analysing the entire spectra, the spectral data was split into different portions and each section separately analysed.

PC Number	% Variance	Total % Variance
1	99.615	99.615
2	0.192	99.807
3	0.058	99.865
4	0.052	99.918
5	0.042	99.960
6	0.013	99.973
7	0.010	99.983
8	0.008	99.991

Table 3.7.1 Results from PCA for entire spectra (4000-500 cm^{-1})

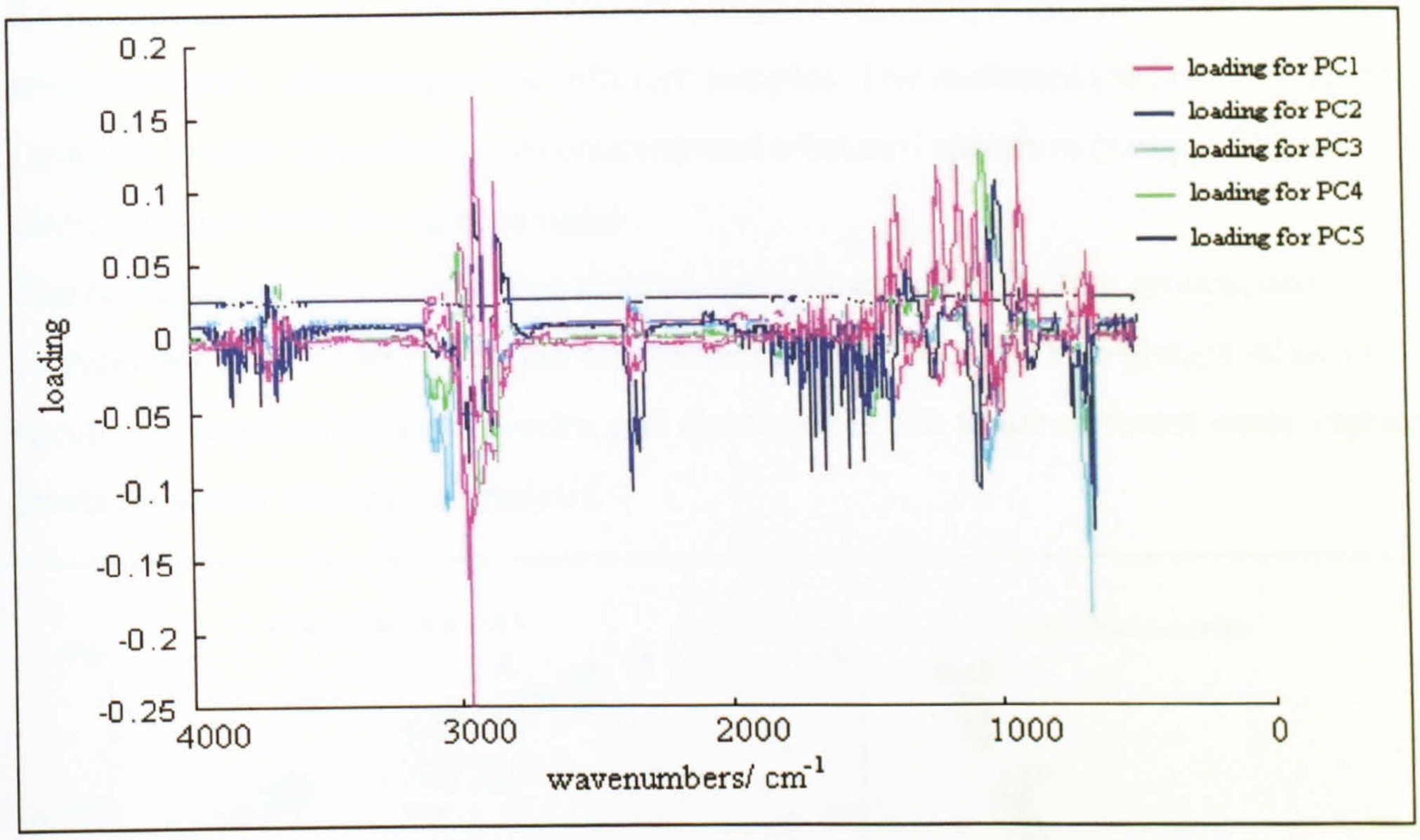


Figure 3.7.2 Loadings obtained from PCA performed on entire spectra

PCA on Region one (4000-3400 cm^{-1})

Region one is where O-H stretches occur therefore absorption peaks in this region are due to water vapour and O-H stretches on alcohols. PCA gave three components; PC1 described 99.961 %; PC2, 0.025 %; and PC3, 0.008 %. Referring to the plot of the loadings (figure 3.7.3) it is clear that PC1 describes mostly absorption due to OH group on the alcohols and PC2 contains mostly variation due to water vapour.

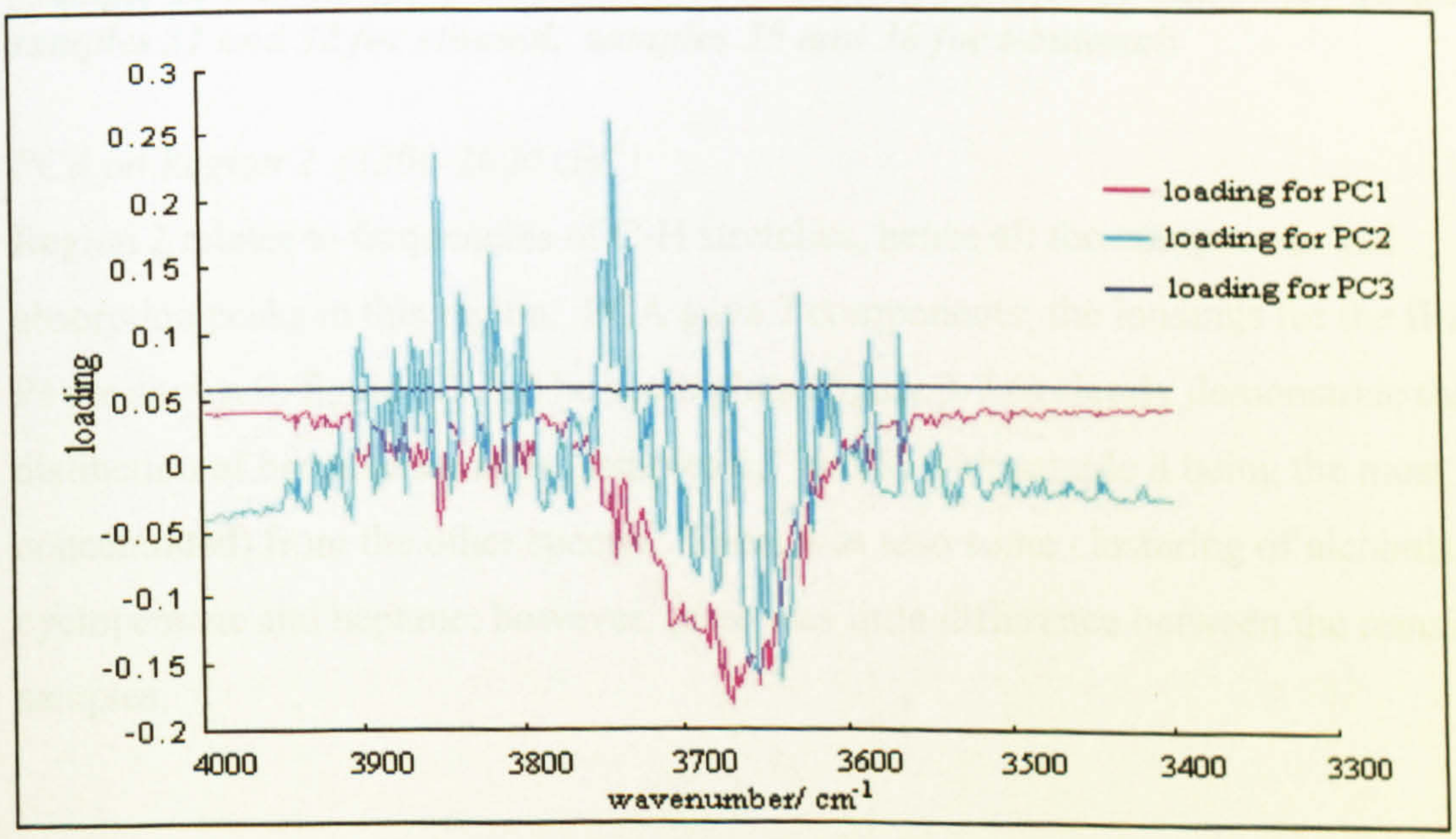


Figure 3.7.3 Loadings from PCA of vapour spectra using just wavenumbers 4000-3400 cm^{-1}

By plotting the sample scores of different components (refer to figure 3.7.4) it can be seen that there is clustering of the different samples. The methanol (sample 24), ethanol (samples 31 and 32) and the most concentrated t-butanol spectrum (sample 35) are clearly distinct from the other samples.

The remaining samples (excluding alcohol spectra) are split into two groups, one containing samples 1 to 15 and the other samples 16 to 34. The two groups relate to spectra collected on different weeks, and clustering is due to the different water vapour levels present at the time of analysis.

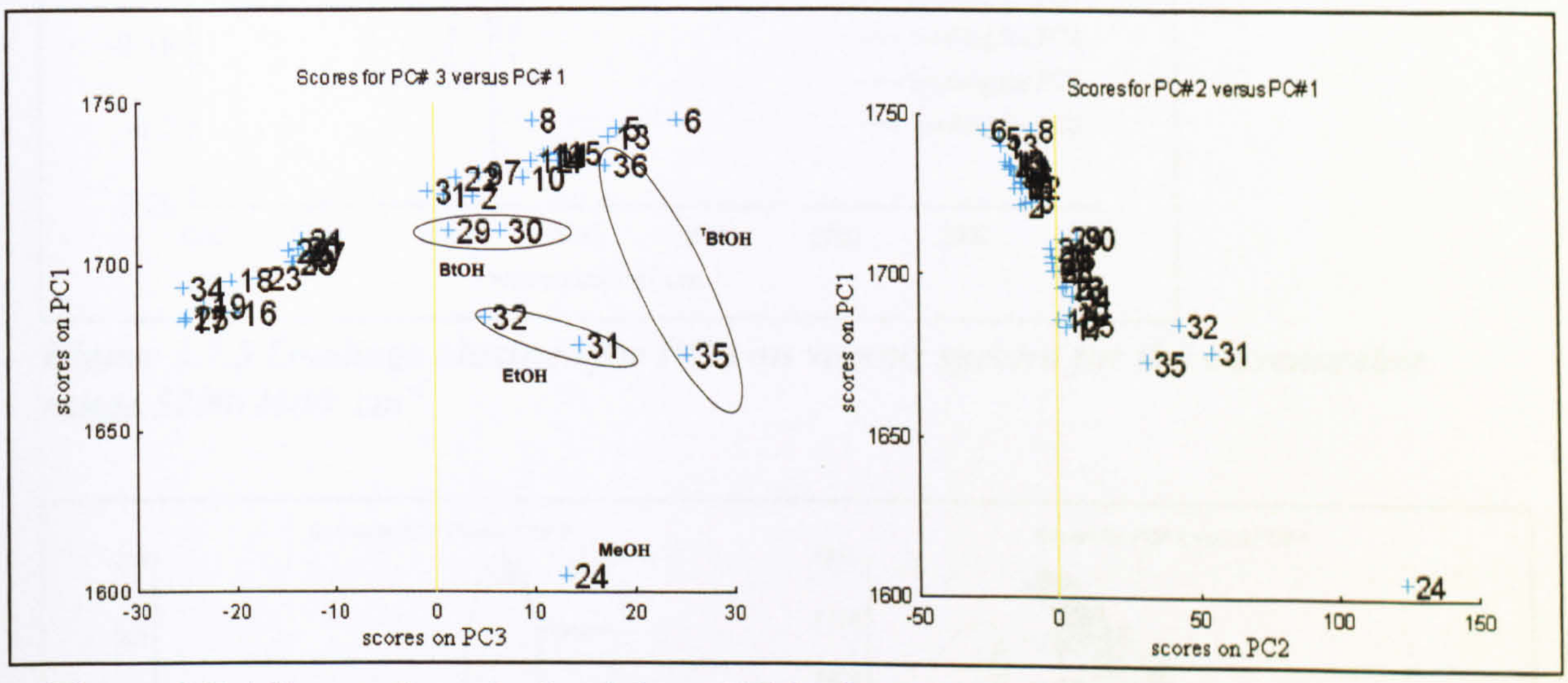


Figure 3.7.4 Score plots obtained from PCA of vapour spectra using just wavenumbers 4000-3400 cm^{-1}
(Sample 24 was the spectra obtained for methanol, samples 29 and 30 for butan-1-ol, samples 31 and 32 for ethanol, samples 35 and 36 for t-butanol)

PCA on Region 2 (3200-2600 cm^{-1})

Region 2 relates to frequencies of C-H stretches, hence all the compounds had absorption peaks in this region. PCA gave 7 components; the loadings for the first five Ps are shown in figure 3.7.5. The score plots (figure 3.7.6) clearly demonstrate the distinction of benzene samples (samples 6,7 and 8, with sample 8 being the most concentrated) from the other spectra. There was also some clustering of alcohols, cyclopentane and heptane; however, there was little difference between the remaining samples.

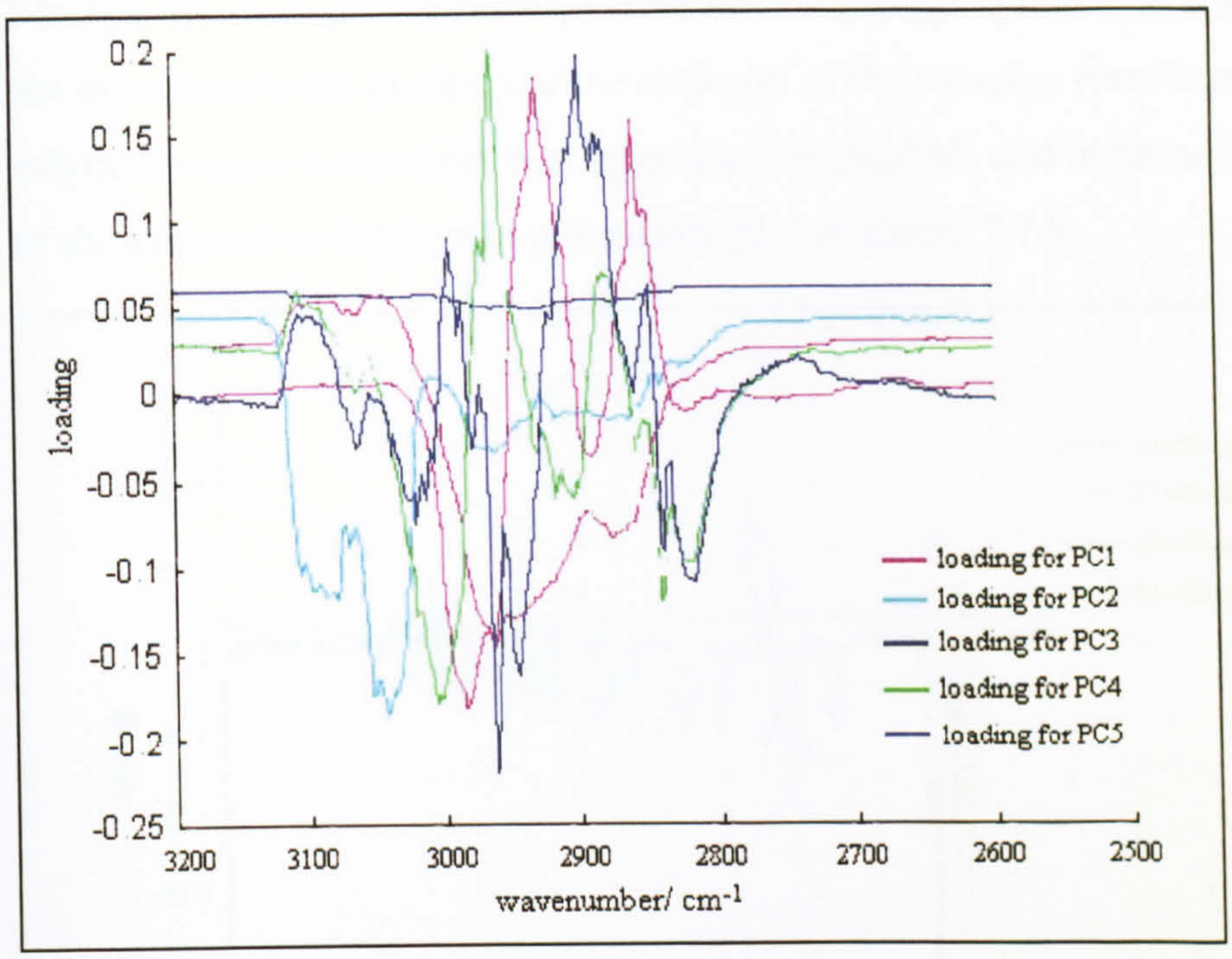


Figure 3.7.5 Loadings obtained for PCA on vapour spectra for the wavenumber range 3200-2600 cm^{-1}

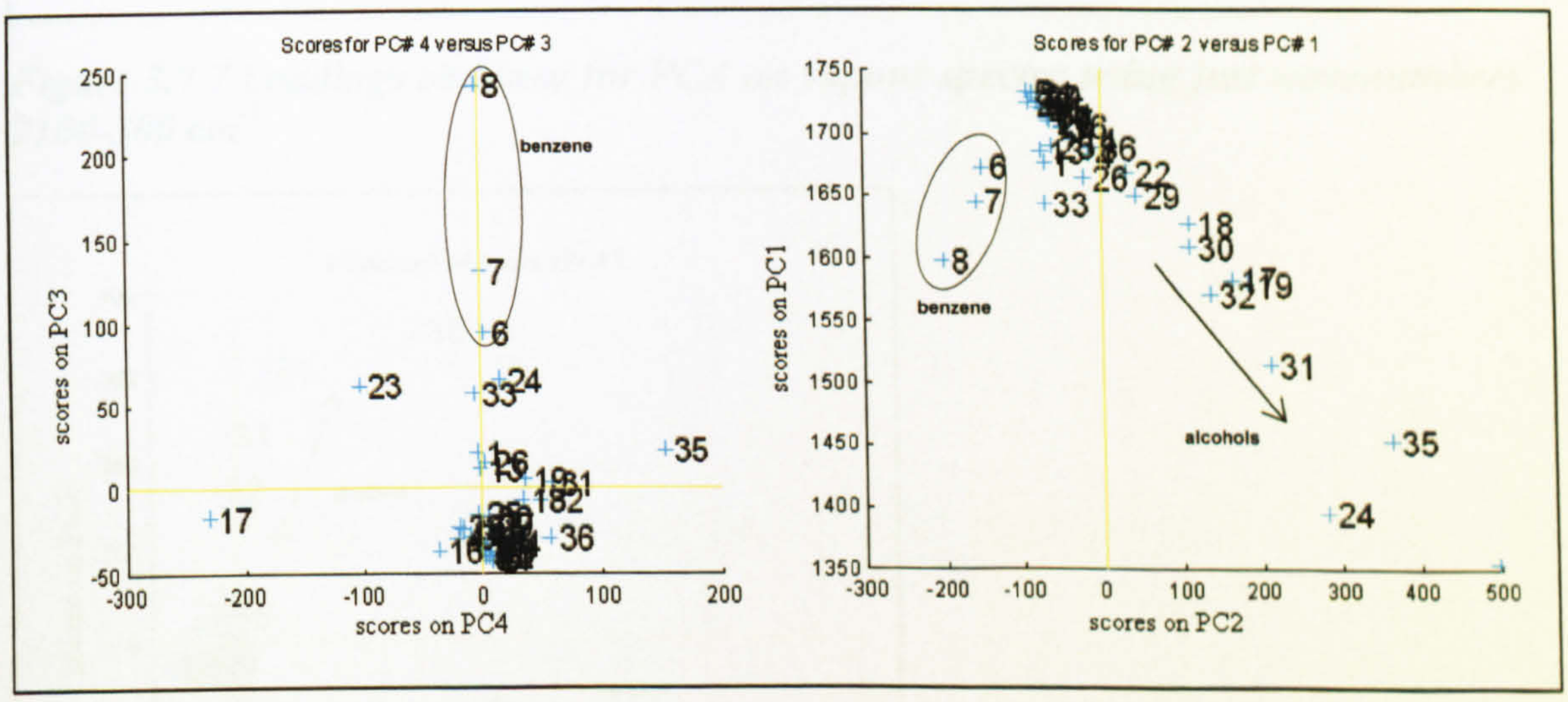


Figure 3.7.6 Score plots obtained from PCA on vapour spectra for region 3200-2600 cm^{-1}

Region 3 (2100-500 cm^{-1})

Region three also gave 7 principal components, this region is where characteristic C-C, C-H and C-O stretches occur, as well as H-O stretches (from water).

Loadings for the first four PCs are displayed in figure 3.7.7.

Most of the score plots gave little evidence of clustering between compounds containing the same functional groups and the majority of the samples were centred around the origin. In some cases there was separation of alcohols and benzenefrom the main bulk of the samples as displayed by the score plot in figure 3.7.8.

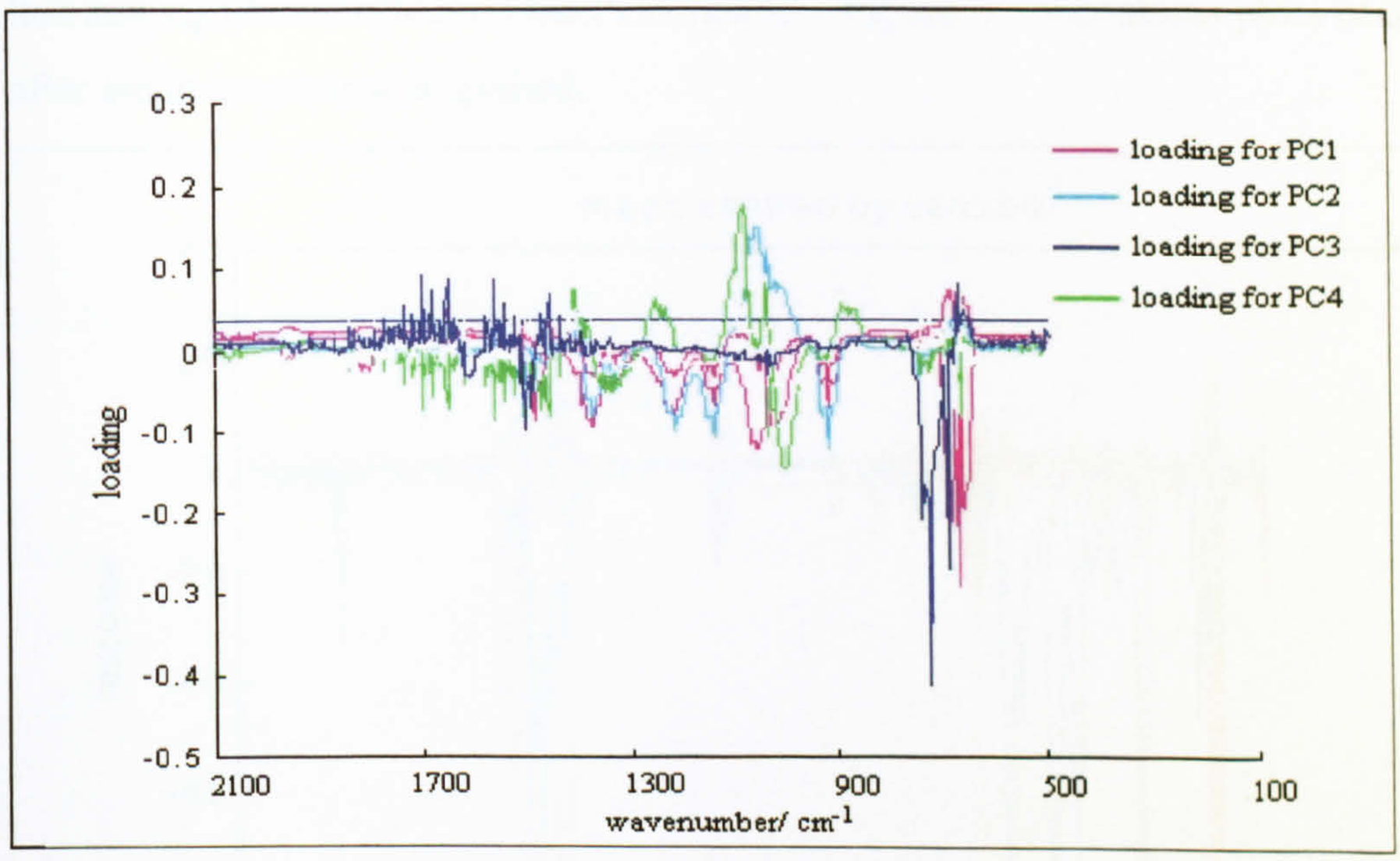


Figure 3.7.7 Loadings obtained for PCA on vapour spectra using just wavenumbers 2100-500 cm^{-1}

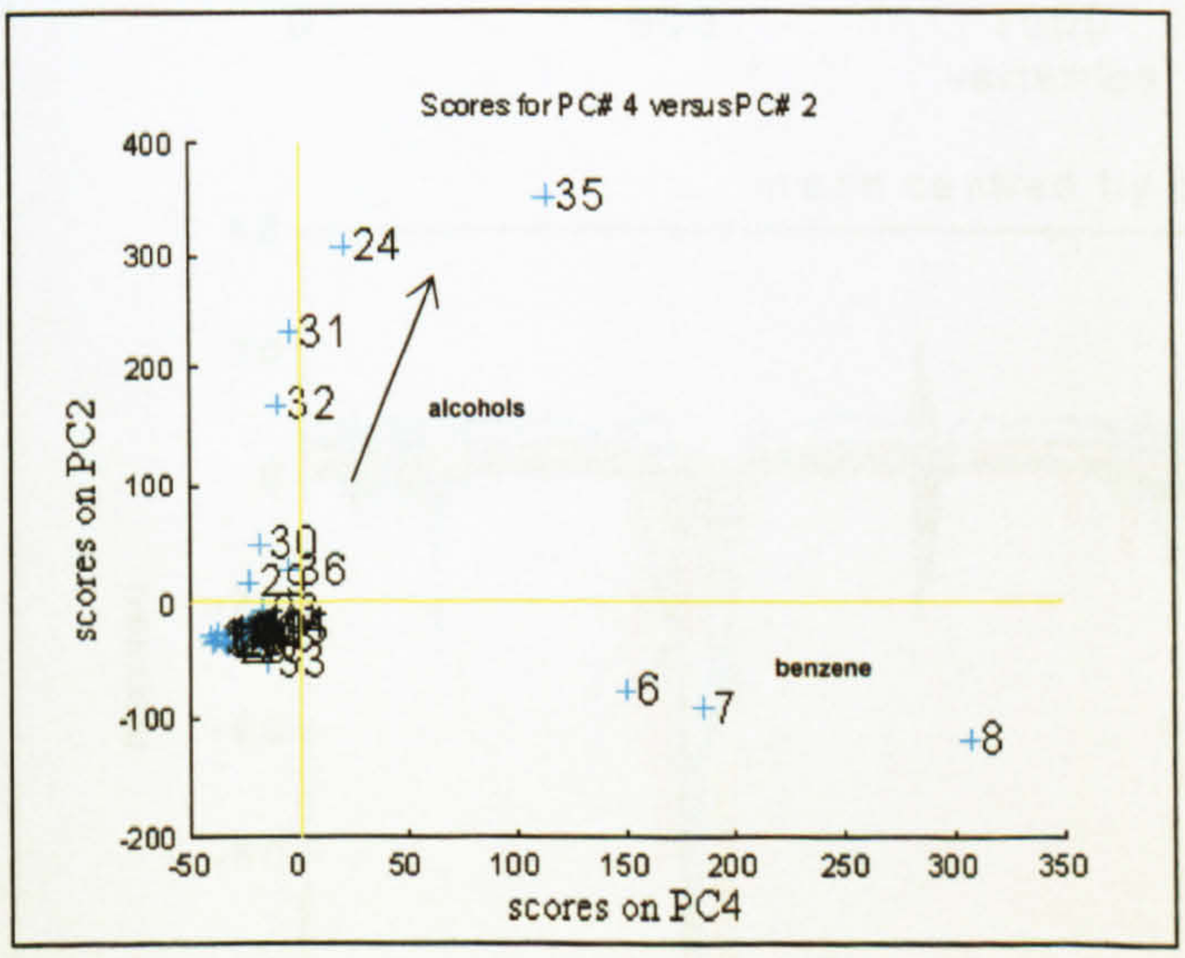


Figure 3.7.8 Score plot obtained from PCA on vapour spectra for region 2100-500 cm^{-1}

Region 4 (1300-500 cm^{-1})

Region 4 corresponds to the region where characteristic C-C, C-H and C-O stretches occur (as in the case of region 3, but avoiding the water stretches in the region 2100-

1300 cm^{-1}). PCA resulted in seven principal components and similar score plots wre obtained that were similar to previous region 3.

3.7.1.2 Mean centring pre-treatment of data

Mean centring by sample and by variable was applied to see whether data pre-treatment had any significant effect on the PCA results. Figure 3.7.9 contains plots of the spectra after mean centring was applied.

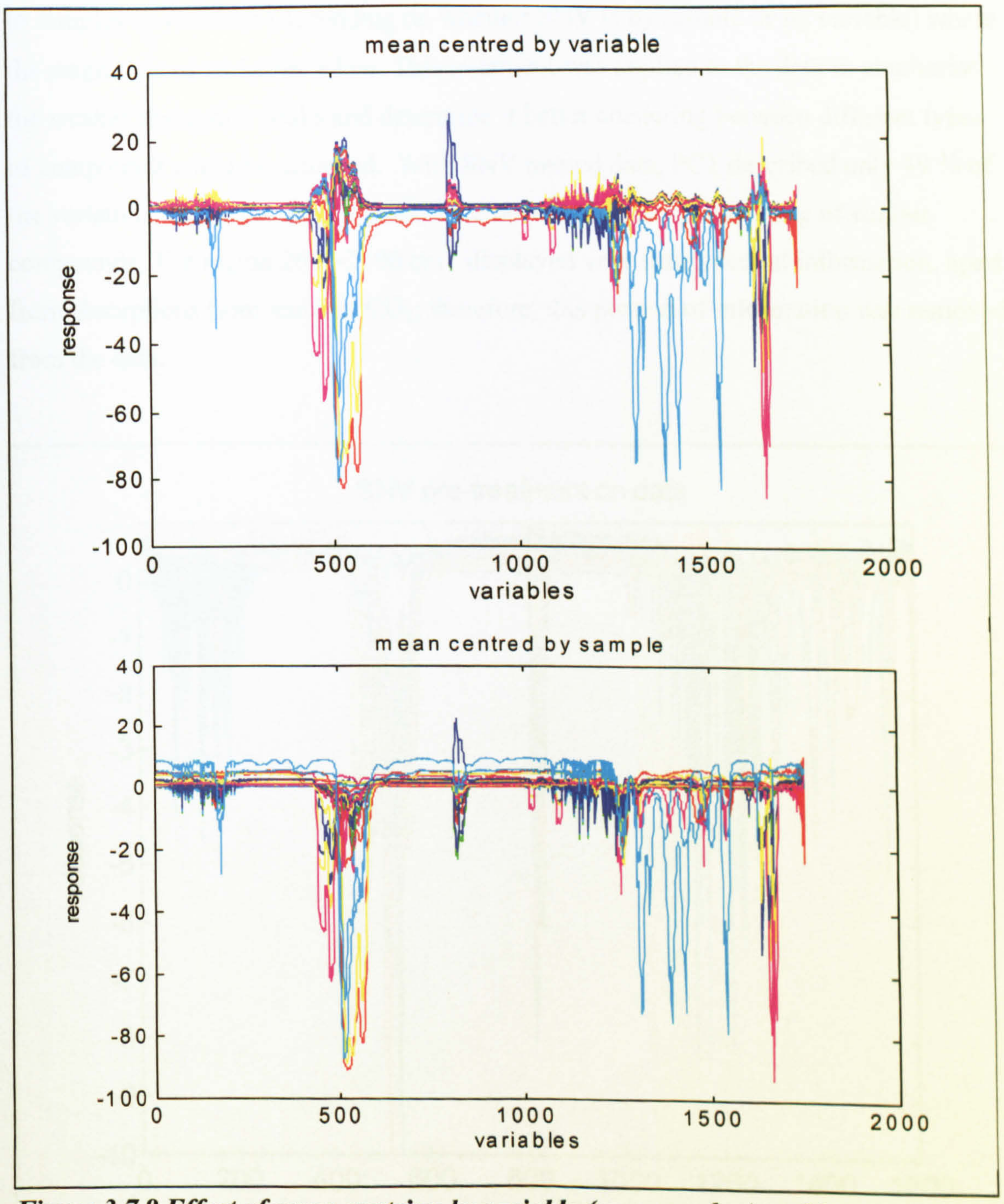


Figure 3.7.9 Effect of mean centring by variable (wavenumber) and by sample on vapour spectra

As with the analysis on the untreated spectra, PC1 described the majority of variation (over 80 %) and the remaining components gave little information to distinguish spectra of different compounds; thus indicating that mean centring offered no significant improvement over untreated spectra.

3.7.1.3 Standard Normal Variate pre-treatment of data

In SNV treatment data is divided by the standard deviation, thus giving added weighting to samples or variables (depending on whether SNV is by sample or by variable) where the magnitude of variation is low. This treatment was applied to the data to emphasise the weaker absorption peaks and determine if better clustering between different types of compounds could be achieved. With SNV treated data, PC1 described only 49 % of the variation. Score plots of PC1 and PC2 now show clearer clustering of similar compounds. The region 2600-2100 cm^{-1} displayed very little spectral information, apart from absorptions from ambient CO_2; therefore, this portion of information was removed from the data.

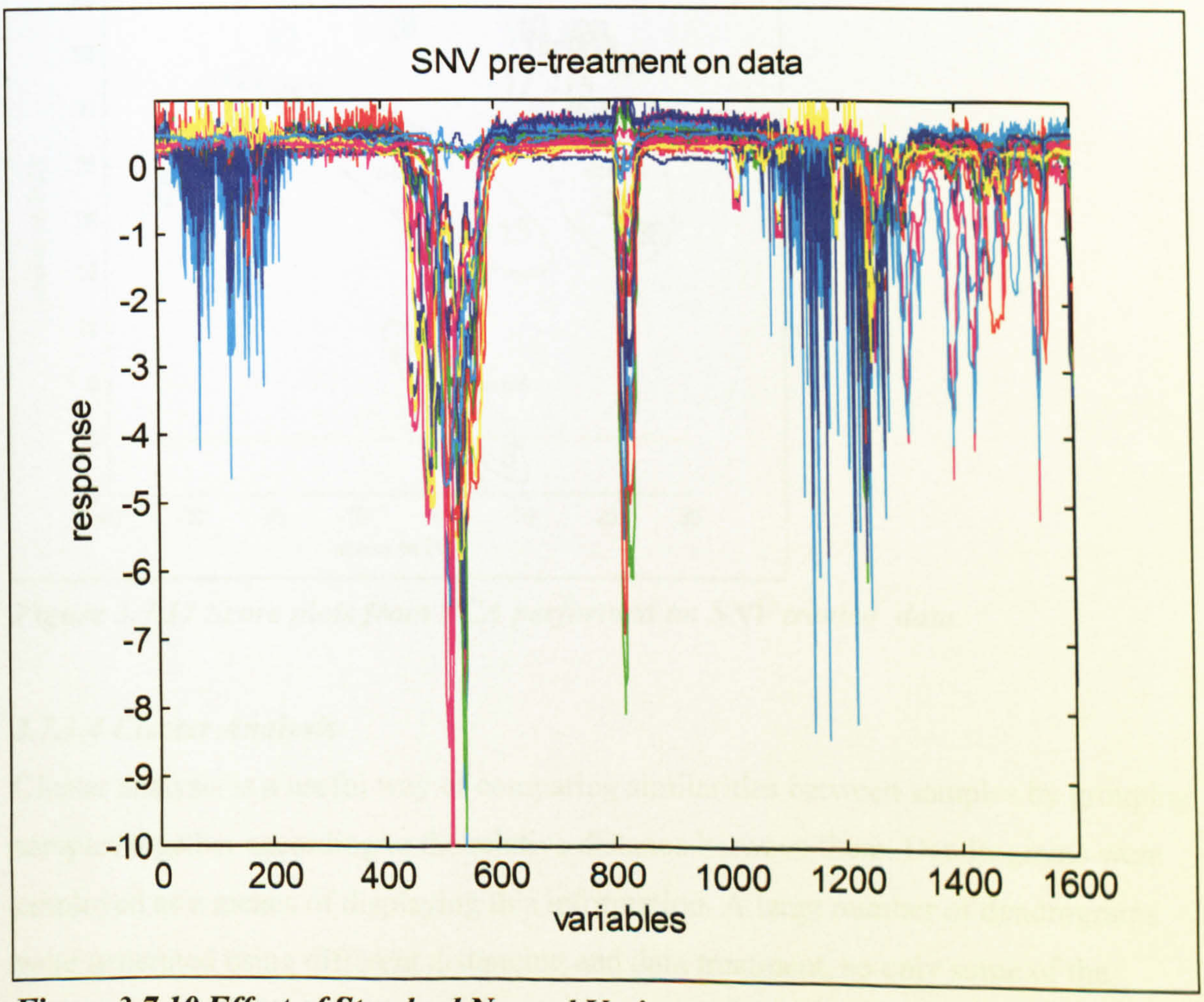


Figure 3.7.10 Effect of Standard Normal Variate treatment on Spectra

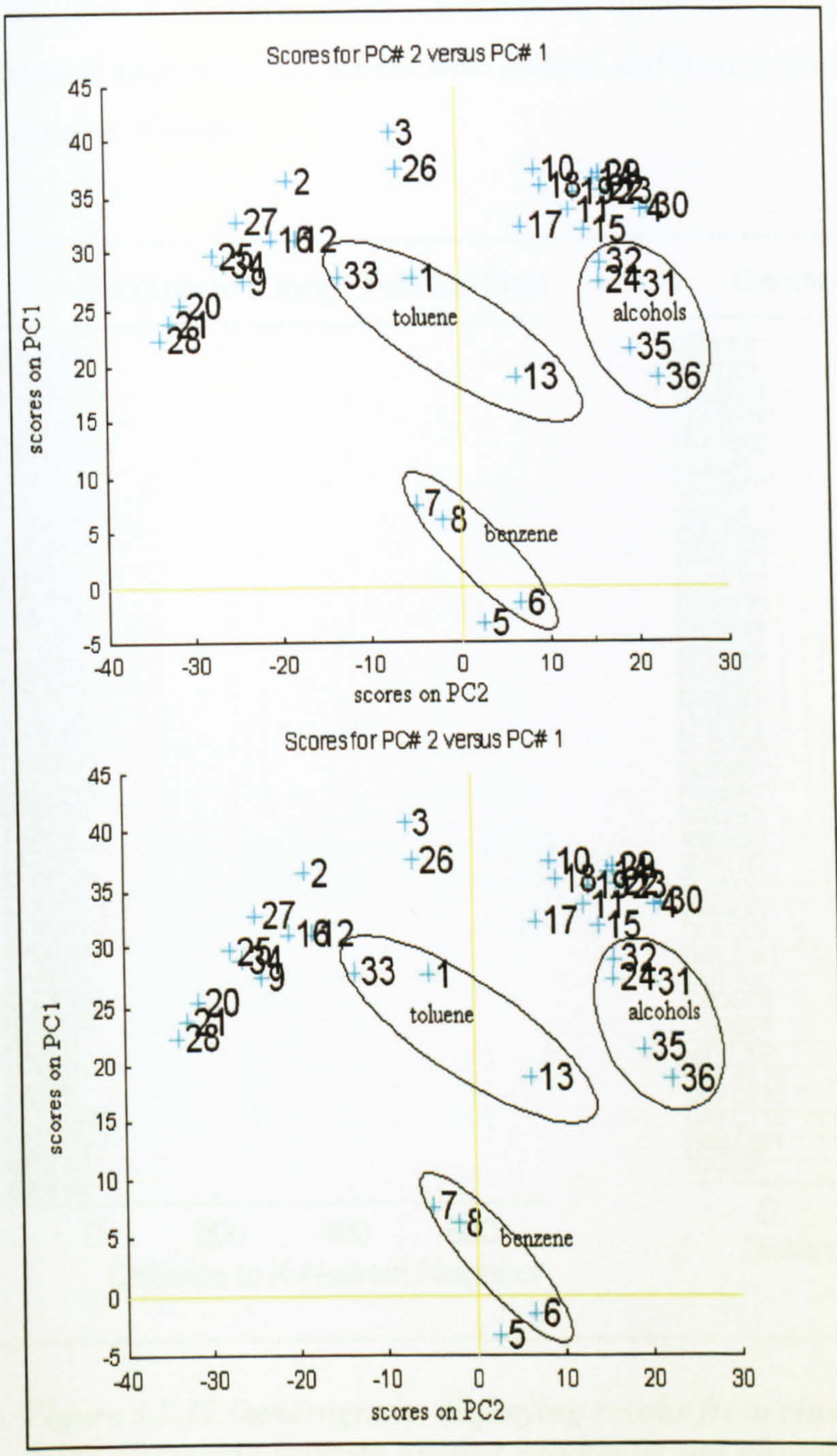


Figure 3.7.11 Score plots from PCA performed on SNV treated data

3.7.1.4 Cluster Analysis

Cluster analysis is a useful way of comparing similarities between samples by grouping samples together according to the relative distance between them. Dendrograms were employed as a means of displaying this information. A large number of dendrograms were generated using different distancing and data treatment, so only some of the dendrograms obtained are displayed in figures 3.7.12 and 3.7.13.

As can be seen, spectra of different concentrations of benzene are clustered together as are those of p-xylene. There is some grouping of similar compounds for example ethanol and methanol, but not with butanol and there is no clustering of the different isomers of xylene.

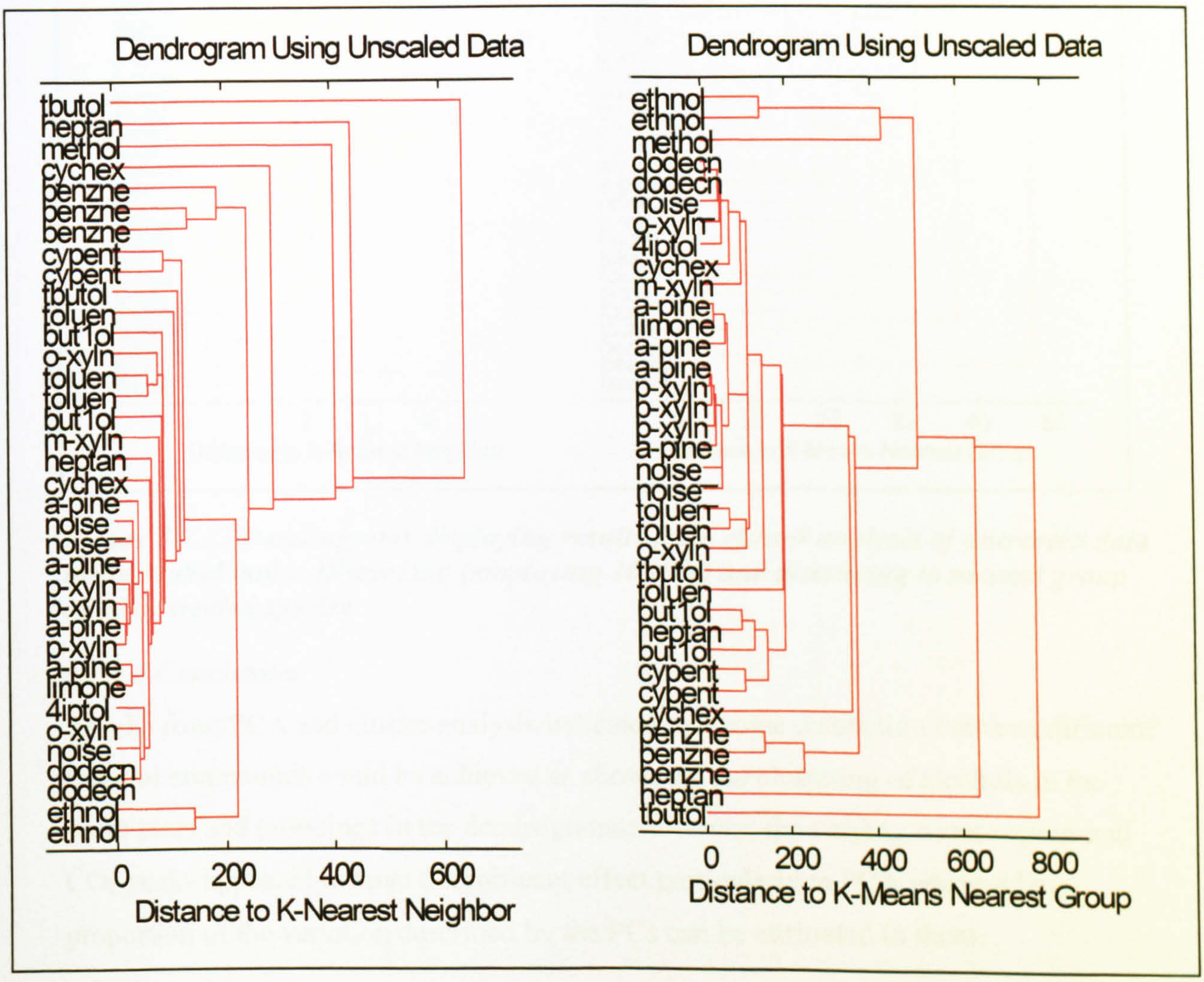


Figure 3.7.12 Dendrograms displaying results from cluster analysis of untreated data using distancing between nearest neighbour and distancing to nearest group

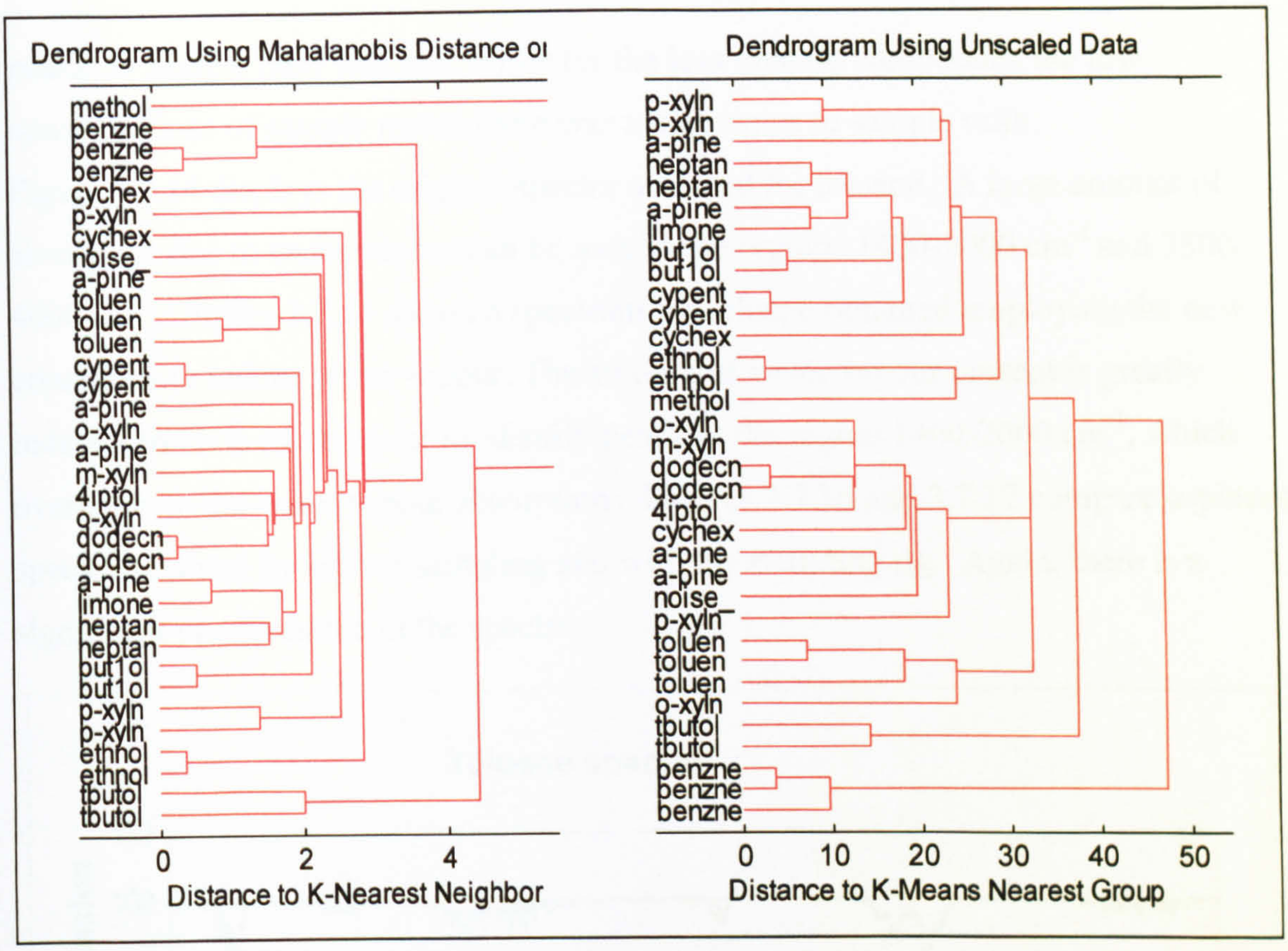


Figure 3.7.13 Dendrograms displaying results from cluster analysis of untreated data using Mahalonobis Distancing (employing 10 PCs) and distancing to nearest group on SNV treated spectra

3.7.1.5 Conclusion

Results from PCA and cluster analysis indicated that some distinction between different types of compounds could be achieved as shown by the clustering of alcohols in the score plots and groupings in the dendrograms; however, the varying water vapour and CO_2 peaks appeared to have a significant effect particularly in PCA where a large proportion of the variation described by the PCs can be attributed to them.

3.7.2 Spectral Correction

To remove the possible effect and interference of the varying water and CO_2 peaks, various means of correction where applied to the spectra.

As described in section 3.1, the spectra were ratioed against a cell that was evacuated employing the sampling rig before being filled with air from the surrounding environment and this helped remove some of the water and CO_2 peaks. The concentration of sample was dependent on the vapour phase of the solvent. For the more volatile compounds such as benzene, toluene and methanol the noise (water and CO_2

peaks) to sample ratio was low, while for the less volatile compounds the low concentrations of sample meant there was a high noise to sample ratio.

Figure 3.7.14 displays the original spectra obtained for toluene. A large amount of absorption due to water vapour can be seen in the regions 1400-2000 cm^{-1} and 3500-4000 cm^{-1}. Figure 3.7.15 shows a spectrum for toluene obtained employing the new conditions to reduce water vapour. The amount of water vapour present is greatly reduced and it is much easier to identify peaks in the region 1400-2000 cm^{-1}, which overlap with the water vapour absorptions. Figures 3.7.16 and 3.7.17 compare α-pinene spectra obtained by vapour sampling and with the sampling rig. Again, there is a significant improvement in the spectra.

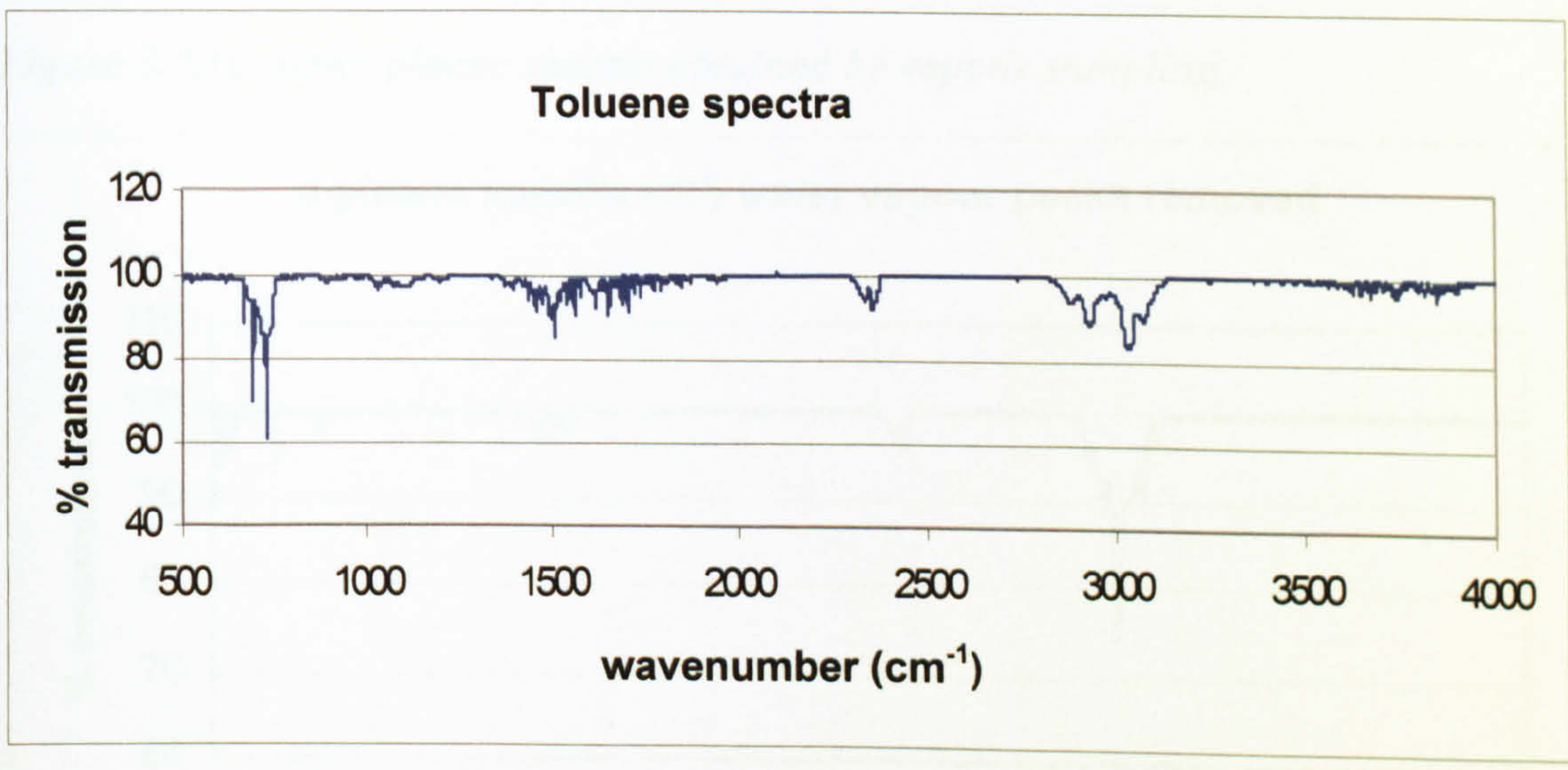


Figure 3.7.1.4 Spectrum obtained for toluene by sampling vapour

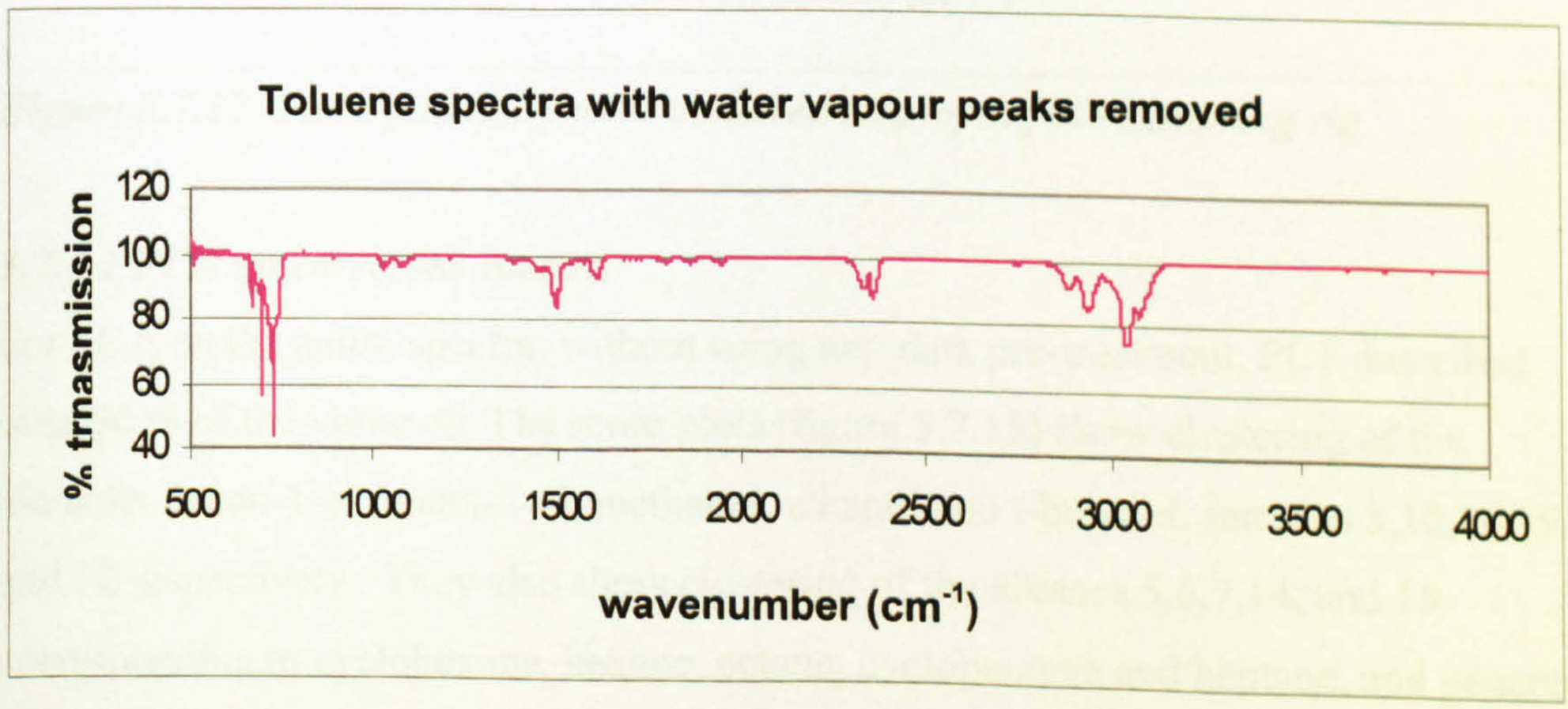


Figure 3.7.15 Spectrum for toluene employing the sampling rig

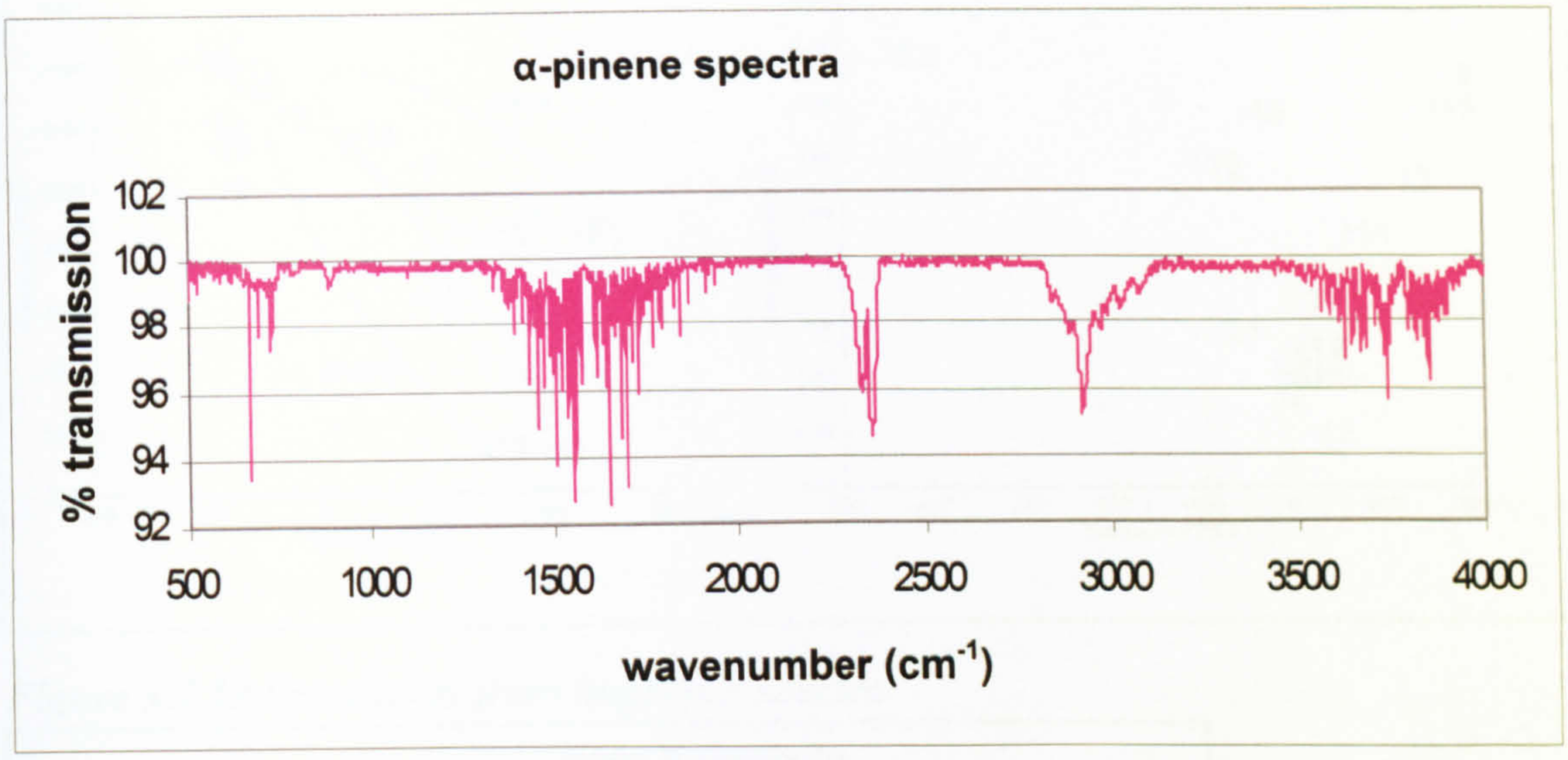


Figure 3.7.16 Alpha pinene spectra obtained by vapour sampling

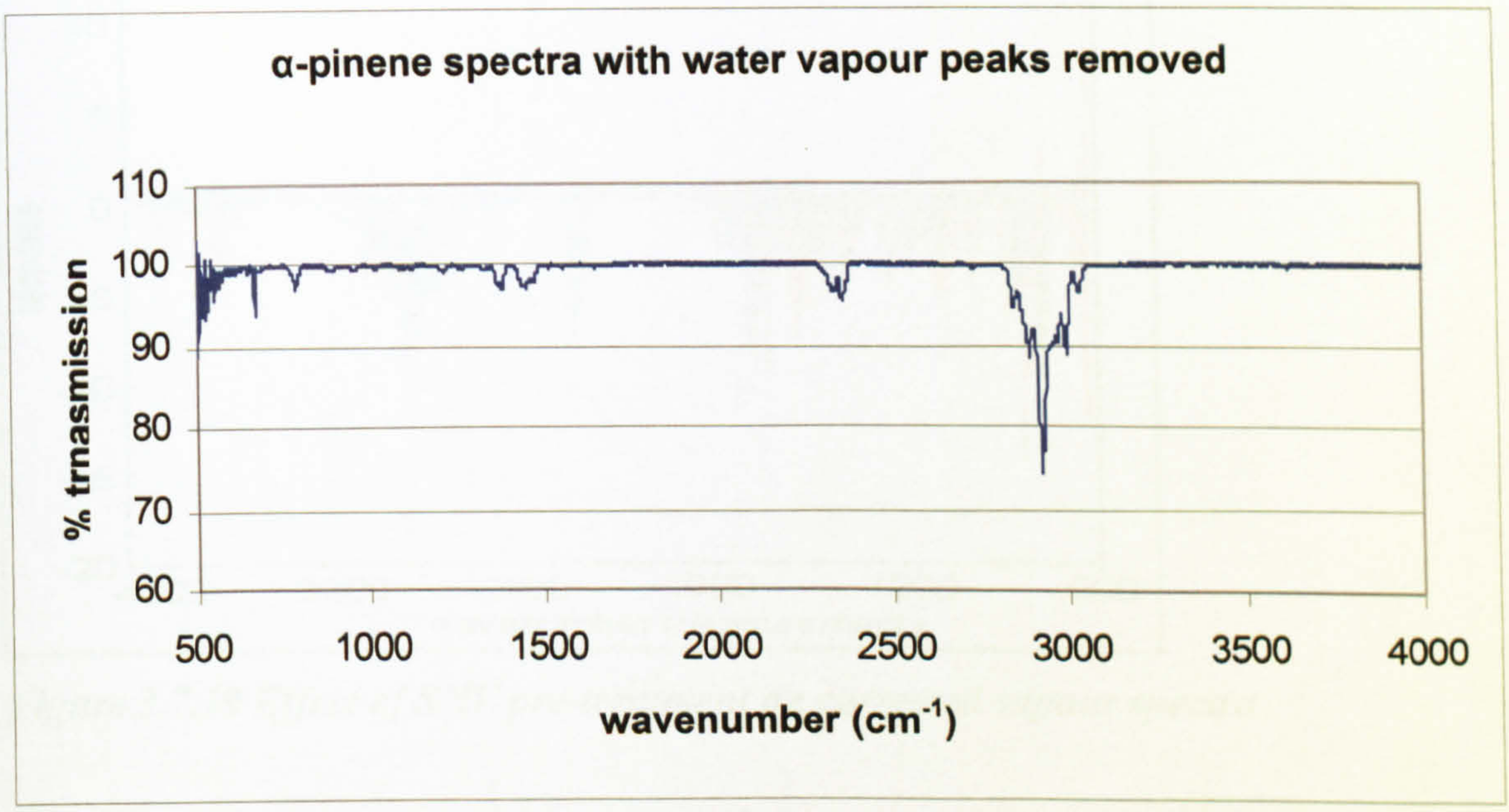


Figure 3.7.17 Alpha pinene spectra obtained employing the sampling rig

3.7.2.1 PCA on corrected spectra

For PCA on the entire spectra, without using any data pre-treatment, PC1 described over 90 % of the variance. The score plots (figure 3.7.18) show clustering of the alcohols, butan-1-ol, butan-2-ol, methanol, ethanol and t-butanol, samples 3,10,16,19,21 and 22 respectively. They also show clustering of the alkanes 5,6,7,14, and 15 corresponding to cyclohexane, hexane, octane, cyclopentane and heptane, and general separation of the rest.

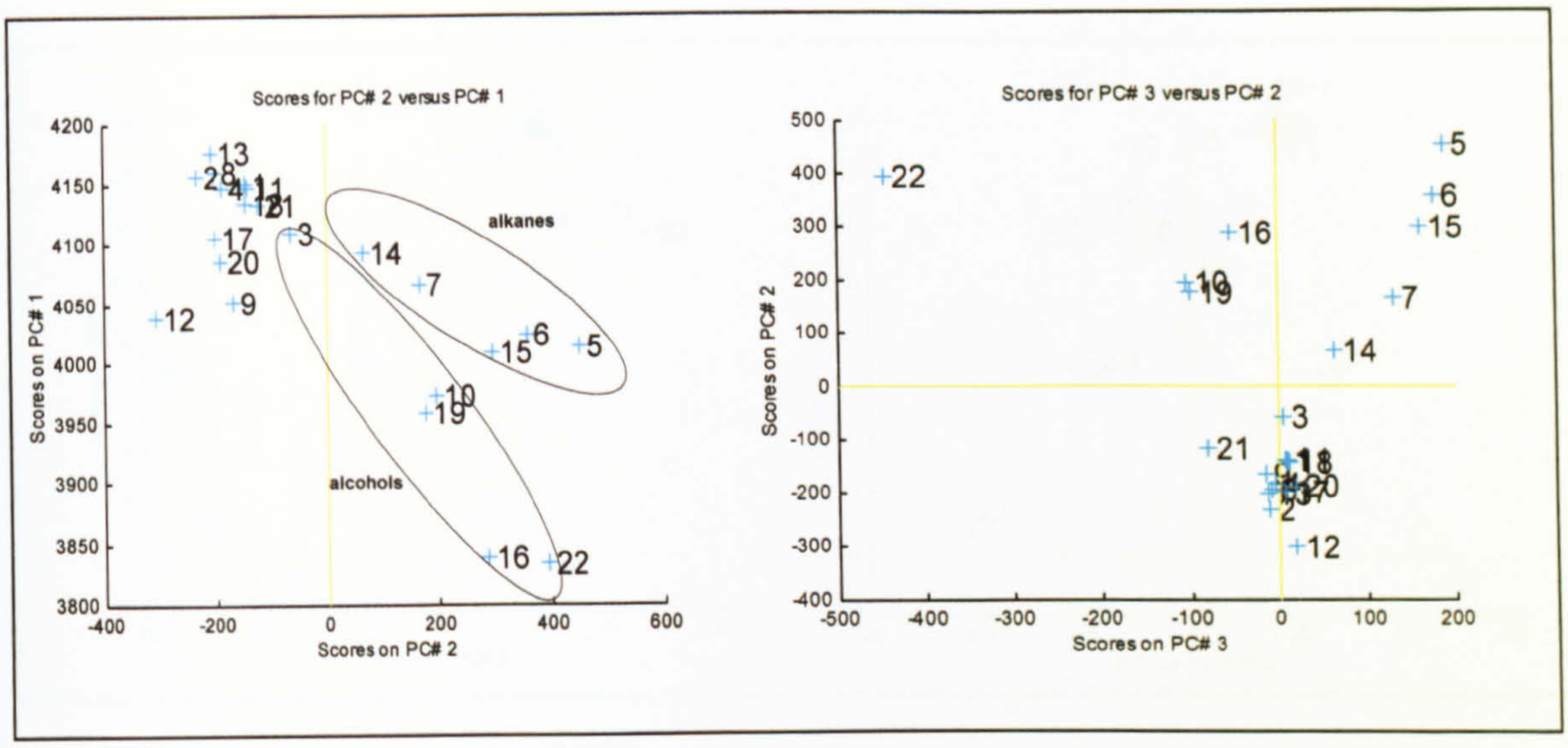


Figure 3.7.18 Score plots from improved spectra

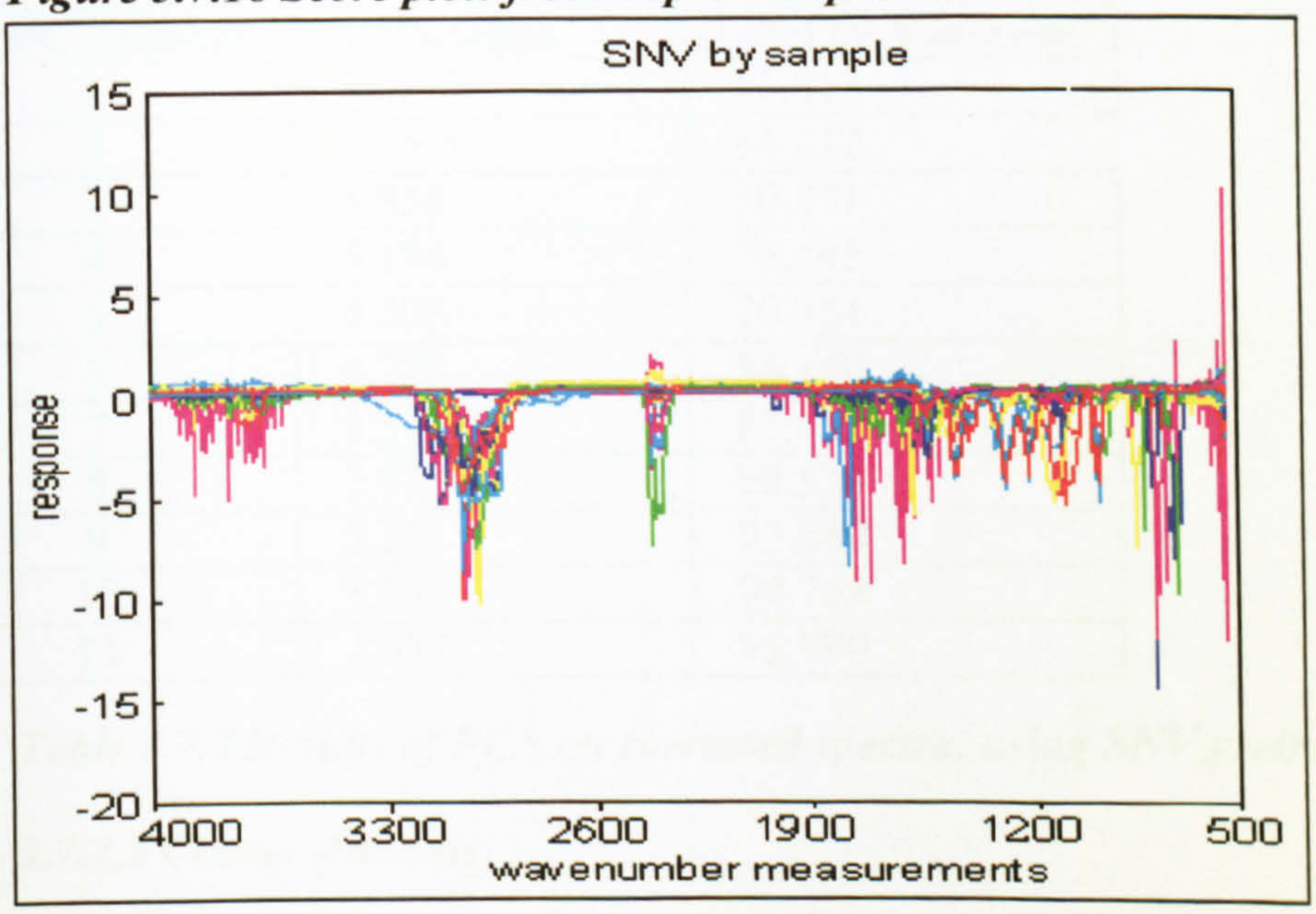


Figure 3.7.19 Effect of SNV pre-treatment on corrected vapour spectra

Figure 3.7.19 shows the spectra after SNV pretreatment, and two of the score plots obtained from PCA (on the SNV treated data) are displayed in figure 3.7.20. Again, score plots display a greater dispersion then the original analysis on spectra collected by vapour sampling. Looking at table 3.7.2, only 52 % of the variation between spectra is described by PC1.

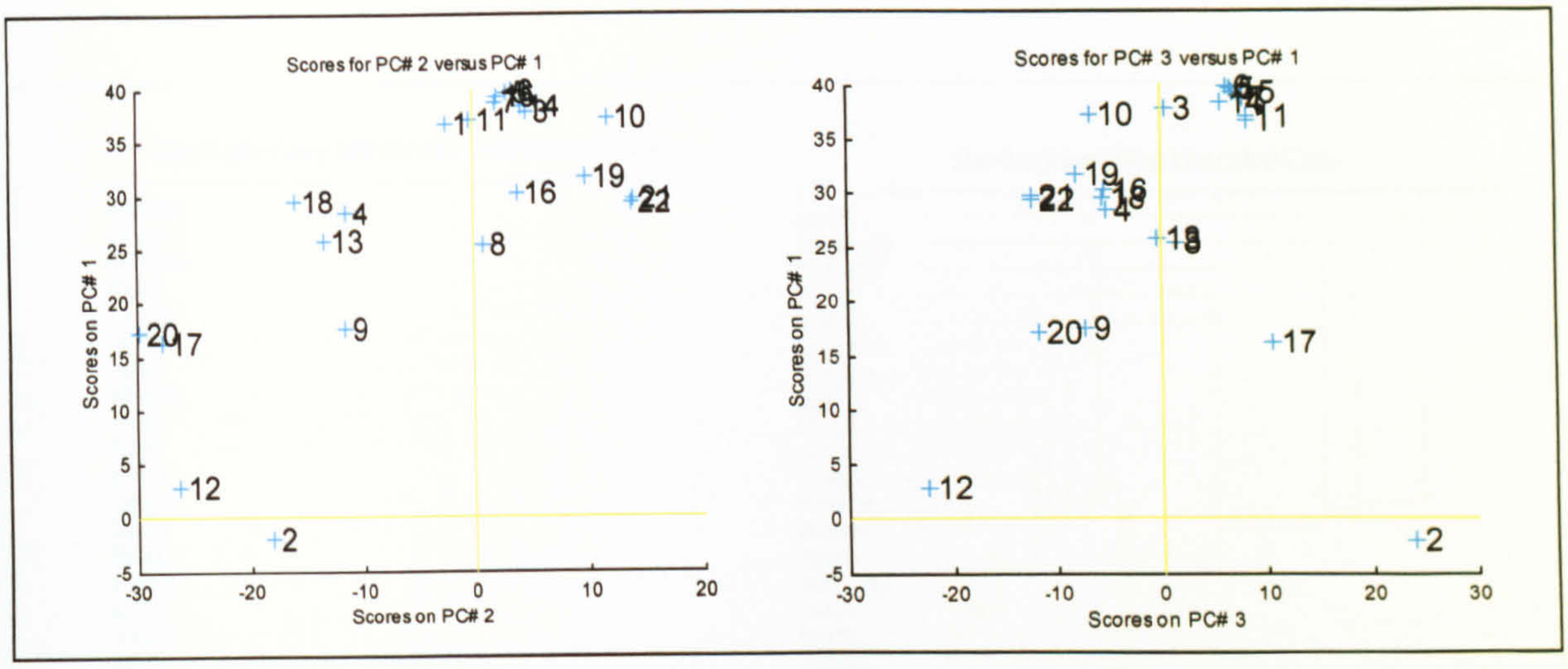


Figure 3.7.20 Score plots for SNV pre-treated vapour spectra

PC Number	% Variance	Total % Variance
1	53.813	53.813
2	10.599	64.412
3	5.838	70.251
4	5.194	75.445
5	4.406	79.851
6	4.288	84.140
7	3.512	87.652
8	2.887	90.539
9	2.321	92.860
10	1.899	94.758
11	1.032	95.790

Table 3.7.2 Results of PCA on corrected spectra, using SNV pretreatment

3.7.2.2 Cluster Analysis

Dendrograms of data using both untreated data and SNV treated data were produced. As before, various methods were used including; K distancing to the nearest group, K distancing to nearest neighbour and using data with and without PCA.

The dendrograms (figure 3.7.21) demonstrate good clustering for alcohol and alkane compounds, so in theory the information can be utilised to identify different types of compounds to a limited extent.

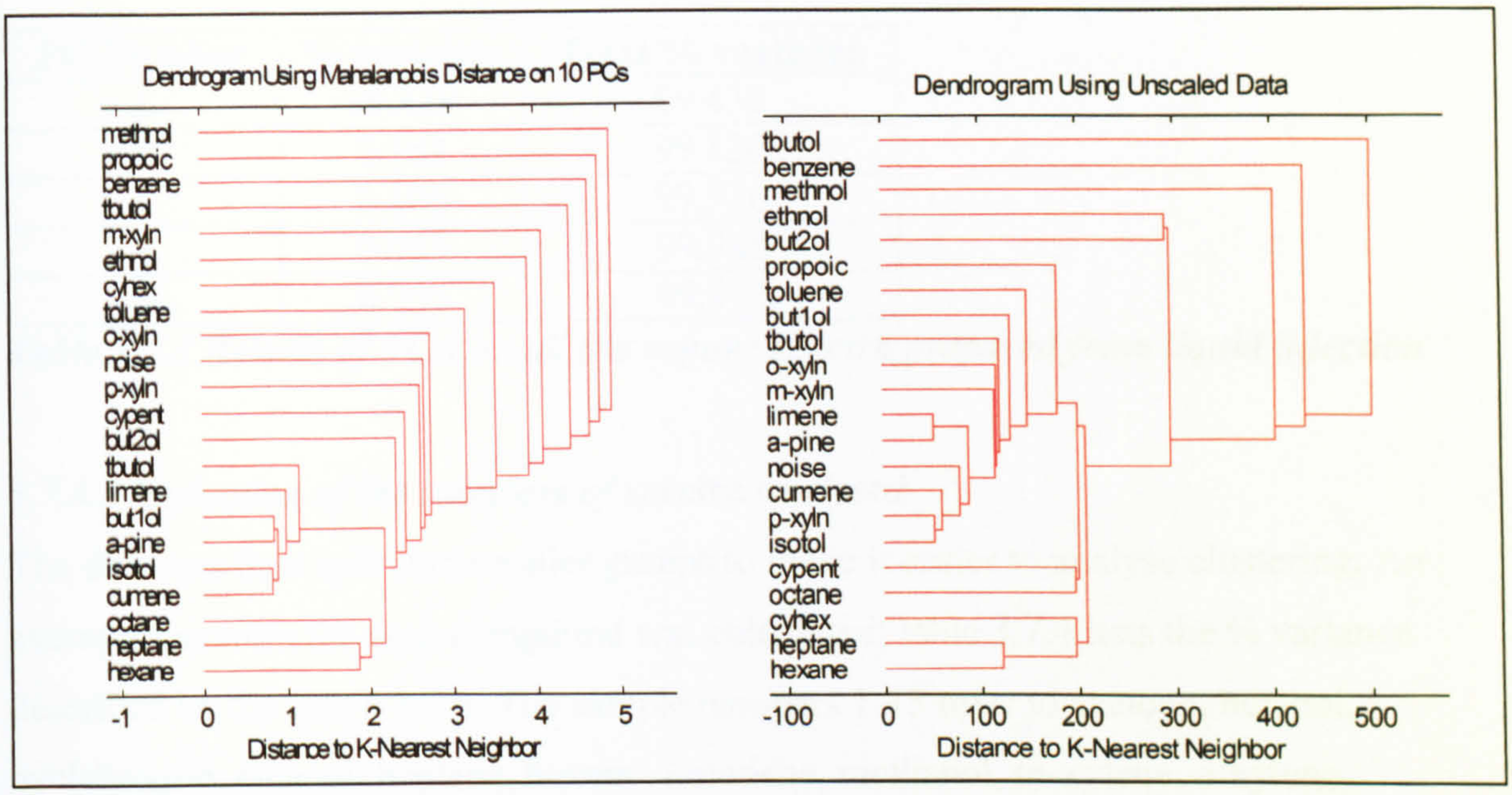


Figure 3.7.21 Dendrograms displaying results for cluster analysis on corrected spectra with no further data pre-treatment

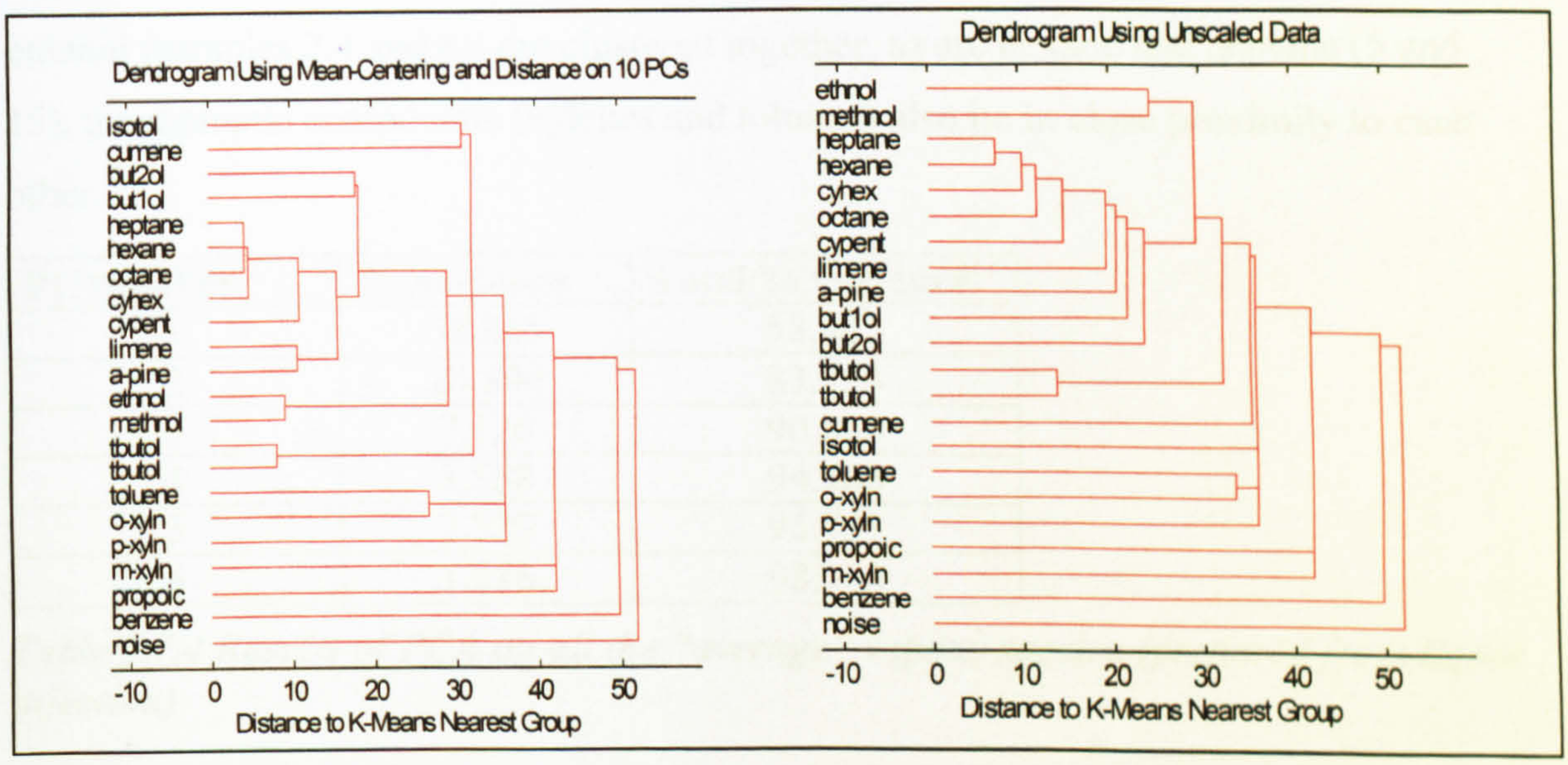


Figure 3.7.22 Dendrograms displaying results for cluster analysis on corrected spectra with SNV pre-treatment

3.7.2.3 Conclusion for corrected spectra

Results obtained for corrected spectra gave better clustering of similar types of compounds than the previous uncorrected spectra. A drawback with this approach to collecting spectra was the very limited control of sample concentrations particularly for the less volatile compounds, this led to the next stage of the work where a different approach to sample preparation was applied.

Pages Missing not Available

PC Number	% variance	Total % variance
1	99.438	99.438
2	0.382	99.820
3	0.106	99.926
4	0.063	99.988
5	0.005	99.993

Table 3.7.3 Results of PCA on all the vapour spectra prepared from liquid injection

3.7.3.3 Reduction of the numbers of spectra analysed

The data was then split into smaller groups to make it easier to analyse clustering. An average spectrum for each compound was calculated; table 3.7.4 lists the % variance described by the first 6 PC's. The sample numbers 1-15 refer to acetone, butanol, cyclohexane, ethanol, heptane, hexene, limonene, methanol, m-xylene, o-xylene, propanoic acid, p-xylene, tert-butyl alcohol, toluene and hexane, respectively. From the score plot of the first two principal components, the alcohols (methanol, ethanol, and butanol (samples 2,4 and 8)) are clustered together, as are hexane and heptane (5 and 15), the aromatic compounds (xylenes and toluene) also lie in close proximity to each other.

PC number	% variance	Total % variance
1	58.865	58.865
2	24.890	83.754
3	7.126	90.880
4	3.588	94.468
5	2.329	96.797
6	1.386	98.183

Table 3.7.4 Results of PCA on all the "average" vapour spectra (prepared from liquid injection)

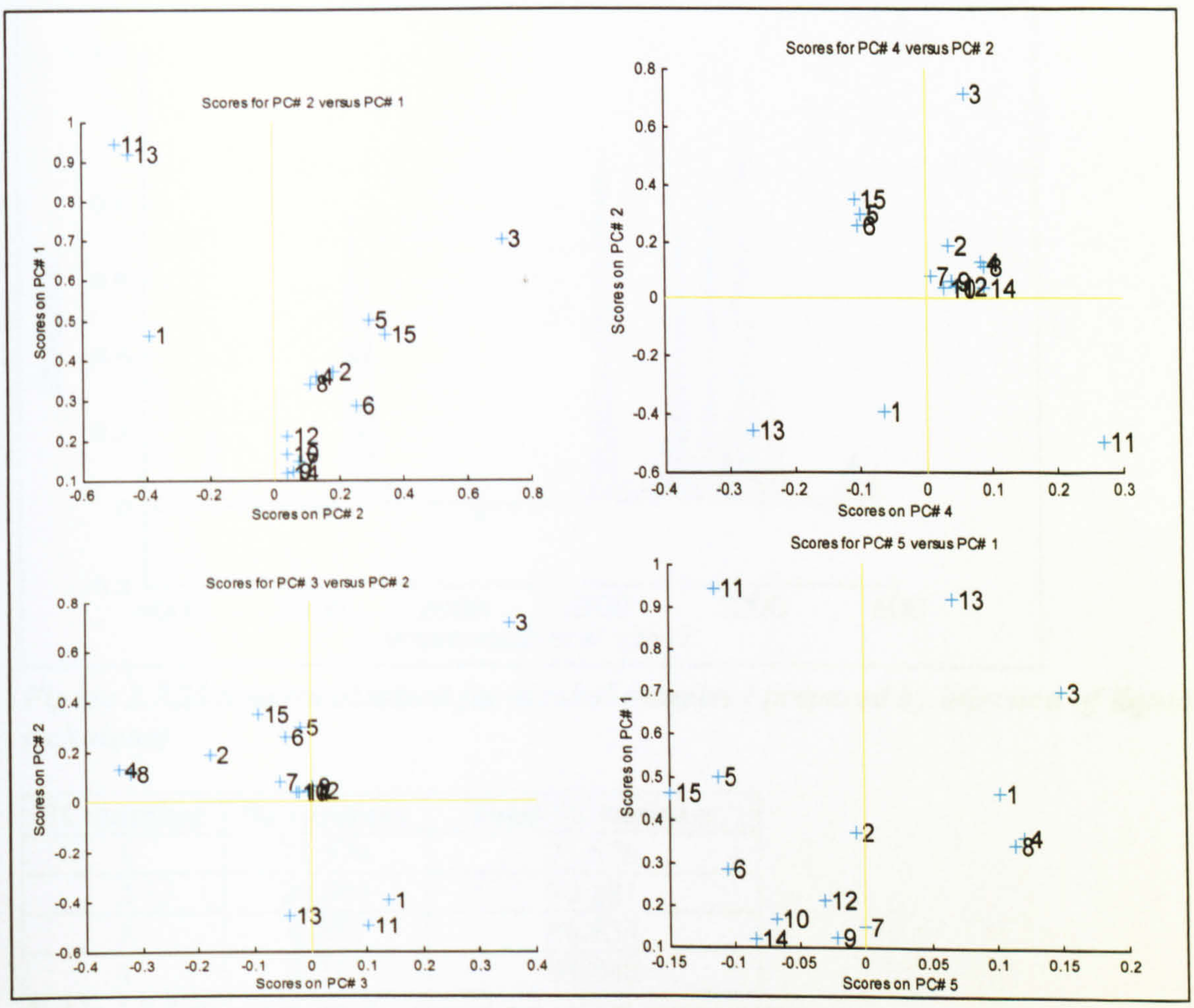


Figure 3.7.24 Score plots obtained for "average" vapour spectra (from samples prepared by injection of liquid technique)

3.7.3.4 Analysis of Alcohols

This consisted of the analysis on just alcohol spectra. Samples 1-8 were different concentrations of butan-1-ol, samples 9-15 were methanol, samples 16-24 were ethanol and samples 25-32 consisted of tert-butyl alcohol. These spectra are shown in figure 3.7.25.

Over 99 % of variation was described by four PCs (table 3.7.5).

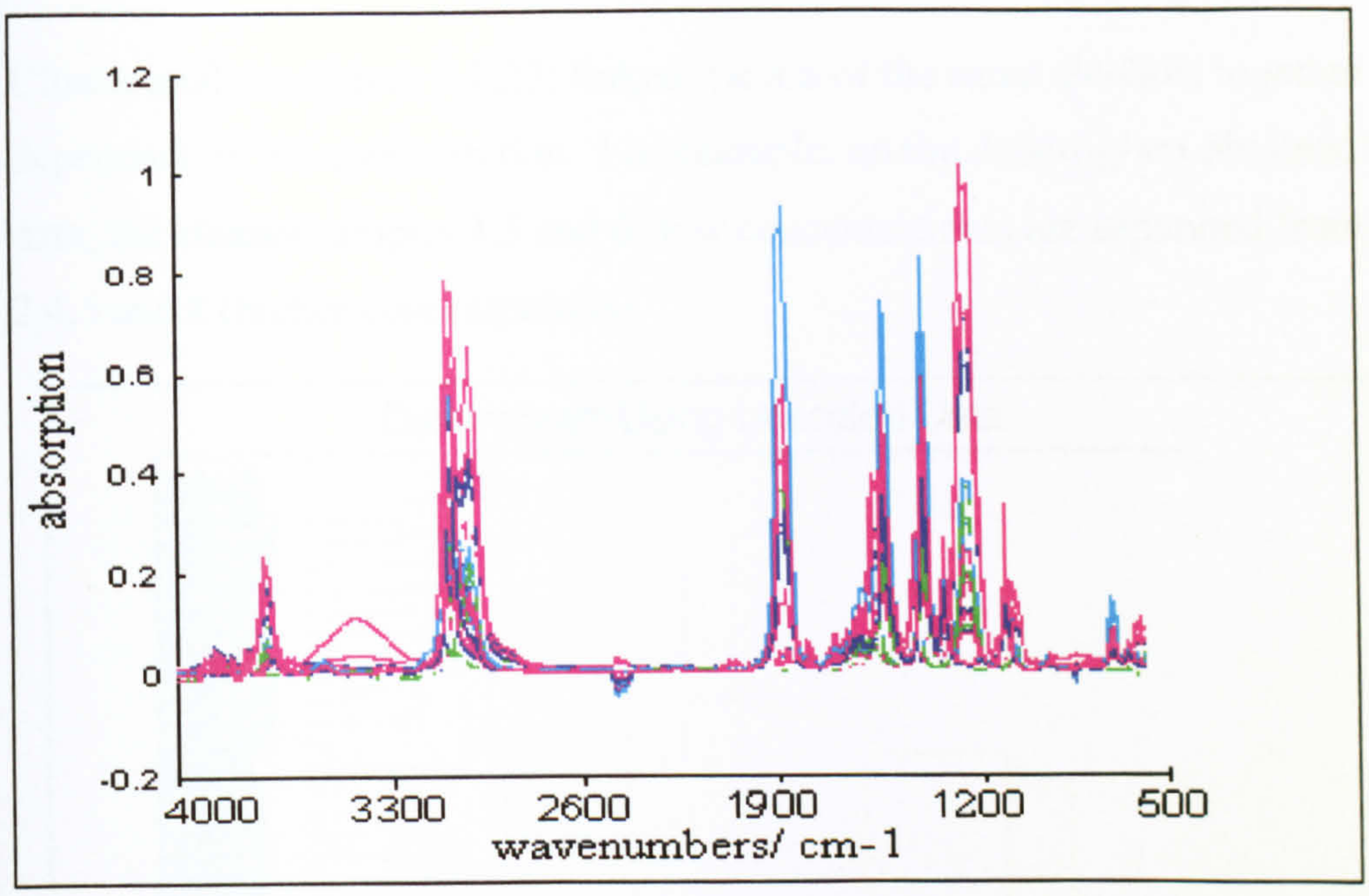


Figure 3.7.25 Spectra obtained for alcohol samples (prepared by injection of liquid technique)

PC number	% variance	Total % variance
1	71.576	71.576
2	20.605	92.181
3	4.653	96.834
4	2.215	99.048

Table 3.7.5 PCA results for first 4 principal components

Score plots (figure 3.7.26) show increasing separation of the alcohols as the concentration increases. This was particularly pronounced in the score plot of PC4 versus PC2 where each alcohol is spread on a different axis.

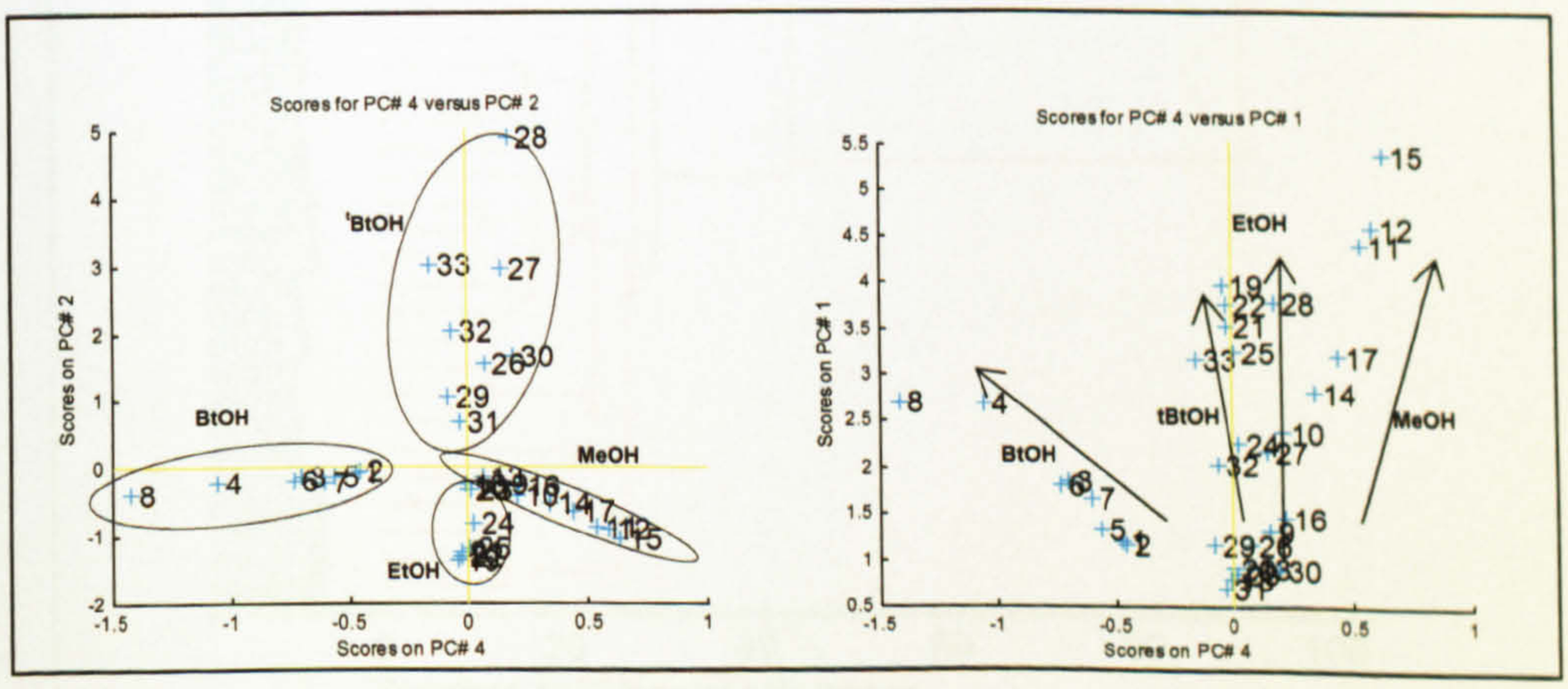


Figure 3.7.26 Score plots obtained from PCA performed on alcohol spectra

Cluster analysis (figure 3.7.27) linked spectra of the same alcohols together but was dependant on the concentration. For example, on the dendrogram obtained for unscaled data, the ethanol samples 1,3 and 6 (low concentrations) are separated from samples 2,4,5 and 8 (higher concentrations).

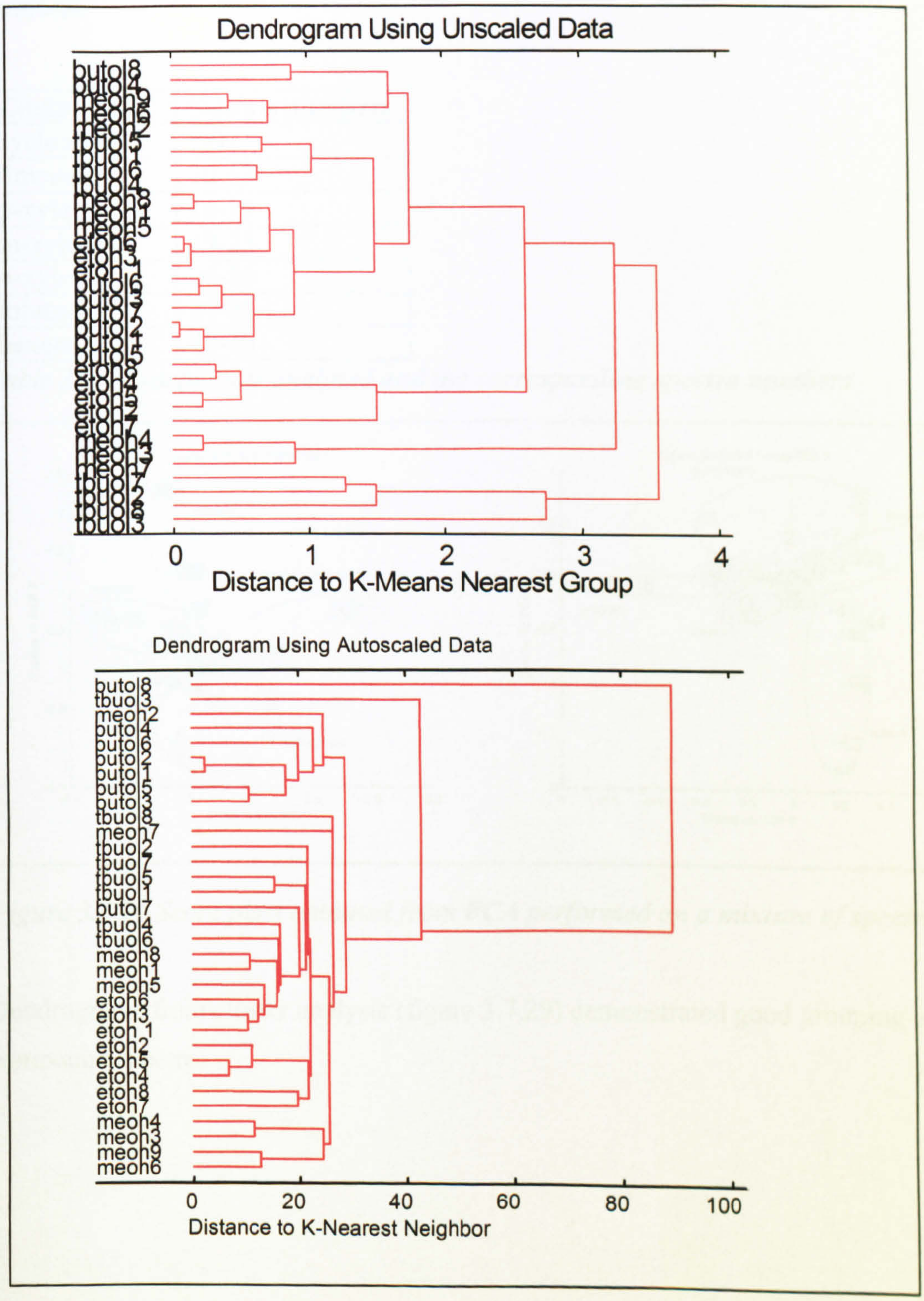


Figure 3.7.27 Dendrograms from cluster analysis on alcohol spectra

3.7.3.5 Analysis of mixtures

A mixture of different types was included in this analysis. This consisted of compounds that gave no clear clustering in previous work and consisted mainly of aromatic and cyclic compounds. Table 3.7.6 lists the compounds used and the corresponding spectra numbers.

Compound	Sample numbers
cyclohexane	1-9
limonene	10-13
p-xylene	14-18
m-xylene	19-25
o-xylene	26-30
toluene	31-39
hexene	40-46

Table 3.7.6 Compounds analysed and the corresponding spectra numbers

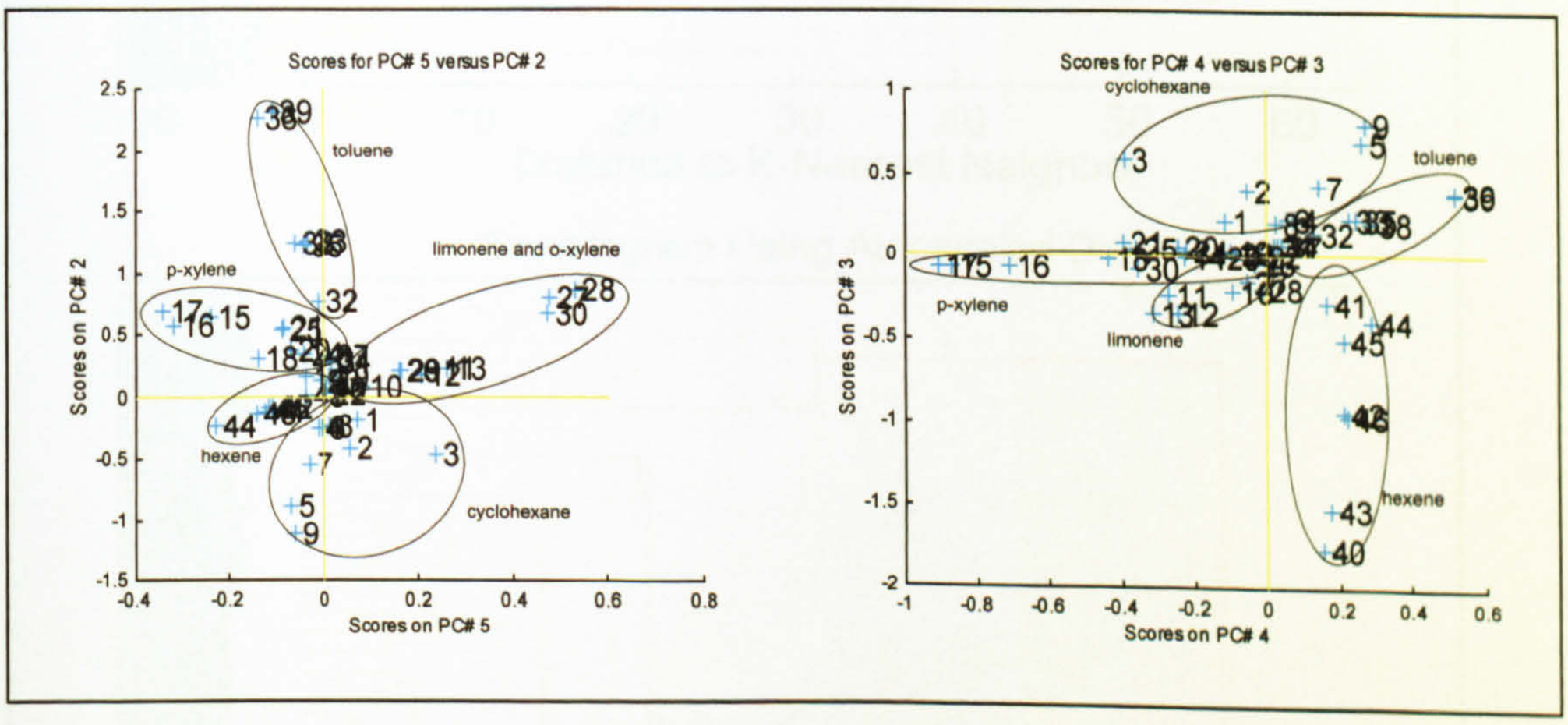


Figure 3.7.28 Score plots obtained from PCA performed on a mixture of spectra

Dendrograms from cluster analysis (figure 3.7.29) demonstrated good grouping of same compound spectra.

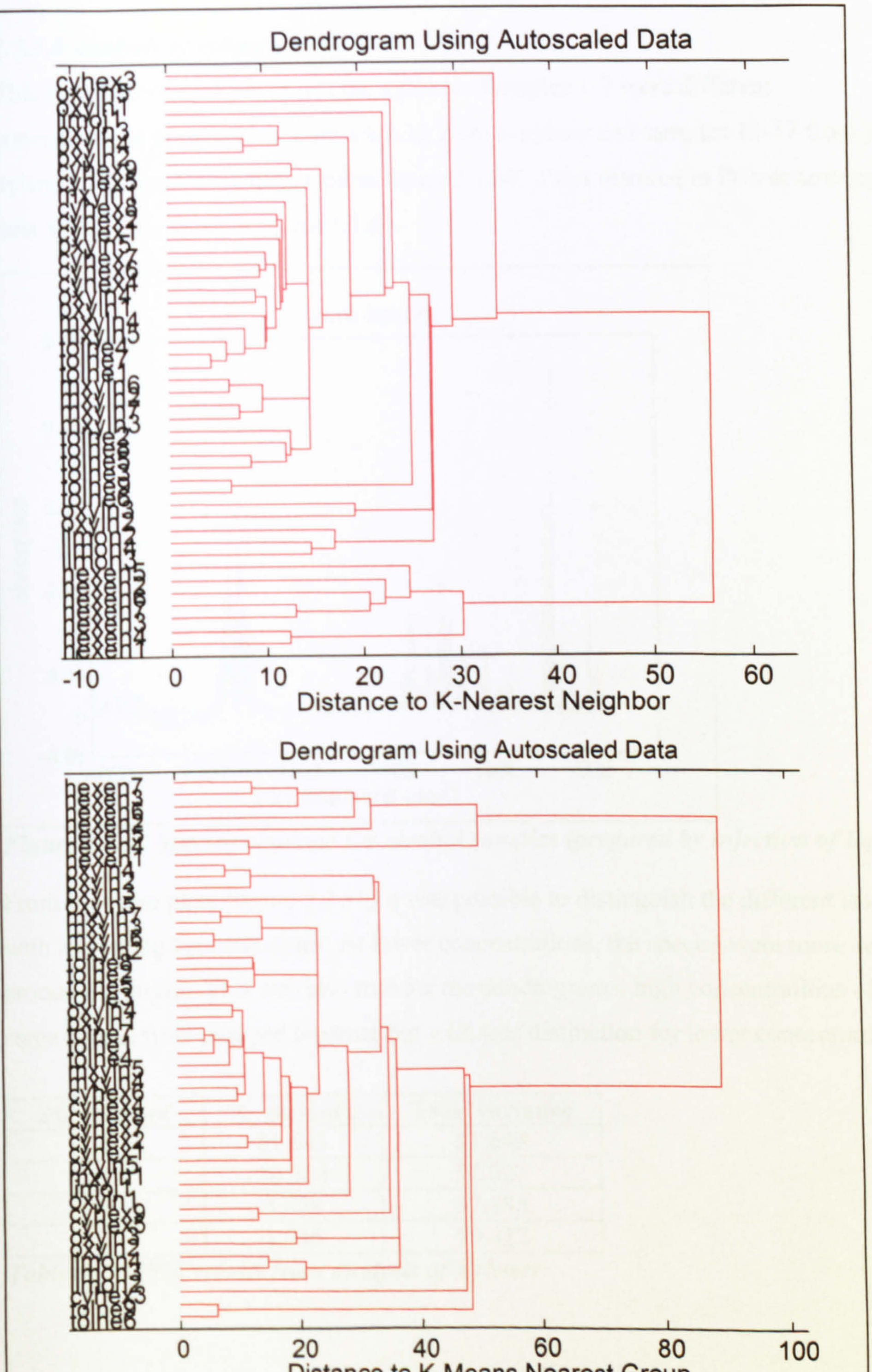


Figure 3.7.29 Dendrograms obtained for analysis of a mixture of spectra

3.7.3.6 Analysis of xylene spectra

This consisted of analysis of just the xylenes. Samples 1-7 were different concentrations of m-xylene, samples 8-12 from o-xylene and samples 13-17 from p-xylene. The spectra are displayed in figure 3.7.30. PCA resulted in PCs describing over 99% of the variance (table 3.7.6).

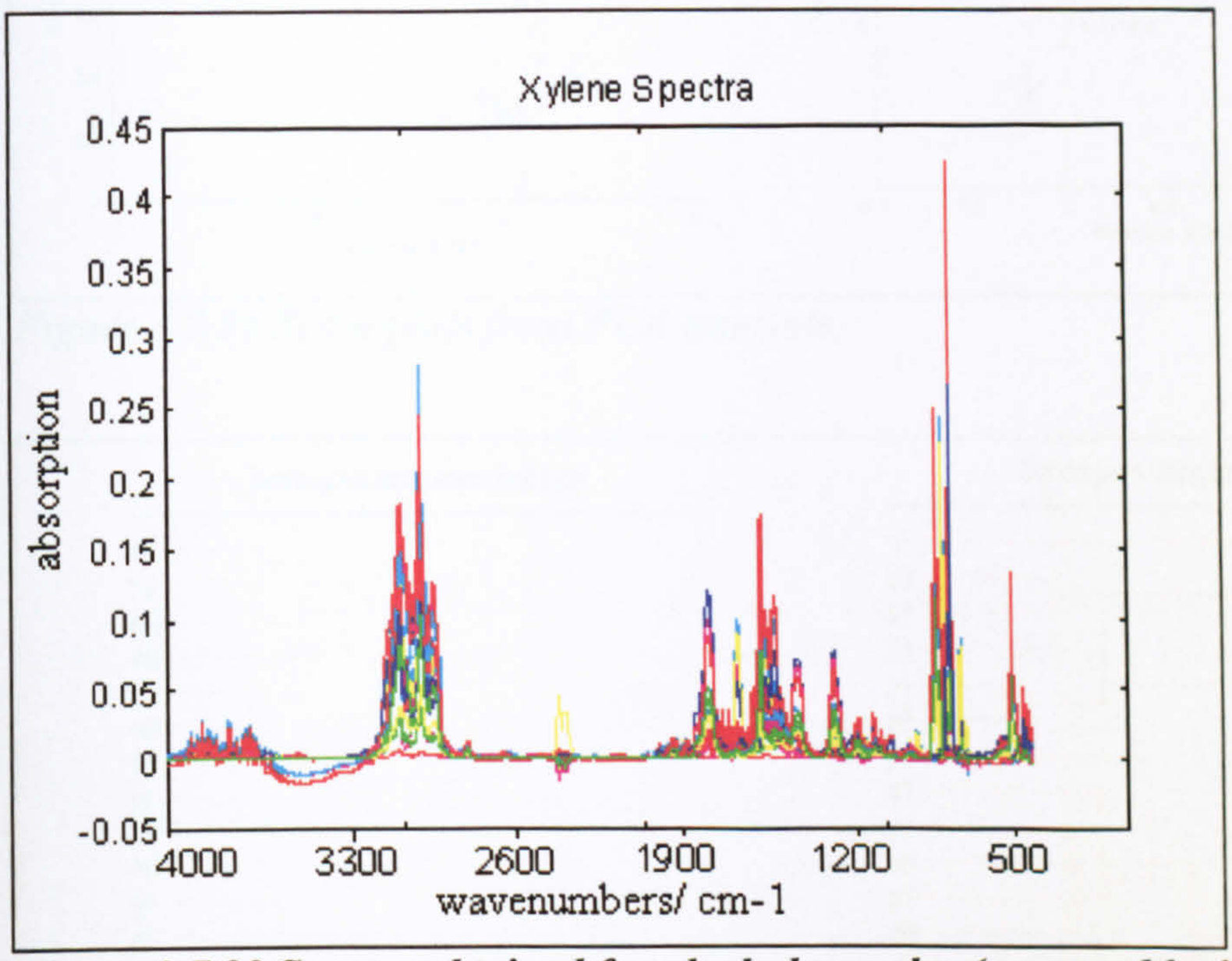


Figure 3.7.30 Spectra obtained for alcohol samples (prepared by injection of liquid)

From the score plots (figure 3.7.31) it was possible to distinguish the different isomers with increasing concentrations. At lower concentrations, the spectra were more centred around the origin. This was also true for the dendrograms; high concentrations of the same isomer were grouped together but with less distinction for lower concentrations.

PC number	% variance	Total variance
1	81.643	81.643
2	10.384	92.027
3	5.858	97.885
4	1.536	99.422

Table 3.7.6 PCA results from analysis of Xylenes

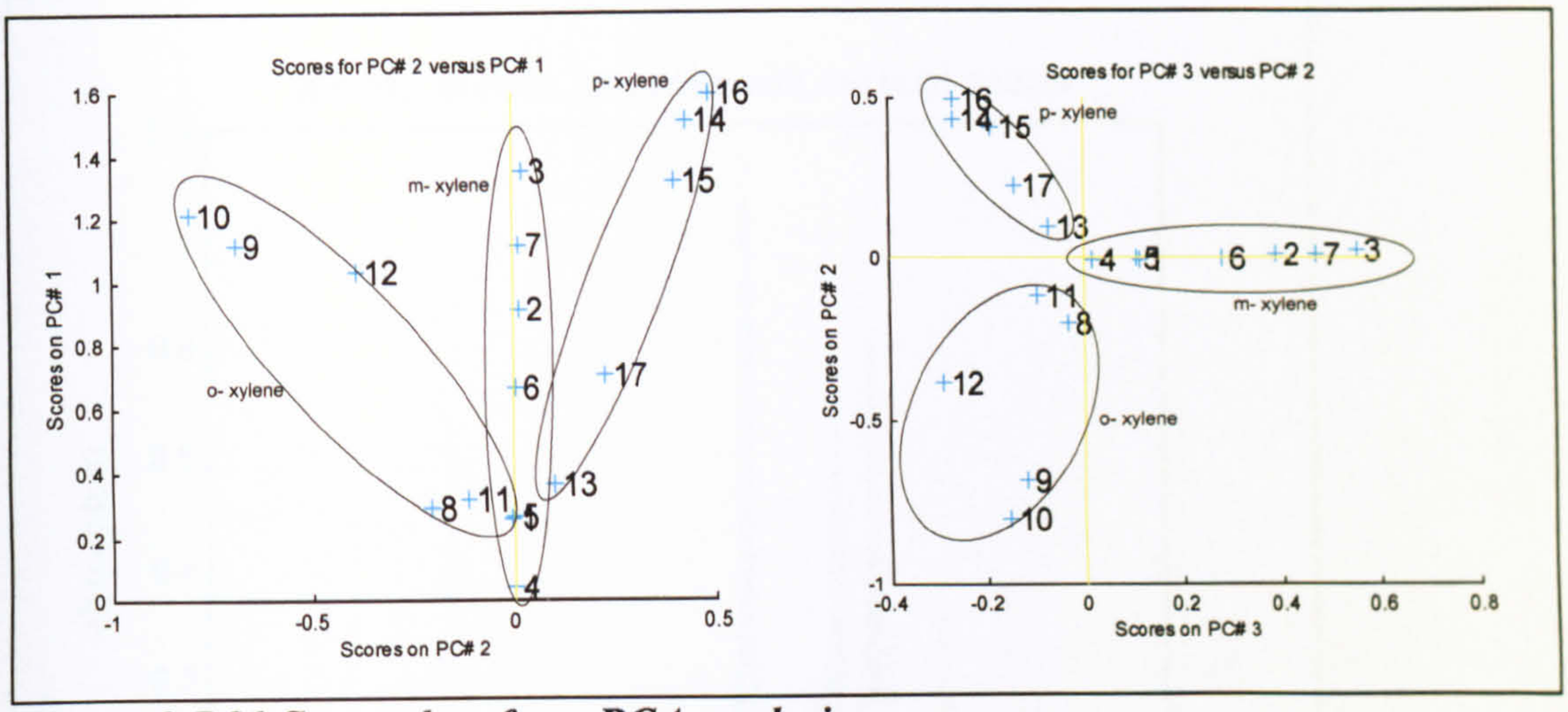


Figure 3.7.31 Score plots from PCA analysis

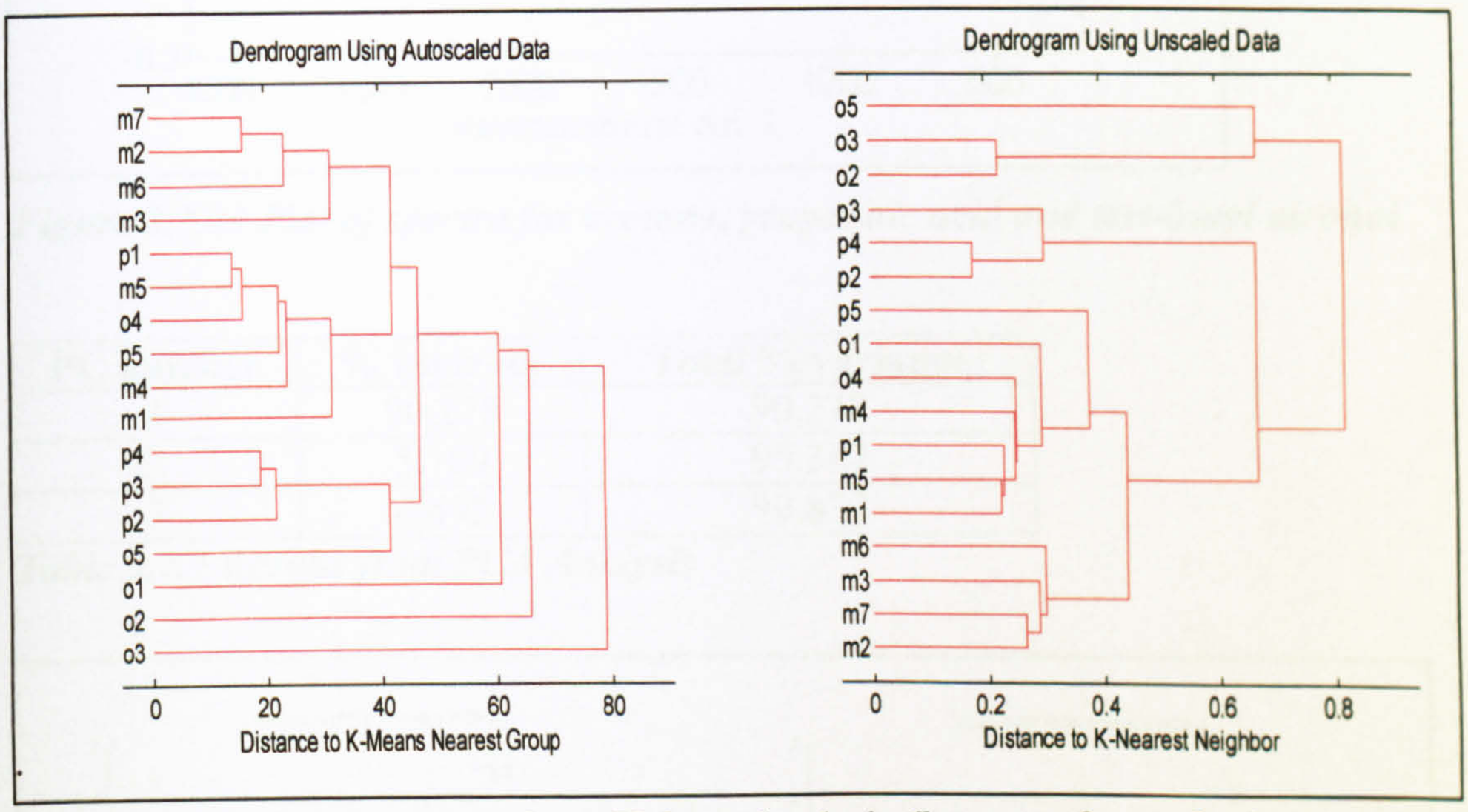


Figure 3.7.32 Dendrograms from Xylene Analysis (letters refer to the isomer)

3.7.3.7 Analysis of acetone, propanoic and tert-butyl alcohol

This analysis was performed on acetone spectra (samples 1-6), propanoic spectra, (samples 7-13) and tert-butyl alcohol (samples 14-21). The spectra are shown in figure 3.7.33, with results from PCA displayed in table 3.7.7. Two of the score plots obtained are shown in figure 3.7.34, the plot of PC2 versus PC3 demonstrates clustering of the different compounds; however, many of the samples are close to the origin making it harder to distinguish between low concentration samples. From the dendrograms (figure 3.7.35), it can be concluded that it is possible to identify propanoic acid (based on the fact that most of the propanoic samples were closely linked), but acetone and tert-butyl were much harder to distinguish.

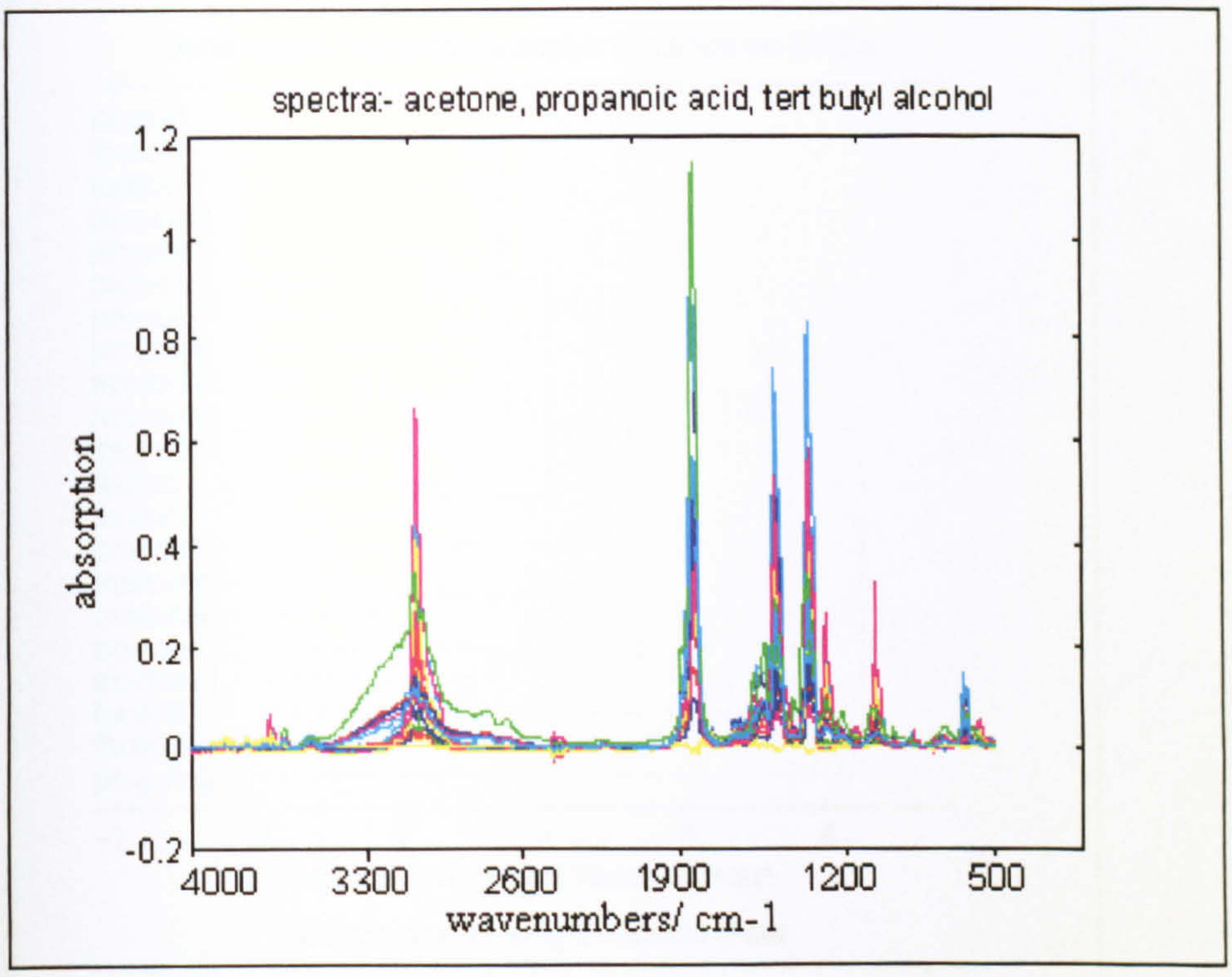


Figure 3.7.33 Plot of spectra for acetone, propanoic acid and tert-butyl alcohol

PC number	% variance	Total % variance
1	90.278	90.278
2	5.109	95.387
3	4.492	99.879

Table 3.7.7 Results from PCA Analysis

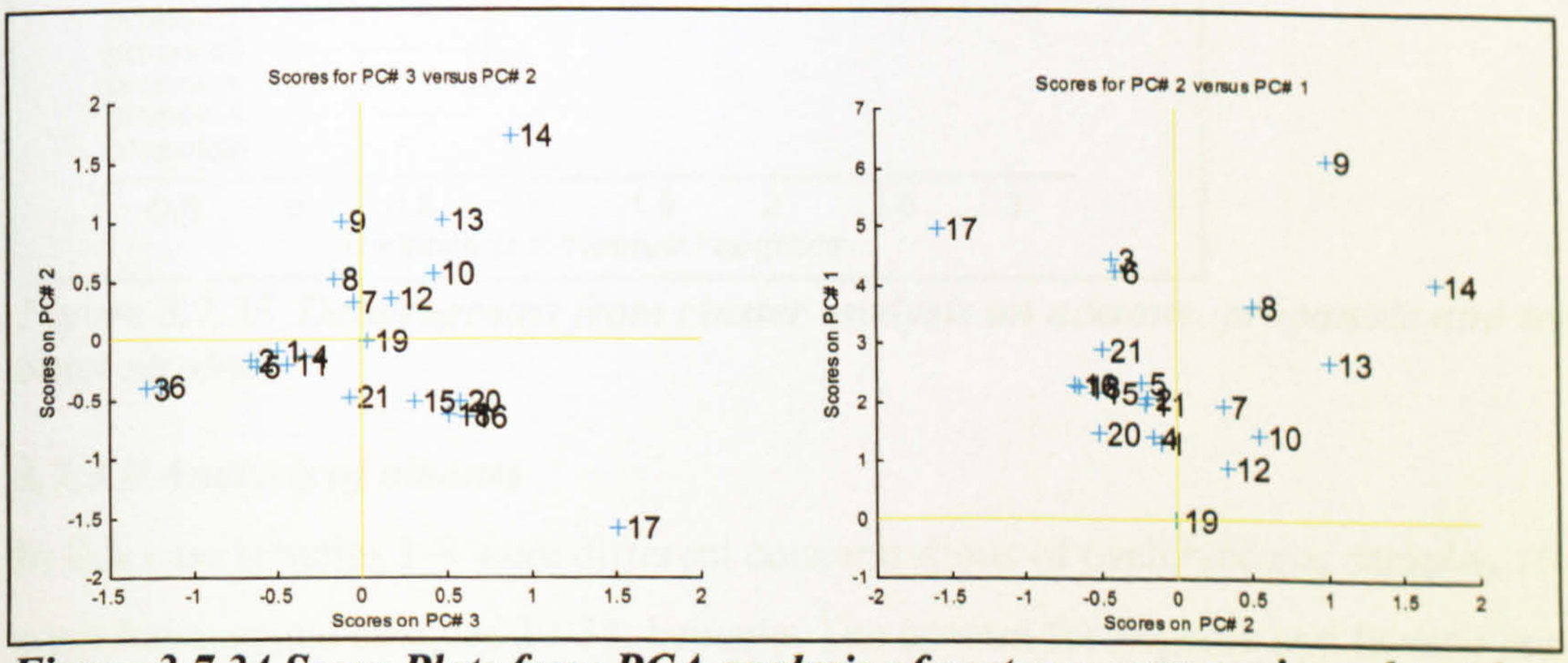


Figure 3.7.34 Score Plots from PCA analysis of acetone, propanoic and tert-butyl alcohol

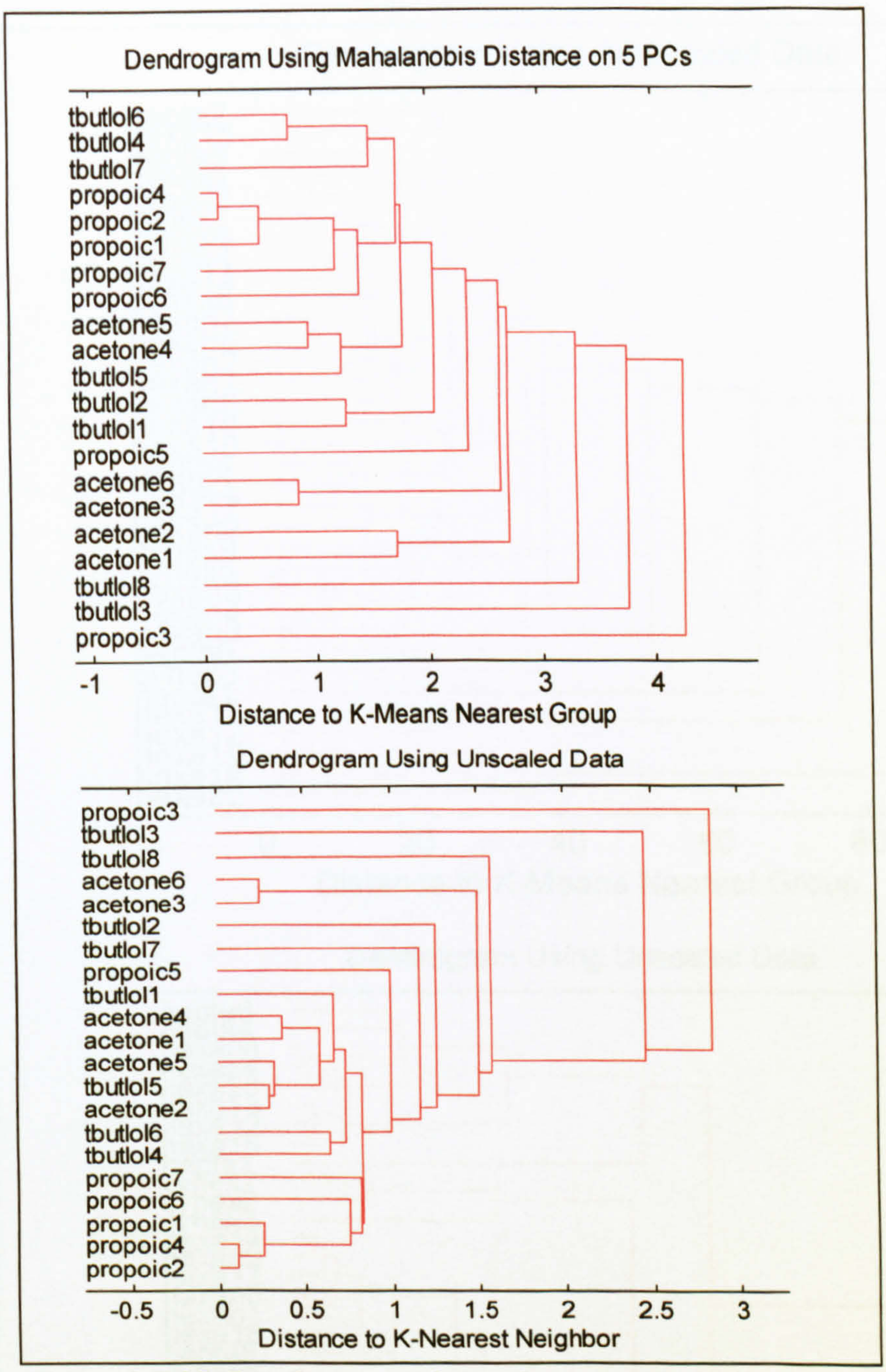
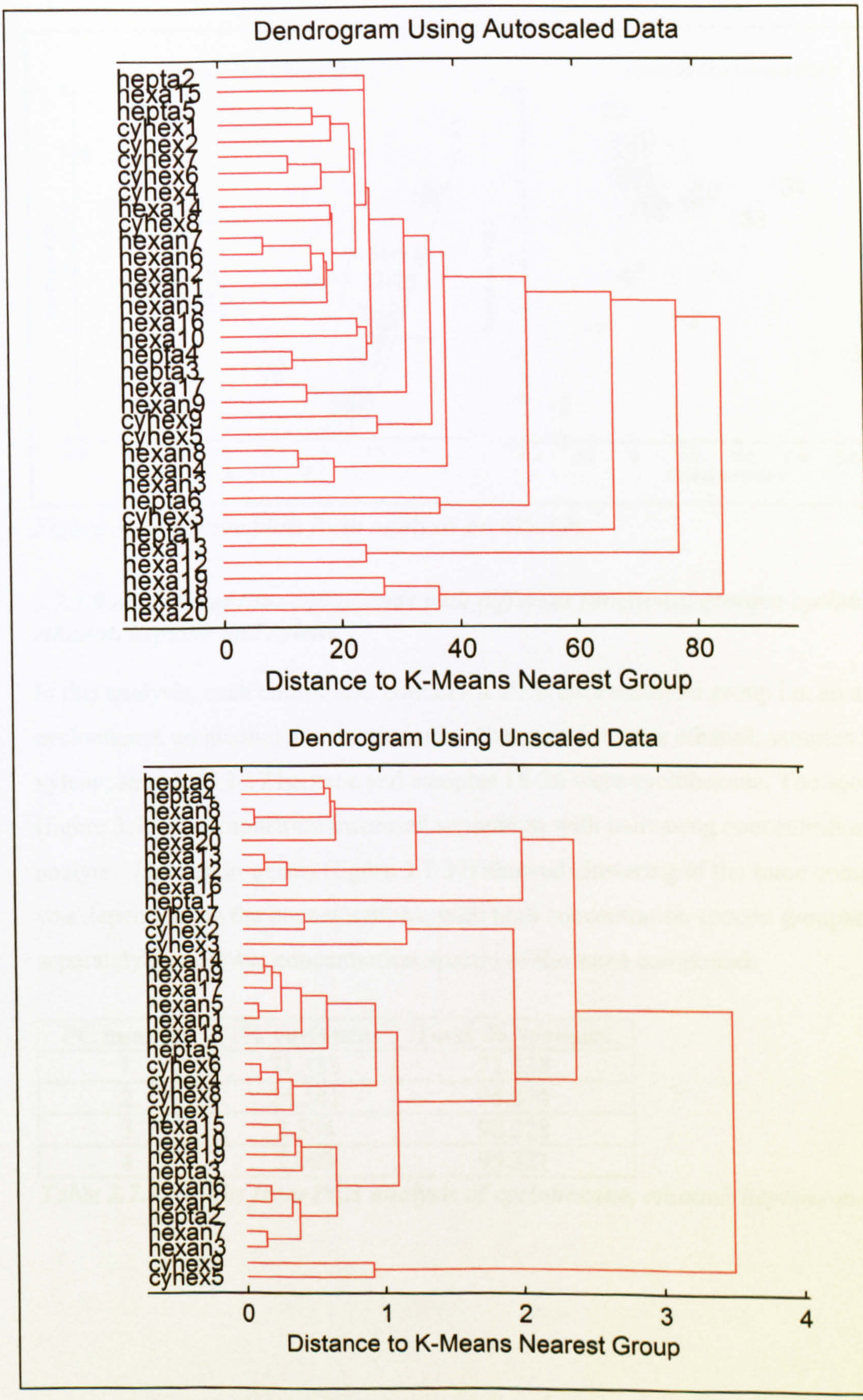


Figure 3.7.35 Dendrograms from cluster analysis on acetone, propanoic and tert-butyl alcohol

3.7.3.8 Analysis of alkanes

In this case,samples 1-9 were different concentrations of cyclohexane; samples 10-29 were hexane; and samples 30-34, heptane. The spectra for heptane and hexane are practically indistinguishable from each other as demonstrated in the score plots (figure 3.7.37) and dendrograms (figure 3.7.36), whereas, cyclohexane was clearly identified in the score plots.



3.7.36 Dendrograms from cluster analysis on acetone, propanoic and tert-butyl alcohol

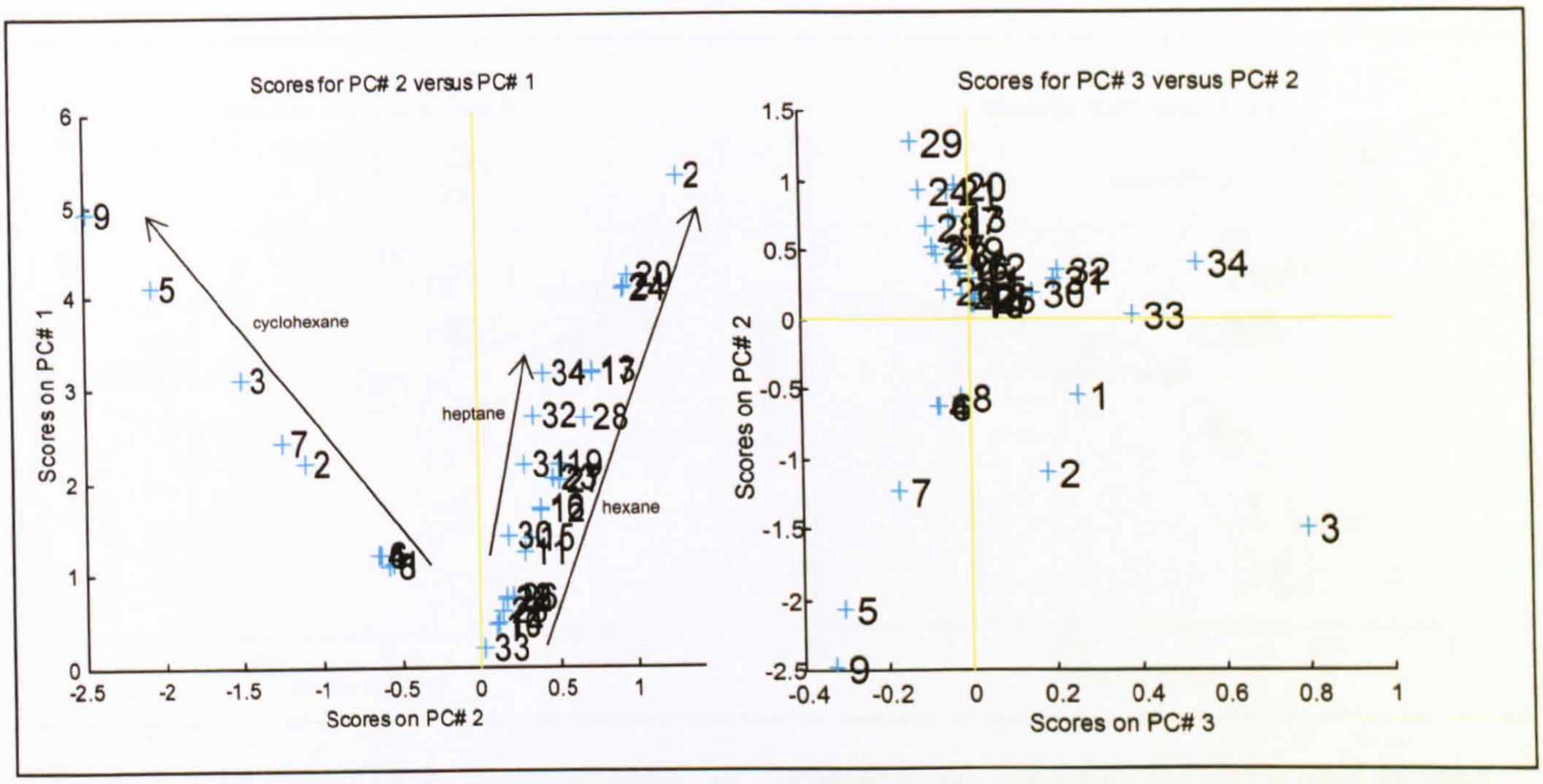


Figure 3.7.37 Score plots from analysis on alkanes

3.7.3.9 Analysis of four compounds with different functional groups; cyclohexane, ethanol, heptane and xylene

In this analysis, each component contains a different functional group i.e. an alkane, a cycloalkane, an alcohol and an aromatic. Samples 1-7 were ethanol, samples 8-12, o-xylene; samples 13-17 heptane and samples 18-26 were cyclohexane. The score plots (figure 3.7.38) demonstrate increased separation with increasing concentrations of each analyte. The dendrograms (figure 3.7.39) showed clustering of the same compound but was dependent on the concentrations, with high concentration spectra grouped separately from lower concentration spectra of the same compound.

PC number	% variance	Total % variance
1	71.333	71.333
2	24.541	95.874
3	2.354	98.228
4	1.093	99.321

Table 3.7.8 Results from PCA analysis of cyclohexane, ethanol, heptane and xylene

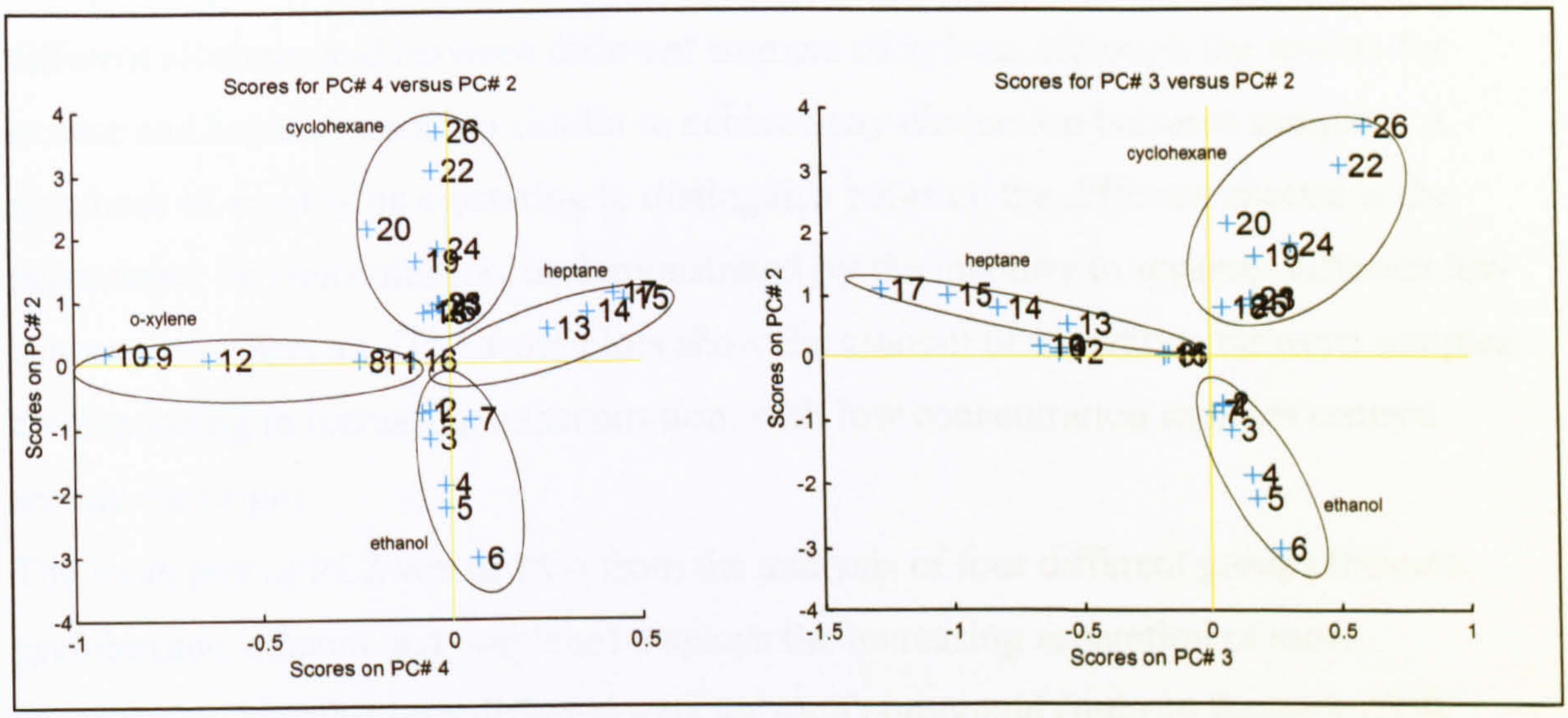


Figure 3.7.38 Score plots from analysis of cyclohexane, ethanol, heptane and xylene

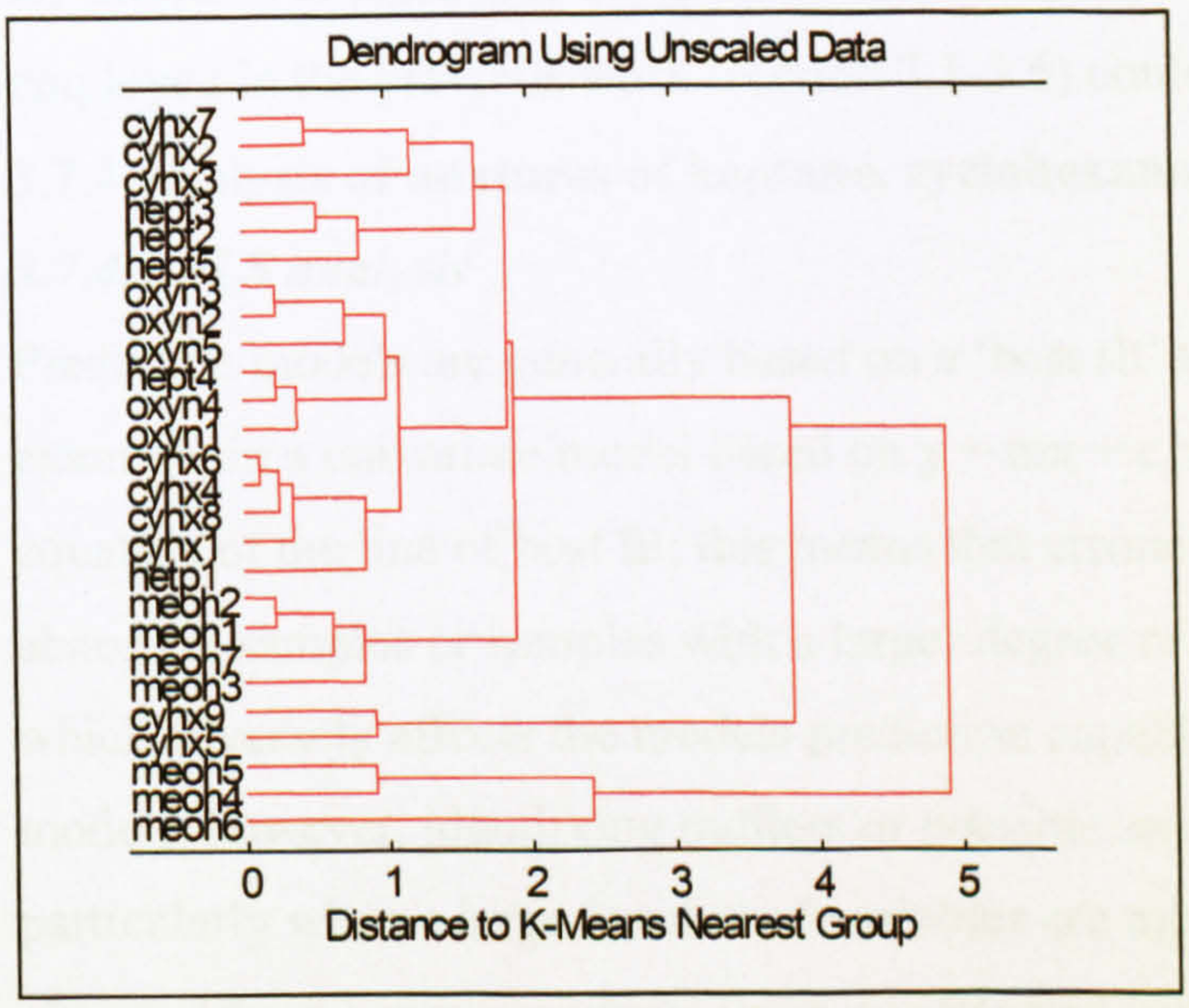


Figure 3.7.39 Dendrogram from cluster analysis of cyclohexane, ethanol, heptane and xylene

3.7.3.10 Conclusion

The injection of different amounts of liquid followed by evapouration allowed spectra of different concentrations for the compounds of interest to be produced. Scores plots produced from the analysis of the entire data set displayed definite clustering of samples. Analysis of the reduced data set (where average spectra were generated to represent each compound) resulted in clusters of alcohols, hexane and heptane, and the aromatic compounds (toluene and xylenes).

Analysis of particular groups demonstrated it was possible to differentiate between different alcohols, and between different isomers of xylene, although the spectra for hexane and heptane were too similar to achieve any distinction between samples. A drawback of employing clustering to distinguish between the different spectra is the dependence on concentration, as demonstrated by the inability to separate different low concentration spectra. The score plots show the amount of separation between samples corresponding to increasing concentration, with low concentration samples centred around the origin.

The score plot of PC2 versus PC4 from the analysis of four different groups (hexane, cyclohexane, ethanol and o-xylene) displays the increasing separation of more concentrated samples on a different axis for each compound (refer to figure 3.7.39). Therefore was decided to focus on these four compounds and build prediction models for mixtures of these four compounds and determine whether the VS-MLR algorithm employed in the previous work (section 3.1-3.6) could be applied in a similar manner.

3.7.4 Analysis of mixtures of heptane, cyclohexane, ethanol and o-xylene

3.7.4.1 PLS analysis

Prediction models are generally based on a 'best fit' to describe trends in data, for example, in a univariate model based on $y = mx + c$, m and c are determined from the equation of the line of best fit; this means that erroneous samples (or outliers) i.e. abnormal samples or samples with a larger degree of error can have a leverage effect, which adversely affects the models prediction capability. This is also true for MLR models; however, identifying outliers or possible erroneous samples can be difficult, particularly when a large number of variables are measured. In this case, the IR spectra obtained from vapour mixtures were investigated using PLS.

The first set of data examined was the FTIR spectra obtained from mixtures of cyclohexane, ethanol, heptane and o-xylene; this consisted of the training data (22 samples each containing 1801 absorption measurements (from 4000 cm^{-1} to 400 cm^{-1} with a resolution of 8 cm^{-1} and measurements taken every 2 cm^{-1})) and the validation data (11 samples collected a couple of days later, using same conditions as training data). Plots of the data with and without mean centring are contained in figure 3.7.40.

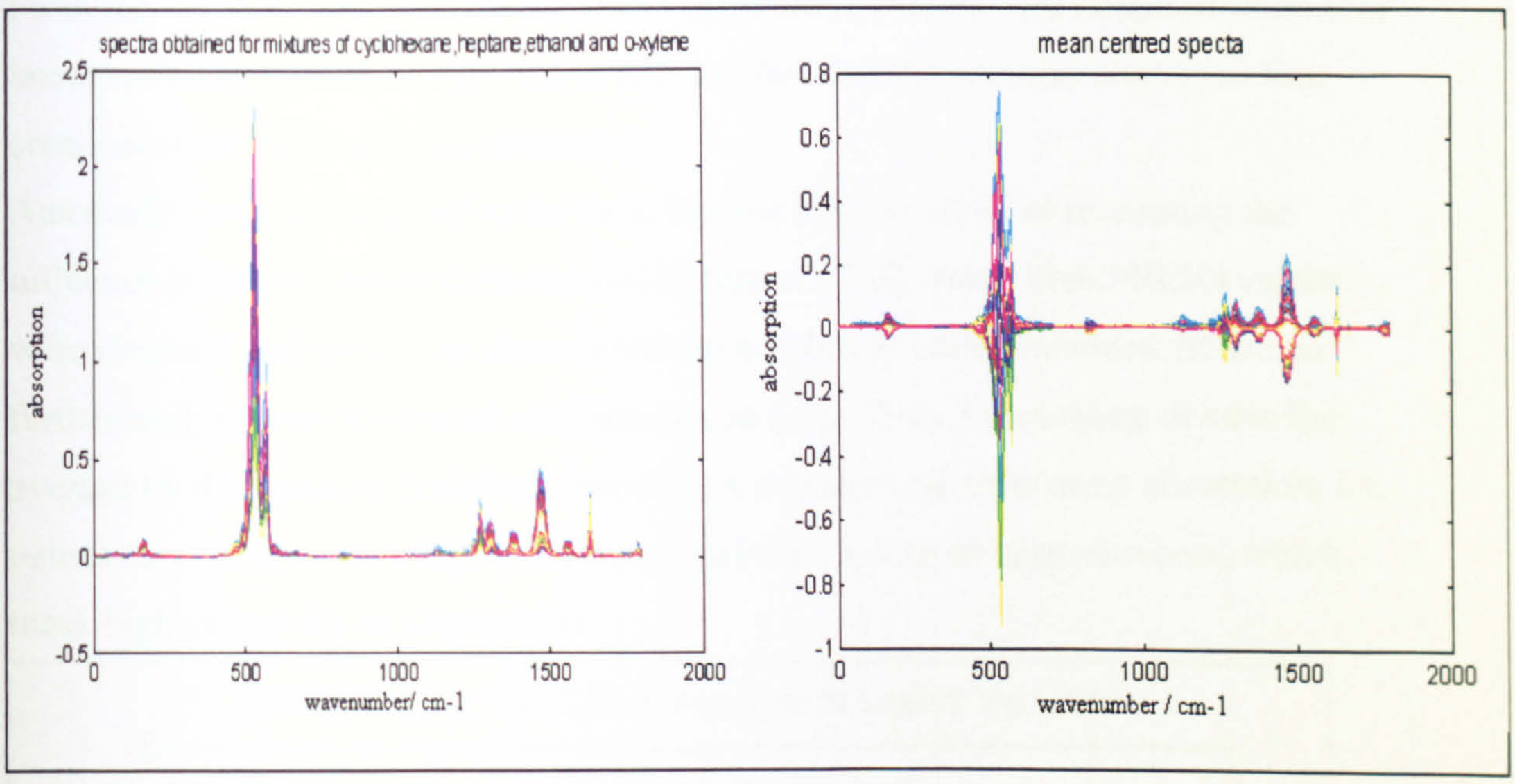


Figure 3.7.40 Spectra obtained for mixtures of cyclohexane, heptane, ethanol and o-xylene, left: absorption spectra, right: absorption spectra after meaning centring applied

PLS was performed, employing "leave one out at a time" cross validation, with no scaling and then with mean centring. PRESS (predicted residual error from sum of squares) was examined for models employing up to 10 latent variables. Figure 3.7.41 shows PRESS as a function of the number of latent variable employed in the model.

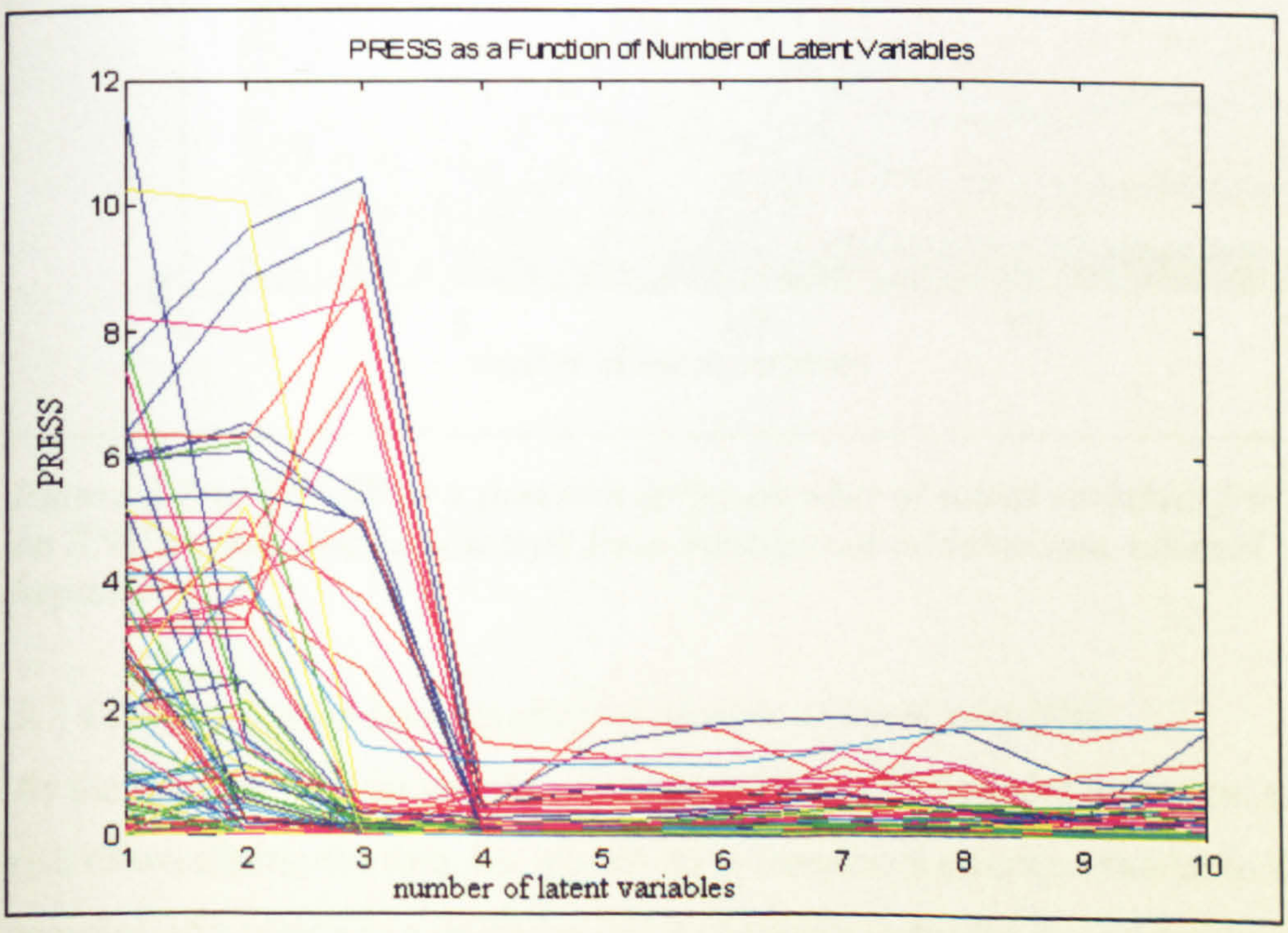


Figure 3.7.41 PRESS as a function of the number of latent variables for PLS models on spectra obtained from mixtures of cyclohexane, ethanol, o-xylene and heptane

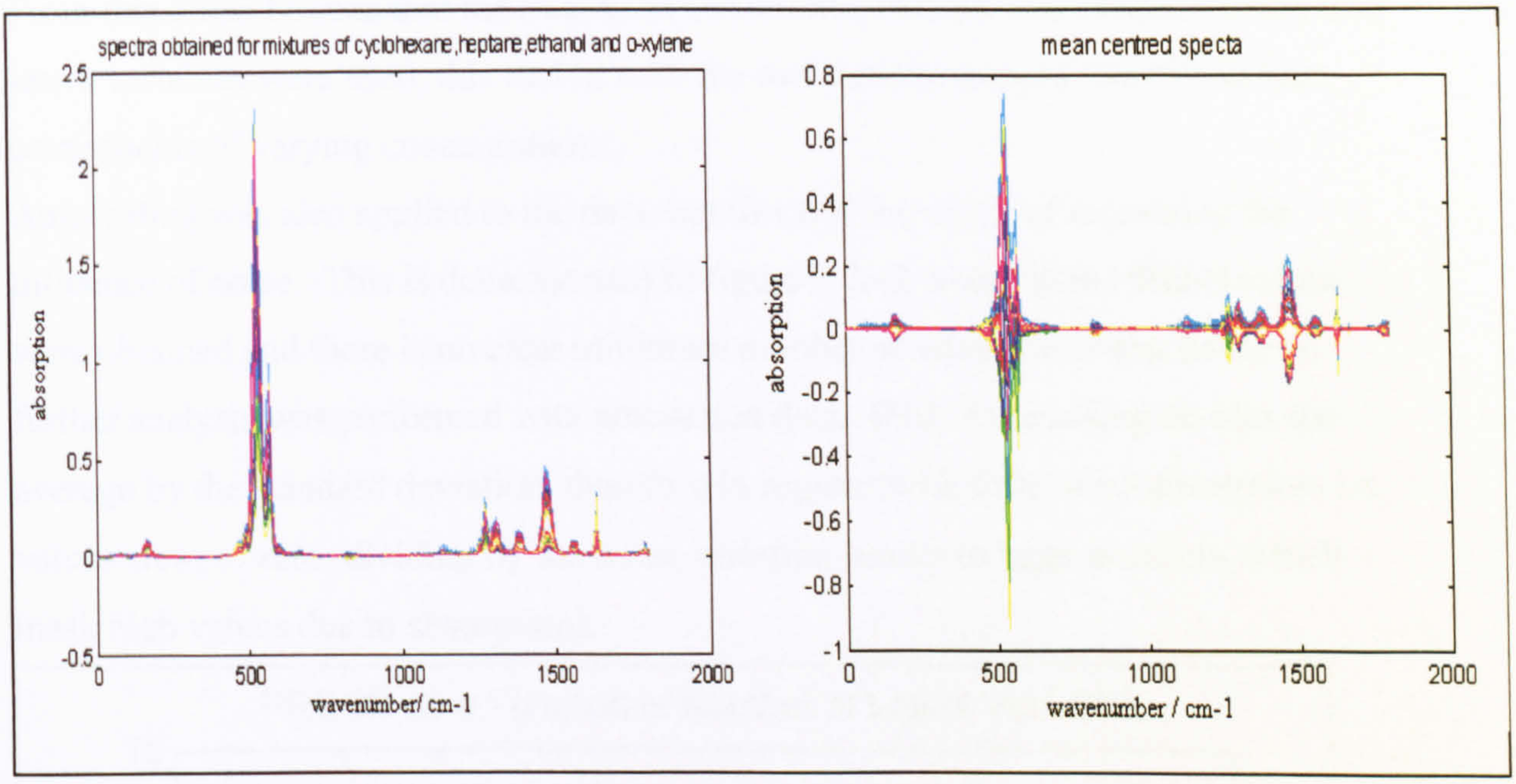


Figure 3.7.40 Spectra obtained for mixtures of cyclohexane, heptane, ethanol and o-xylene, left: absorption spectra, right: absorption spectra after meaning centring applied

PLS was performed, employing "leave one out at a time" cross validation, with no scaling and then with mean centring. PRESS (predicted residual error from sum of squares) was examined for models employing up to 10 latent variables. Figure 3.7.41 shows PRESS as a function of the number of latent variable employed in the model.

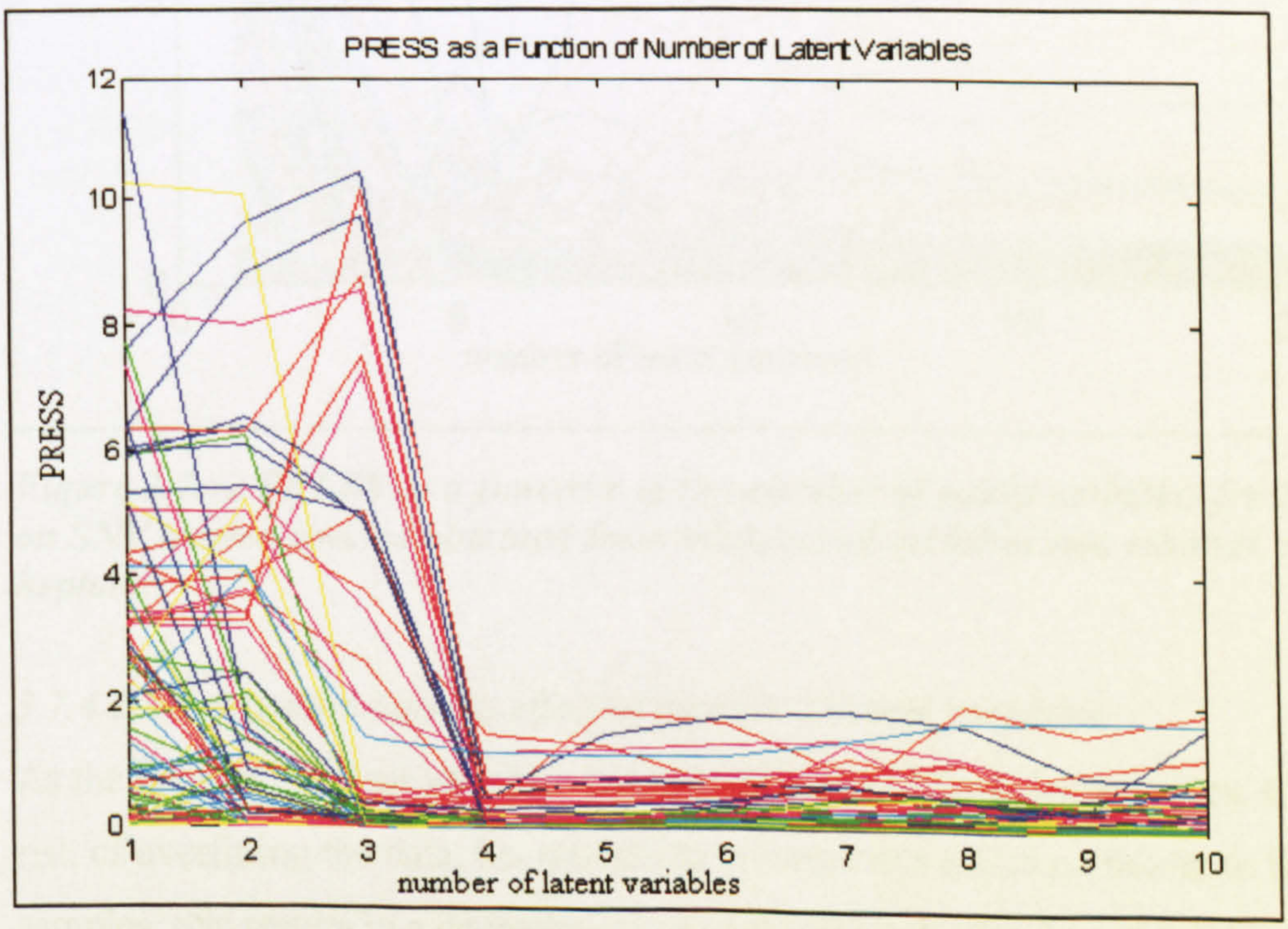


Figure 3.7.41 PRESS as a function of the number of latent variables for PLS models on spectra obtained from mixtures of cyclohexane, ethanol, o-xylene and heptane

From this it can be seen that the minimum cumulative PRESS was obtained when four latent variables were used, this tied in with the fact that the samples contained four components of varying concentrations.

Autoscaling was also applied to the data, but this had the effect of increasing the influence of noise. This is demonstrated in figure 3.7.42 where high PRESS values were obtained and there is no clear minimum number of latent variables; hence no further analysis was performed with autoscaled data. (NB. Autoscaling divides the average by the standard deviation, therefore in regions with little or no absorption, i.e. values close to zero, dividing by the noise variation results in large numbers, which mask high values due to absorption).

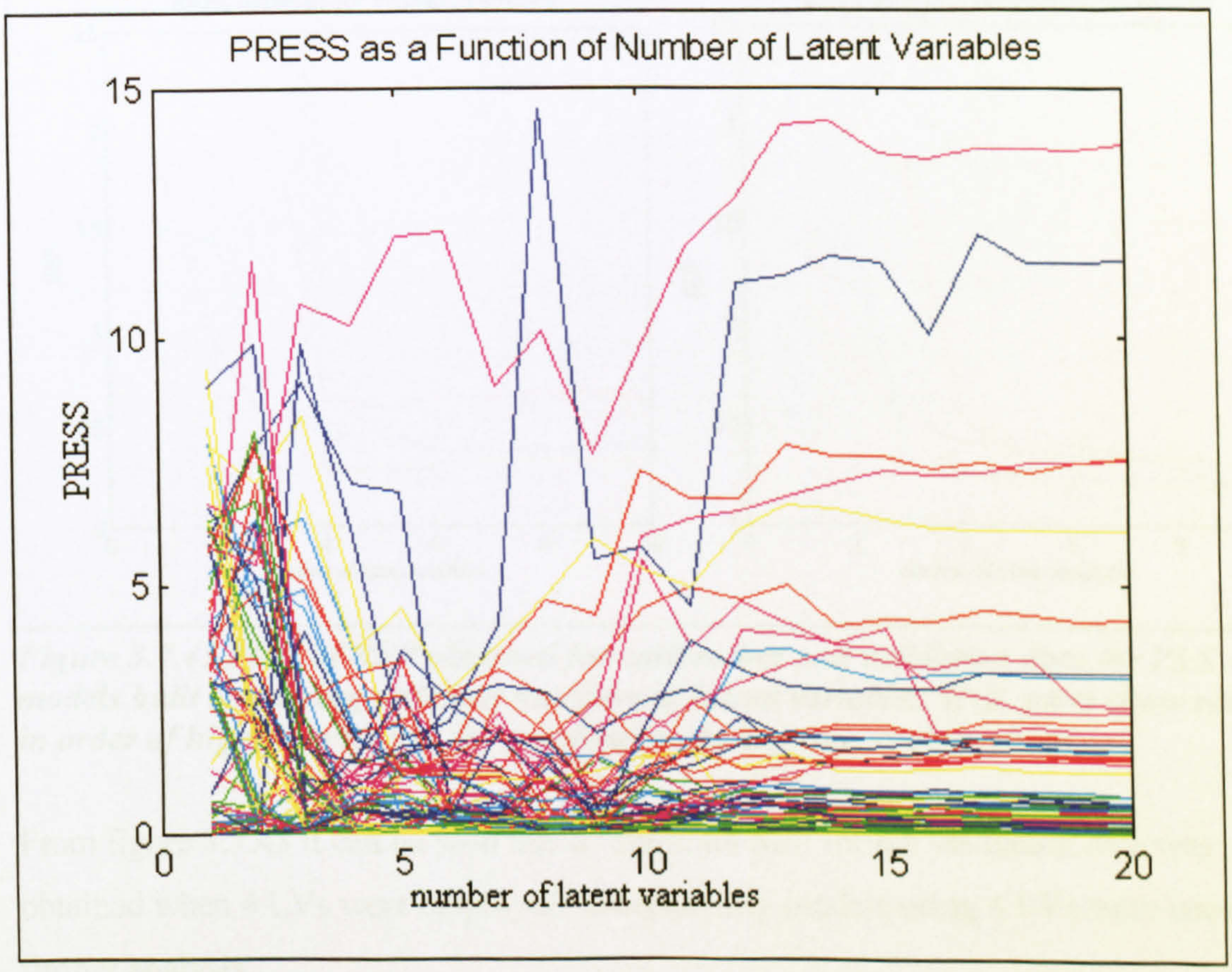


Figure 3.7.42 PRESS as a function of the number of latent variables for PLS models on SNV treated spectra obtained from mixtures of cyclohexane, ethanol, o-xylene and heptane

3.7.4.2 Investigation into the effect of number of latent variables

As the number of latent variables incorporated in a PLS model increases, so does the risk of overfitting the data, i.e. modelling incorporates errors particular to the calibration samples, this results in a decrease in the prediction error for the calibration samples (i.e. the data employed to build the PLS model), but an increase in the prediction error for

any new independent validation data. Figure 3.7.43 displays the SEPs obtained for the calibration and validation data from PLS models built employing different numbers of latent variables. This shows that the prediction error for the calibration data decreases, as the number of latent variables increases, initially due to the latent variables describing variation arising from the analytes of interest, then later due to overfitting of the data. The prediction error for the validation data initially decreases, as the latent variables included in the model describe variation due to the analytes of interest, till it reaches a minimum and then increases as latent variables included in the model start to describe variation unique to the original calibration data.

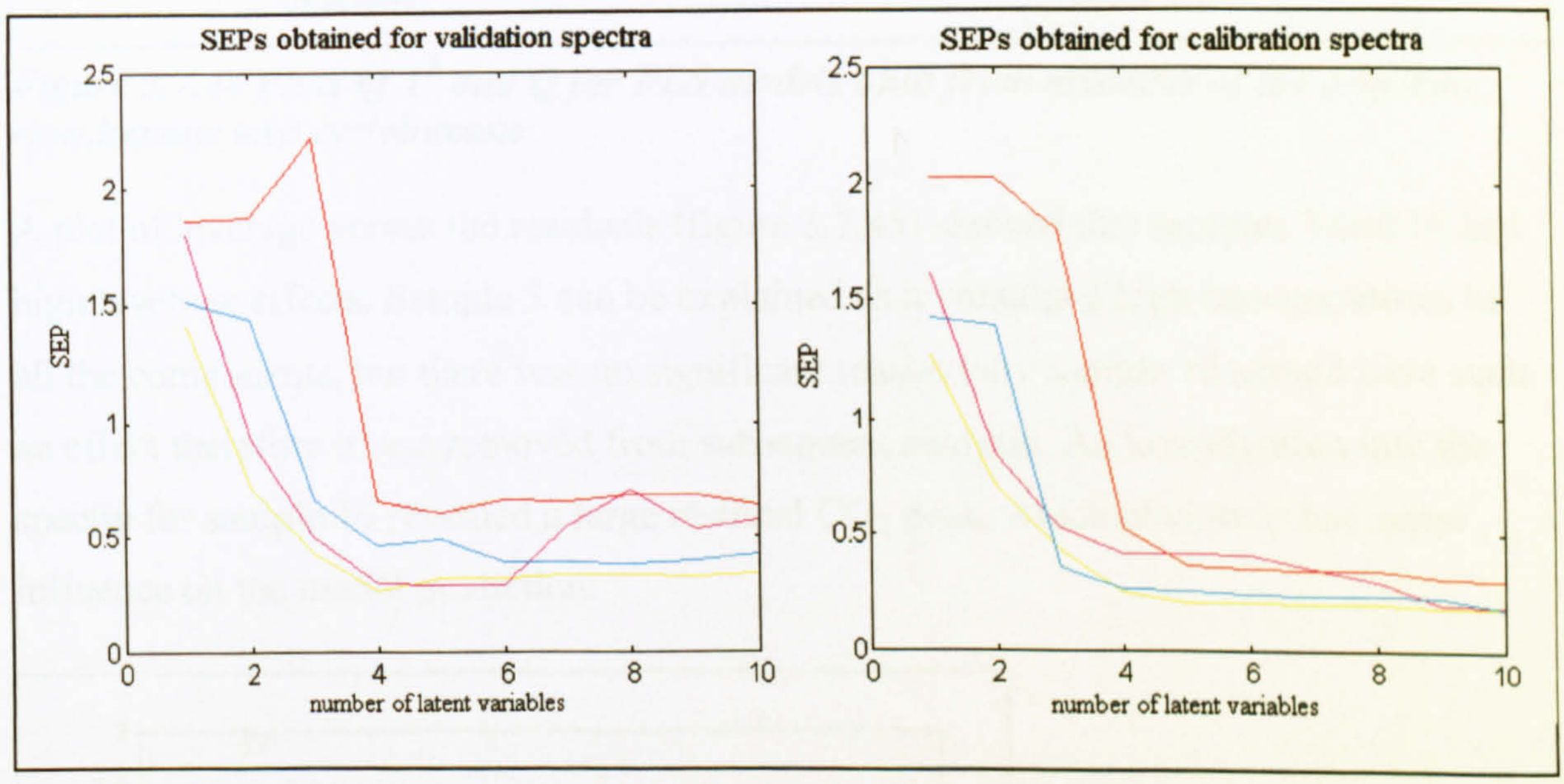


Figure 3.7.43 Plots of SEP obtained for calibration and validation data for PLS models built employing different numbers of latent variables. (NB. plots show results in order of highest to lowest for o-xylene, etoh, heptane, cyclohexane)

From figure 3.7.43 it can be seen that a minimum SEP for the validation data was obtained when 4 LVs were employed, subsequently models using 4 LVs were used for further analysis.

Plots of T^2 and Q (figure 3.7.44) for each of the samples were employed as a means of detecting outliers. In both untreated data and with mean centred data models, a large Q value was obtained for sample 3.

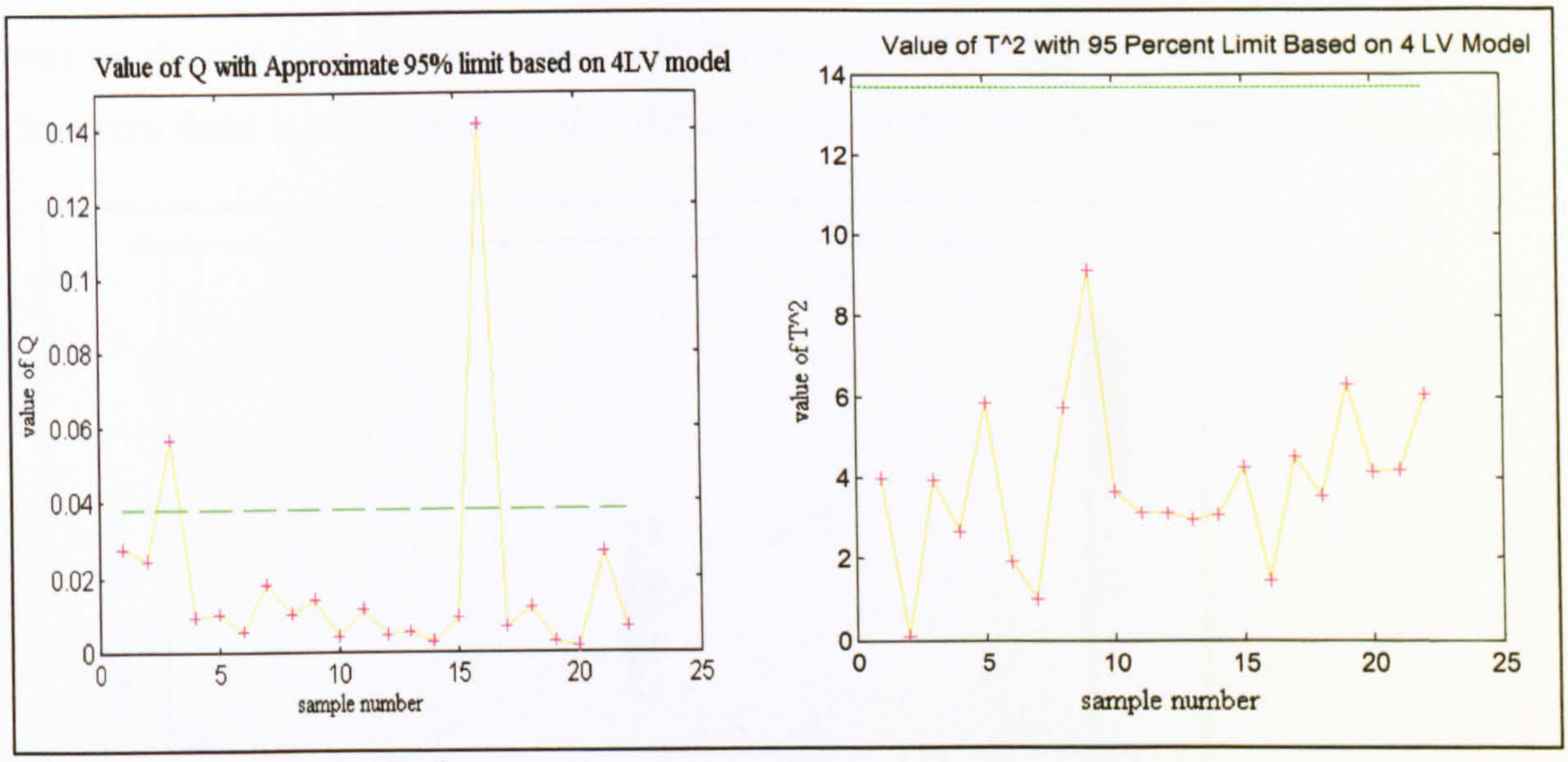


Figure 3.7.44 Plots of T^2 and Q for PLS models built from mixtures of for o-xylene, etoh,hexane and cyclohexane

A plot of leverage versus the residuals (figure 3.7.45) showed that samples 3 and 16 had high leverage effects. Sample 3 can be explained as it contained high concentrations of all the components, but there was no significant reason why sample 16 should have such an effect therefore it was removed from subsequent analysis. An investigation into the spectra for sample 16 revealed a large residual CO_2 peak, which obviously had some influence on the model prediction.

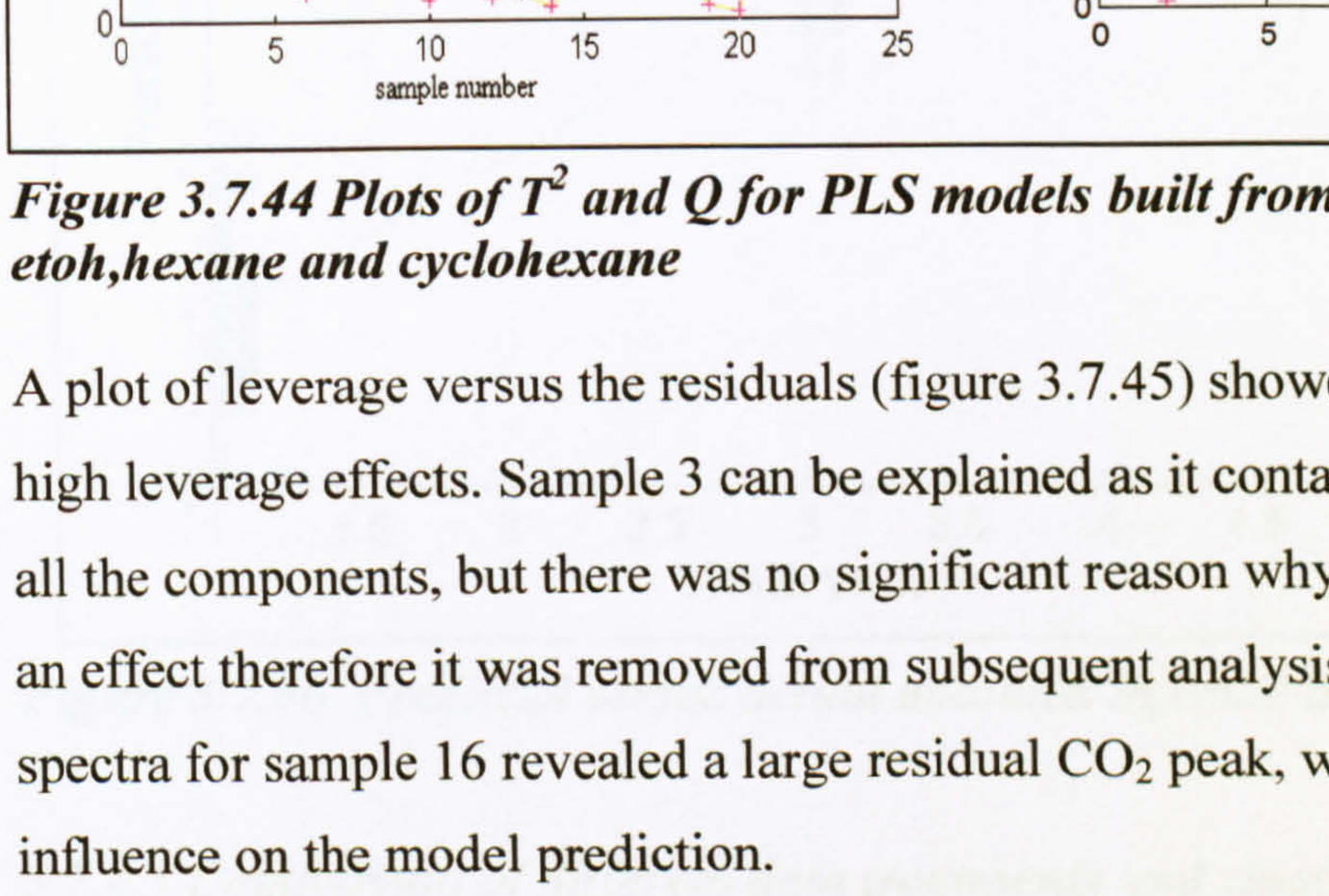


Figure 3.7.45 Plot showing leverage of each sample on the PLS model

The plot of predicted versus actual values for the model showed poor predictions for sample 10 and 11 (figure 3.7.46). The poor prediction for sample 10 turned out to be due to a wrong value entered into the concentration matrix. The error for sample 11

was for the o-xylene concentration, which is less volatile then the other 3 compounds, therefore there is the possibility that the o-xylene in this sample had not fully vaporised.

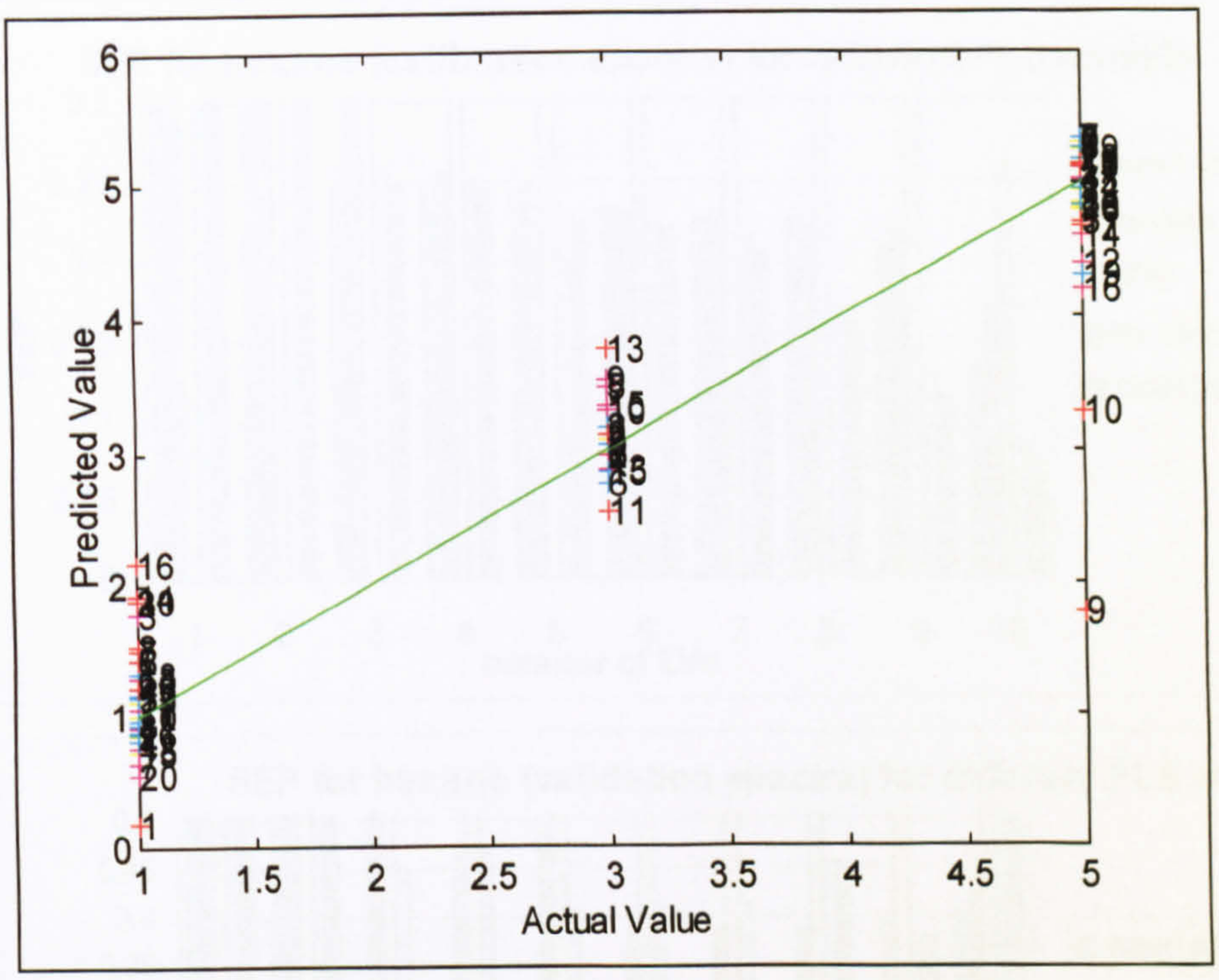


Figure 3.7.46 Predicted versus actual amounts injected into the samples

3.7.4.3 Comparison of different data treatments and single component versus all the component models

Previously the PLS models were built to predict concentrations all the components at the same time. To investigate whether individually predicting each component had any effect of the prediction error, PLS models were built separately for each component, i.e. PLS models built using just the concentration information of the component of interest. This was performed on the absorption spectra, and then on the spectra after mean centring was applied. The SEPs obtained for the validation data from PLS models employing different numbers of latents variables were compared to determine which models gave the best prediction results. Figure 3.7.47 compares the SEPs of heptane obtained for both the calibration and validation spectra from different models for up to 10 lvs.

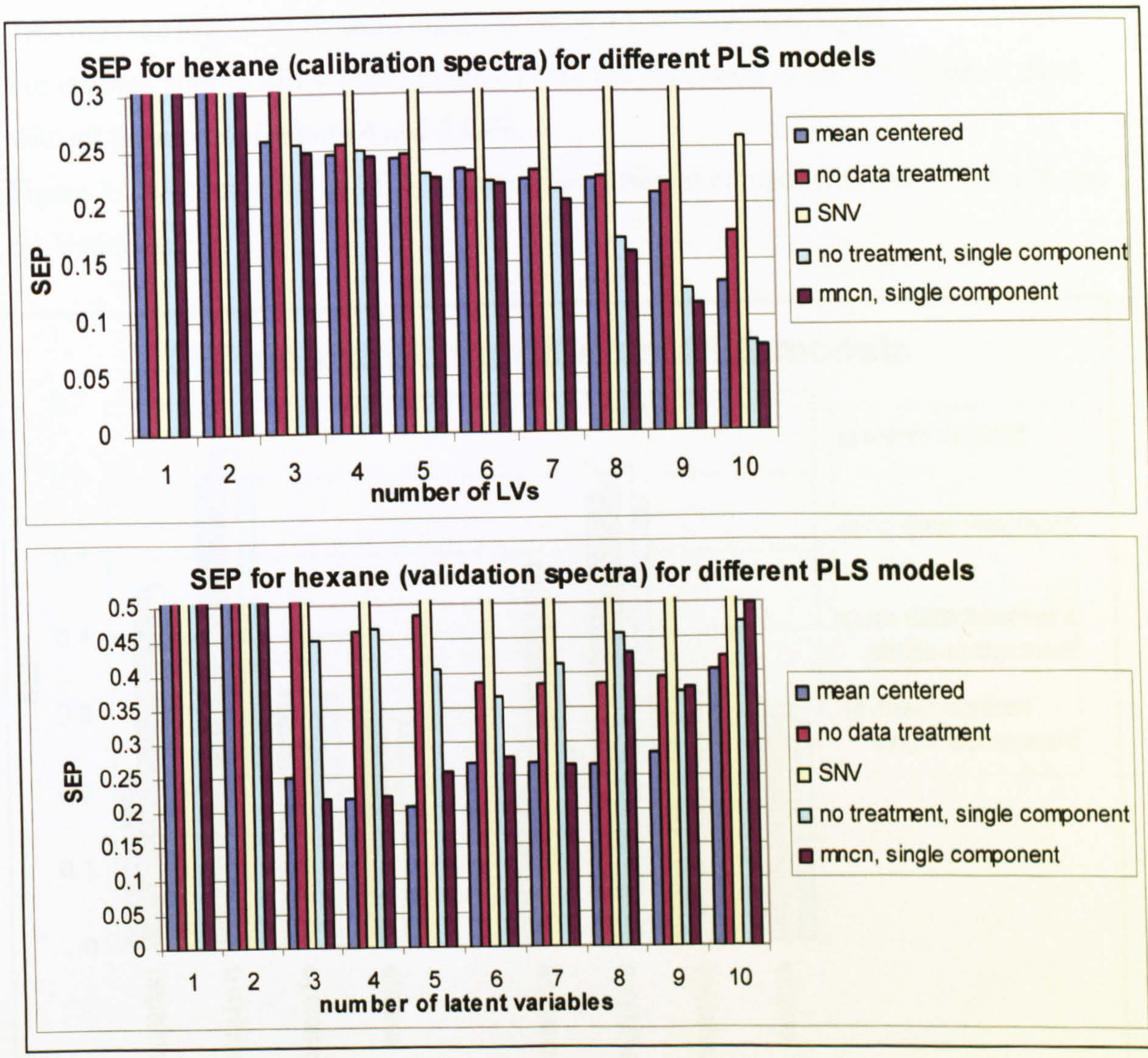


Figure 3.7.47 SEPs of hexane obtained for both the calibration and validation spectra from different models for up to 10 LVs ('single component' refers to models built only to predict hexane, otherwise models were built to predict all four component simultaneously)

The results showed that SNV treatment of the data had a detrimental effect on the effectiveness of the PLS models for prediction of all the components.

For heptane (shown in figure 3.7.47), the minimum SEP was obtained when mean centring was applied to the data (with all the components were modelled for simultaneously) and 5 LVs were employed.

In the case of o-xylene (not shown) there was no significant difference in whether the other component were modelled simultaneously or not. The lowest SEPs were obtained for the mean centred data when 5 LVs were employed. There was no signifcant

difference in whether the other components were modelled simultaneously or not for cyclohexane; however, better results were obtained when no data treatment was applied. This time the lowest SEPs were obtained when 4 LVs were employed.

For ethanol, similar SEPs were obtained (with the exception of the SNV treated data) with all the models for both 4 and 5 LVs.

Figure 3.7.48 compares the SEPs obtained for different components from the 4 LV and 5 LV models.

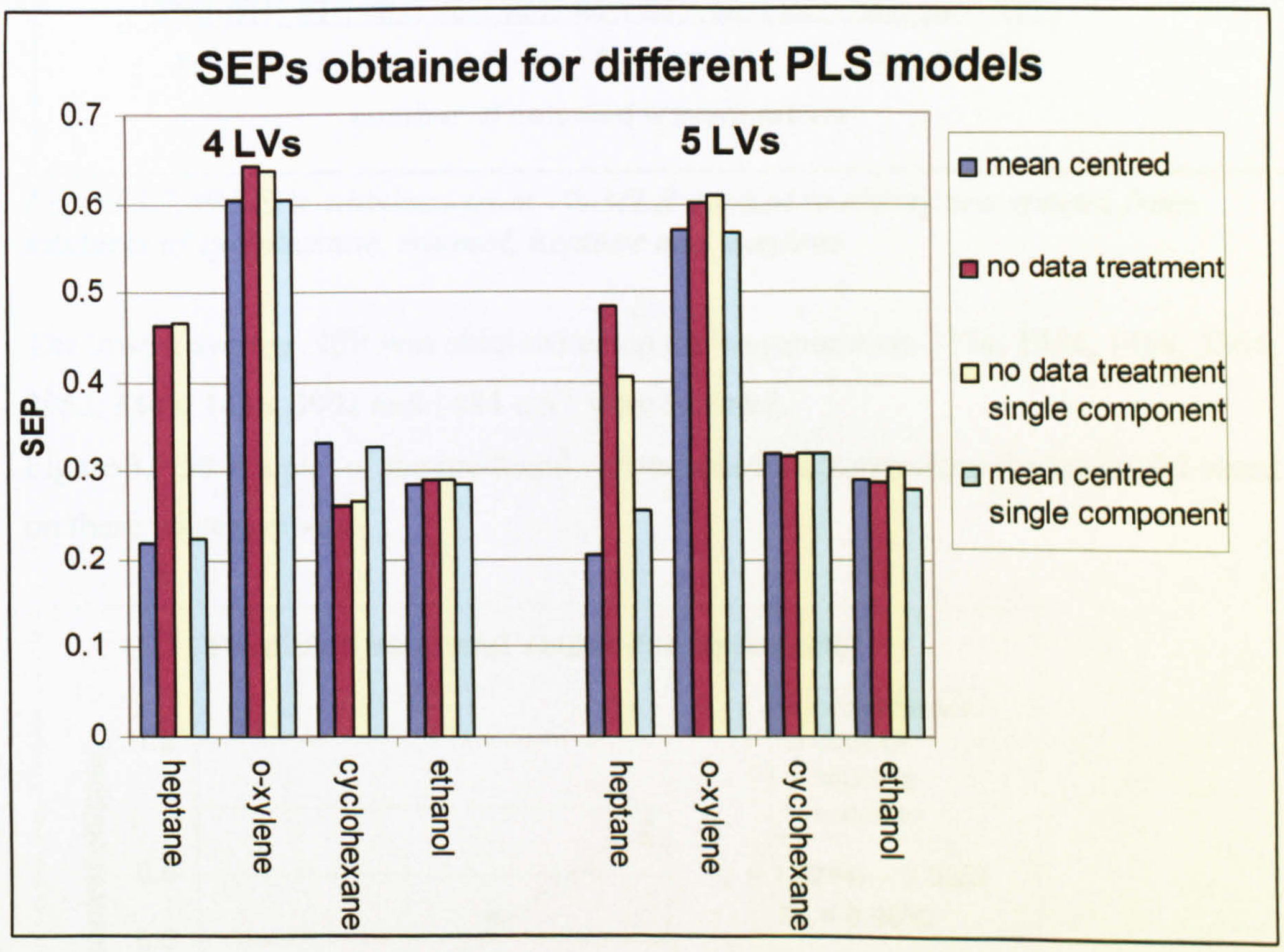


Figure 3.7.48 SEPs obtained for different components from 4 LV and 5 LV models

3.7.4.4 VS-MLR

Figure 3.7.49 displays the SEPs obtained from different selection runs using the raw absorption data. Comparing these results to the previous results (figure 3.7.48) obtained from the PLS models it can be seen that similar SEPs are achieved with the selected wavenumber sets.

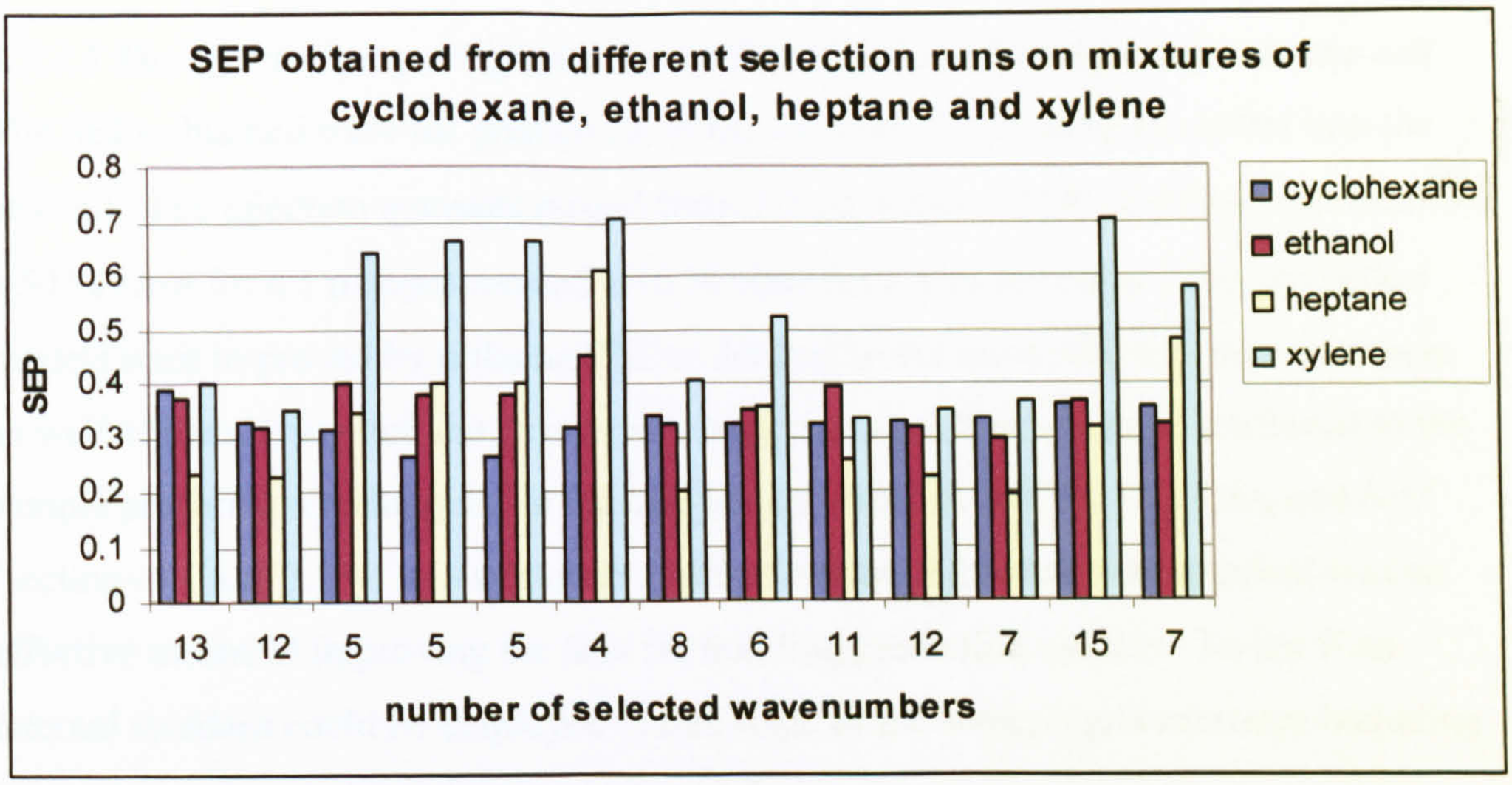


Figure 3.7.49 SEPs obtained from VS-MLR applied to absorption spectra from mixtures of cyclohexane, ethanol, heptane and o-xylene

The lowest average SEP was obtained when the wavenumbers 3754, 1454, 1458, 3864, 2952, 1448, 1456, 3902 and 1444 cm^{-1} were selected.

Figure 3.7.50 is a plot of the predicted versus actual concentrations for the model based on these wavenumbers.

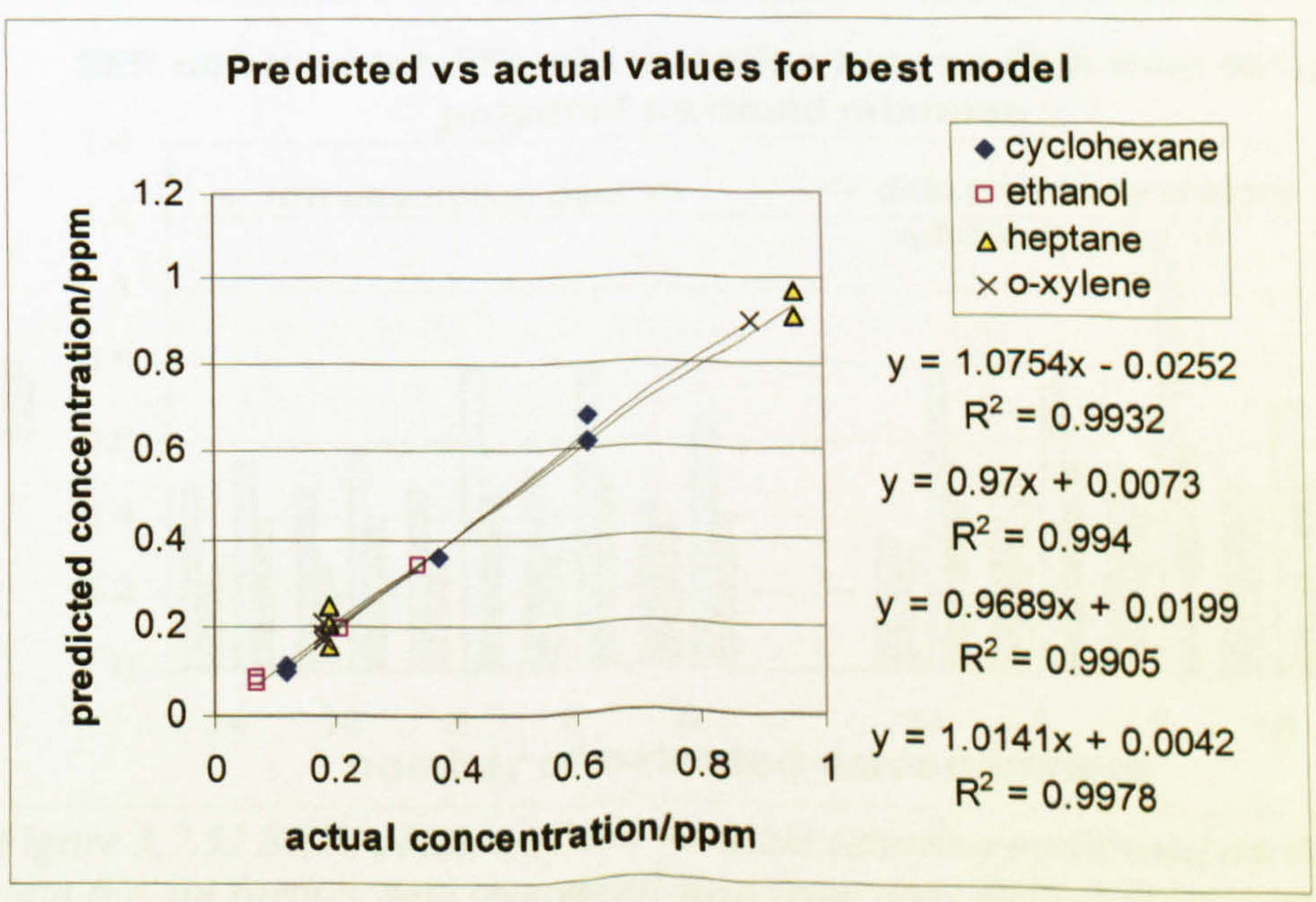


Figure 3.4.50 Predicted versus actual concentrations for model built from selected wavenumbers: 3754, 1454, 1458, 3864, 2952, 1448, 1446, 1456, 3902 and 1444 cm^{-1}

3.7.4.5 The effect of preparing mixtures in liquid form before injecting into the cell

The SEPs obtained were for predictions in the amount of component injected into the gas cell. The injection amounts ranged from 1-5 μl, hence a SEP of 0.5 corresponds to a 50 % error for a 1 μl injection and a 10 % error for a 5 μl injection. The prediction models were improved by utilising PLS to determine the most effective data treatment as well as identifying outliers; however, a large amount of error can be attributed to the sample preparation technique. In the analysis of mixtures of CH_4, CO, CO_2 and N_2O (sections 3.1 and 3.2) it was demonstrated that employing an internal standard was an effective means of improving the data for building prediction models. To see if an internal standard could be employed in this stage of the work, liquid mixtures including acetone as an internal standard were prepared.

Figure 3.7.51 compares SEPs obtained from variable selection runs performed on the raw absorption data (i.e. no further data treatment) with those from data divided by an acetone peak. From this it can be seen that higher SEPs were obtained when the data was divided by the reference peak, hence it can be concluded that the addition of an internal standard was not successful in lowering the prediction errors.

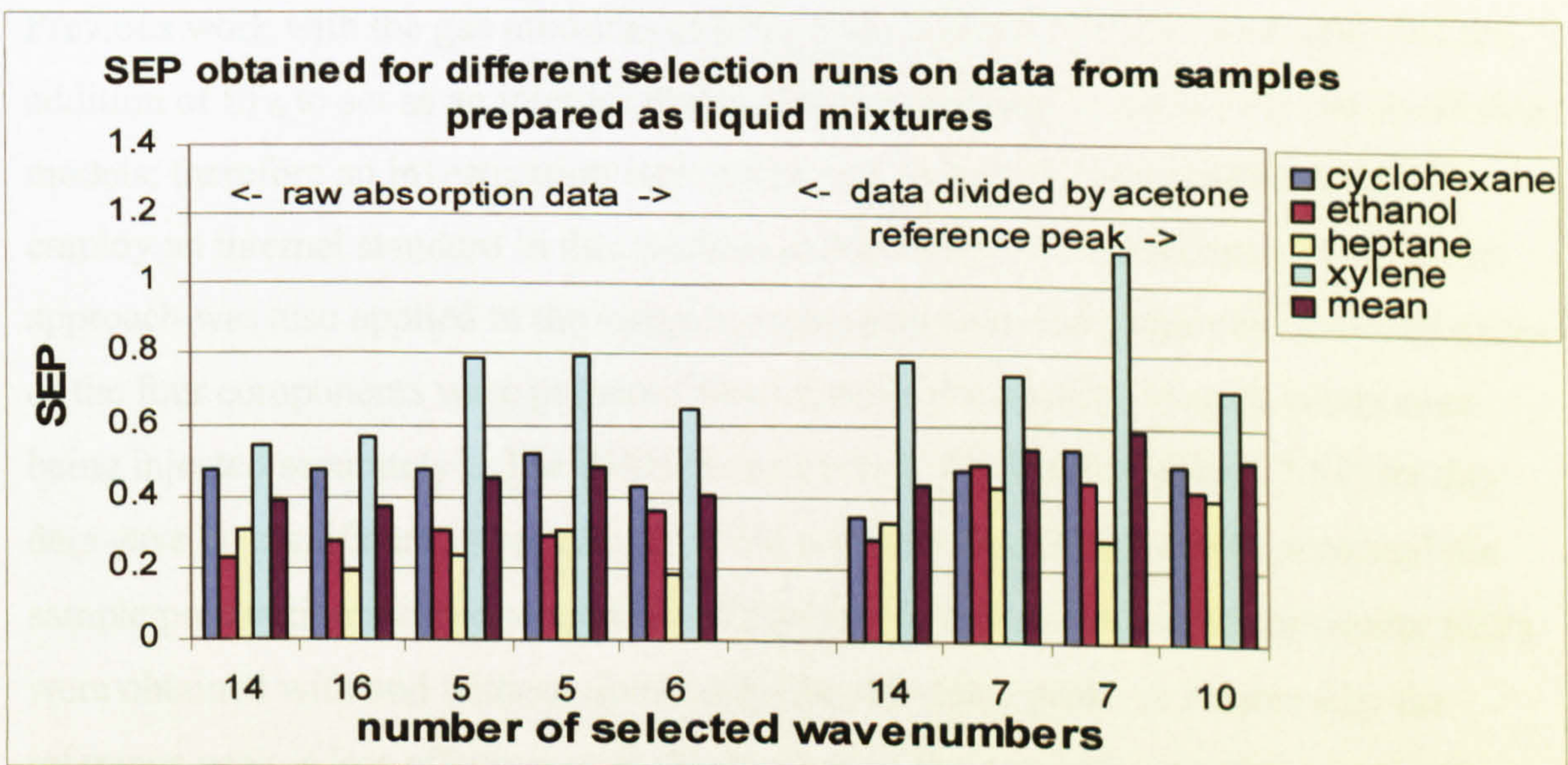


Figure 3.7.51 SEPs obtained from variable selection performed on the raw absorption data (i.e. no further data treatment) and from data divided by an acetone reference peak

3.7.4.4 Conclusion

PLS investigations on the data determined the optimum number of latent variable to employ was four; this corresponded to the fact that four compounds (heptane, cyclohexane, ethanol and xylene) of varying concentrations were contained in the sample. It was also possible to identify that sample 16 had a abnormally high CO_2 absorption peak and samples 10 and 11 gave poor predictions for o-xylene concentrations. It followed that sample 11 had a wrong concentration value entered in the concentration matrix and sample 10 was assumed to be due to experimental error.

A comparison of SNV, mean centring, and absorption spectra with no further treatment showed that mean centring was the most effective means of lowering errors in prediction, whereas SNV had a detrimental affect. It was also found that there were only slight differences between models built to individually predict single components and those built to predict all four components simultaneously.

VS-MLR performed on the data gave similar SEP levels to those obtained from PLS models with the number of variables selected ranging from 4 to 15. The lowest errors in prediction were obtained when the wavenumbers; 3754, 1454, 1458, 3864, 2952, 1448, 1446, 1456, 3902 and 1444 cm^{-1} were selected.

Previous work with the gas mixtures of CH_4, CO, CO_2 and N_2O demonstrated that the addition of SF_6 to act as an internal standard was successful in improving the prediction models; therefore an investigation was carried out to see if it was possible to also employ an internal standard in this work to lower the errors in prediction. A different approach was also applied to the sample preparation technique whereby liquid mixtures of the four components were prepared then injected (as opposed to each component being injected separately). The SEPs obtained from VS-MLR (figure 3.7.51) on this data gave no significant improvement to the previous data thus it can be assumed the sample preparation technique made no difference. It can also be seen that similar SEPs were obtained with and without division by the reference peak. A reason why the reference peak is less effective with this work than the previous gas mixtures (section 3.2) is that this data was collected as absorption spectra therefore was ratioed against a background scan while the gas mixture data relied on the reference peak to compensate for background variations.

4 Conclusions and Future Work

4.1 Conclusions

As stated in the introduction, this work was part of a project to develop a solid state IR gas analyser system based on an array of QWIP detectors. This would lead to the development of simple, inexpensive, portable systems with possible applications including monitoring of gases at landfill sites, process control, stack and exhaust fumes.

The main crux of this work was the determination of suitable sets of wavelengths to measure selected ranges of concentrations in such a detector array, and the building of a prototype system to prove the feasibility of constructing an analyser system at the chosen wavelengths. Throughout this work, chemometrics was applied as the means of analysing data and problem solving.

The method of determining suitable wavelength sets was to apply a variable selection algorithm to FTIR spectra. The chosen method of selection for this work was VS-MLR. The different approaches to variable selection and the advantages of VS-MLR are discussed in section 1.3.4.

The first stage of the work involved the collection of spectra suitable for model building. The gas samples analysed were prepared using a simple injection method as this was a fast, cost effective way to prepare samples of varying concentration combinations

One of the areas investigated was the determination of a method to compensate for background variations without requiring the collection of separate background scans. This involved applying different data treatment techniques and comparing the prediction errors obtained from models built from the differently treated data. Two particular data treatments applied were Savitsky-Golay smoothing and reconstruction of the data employing PCA. Applying the VS-MLR algorithm to the Savitsky-Golay smoothed data generally gave lower SEPs than for the raw untreated data; however an examination of the selected wavenumber sets indicated the majority of the selected sets were unsatisfactory as they either contained too many wavenumbers or wavenumbers that did not relate to absorption regions of the gases of interest (section 3.1.1.2). It was concluded that the main reason for this was that the techniques were not compensating for background variations and that these variations were still having a major influence on the selection process.

In the case of the data reconstructed from PCA the variables selected from the treated data were closer to absorption regions; however, the SEPs obtained were worse than those obtained with raw data (section 3.1.1.3): this lead to the approach of employing an internal standard to provide a reference peak that the data could be ratioed against. The employment of an internal standard initially gave mixed results (section 3.1.1.4), but its worth was shown when analysing replicate samples made during different weeks. This work demonstrated that dividing the data by a reference peak, followed by normalisation, was effective in removing a large amount of the variation caused by changes in the baseline (section 3.1.2.5), and lent weight to the idea of employing an internal standard as a means of compensating for background variation.

Having found a method which successfully reduced the effect of baseline variations, further analysis employing VS-MLR was performed on mixtures of CH_4, CO, CO_2 and N_2O; in this case the concentrations were in the range 0.19 – 2.08 %.

This brought out two weaknesses of the VS-MLR process, first that the values of the convergence levels had a major impact on the final variables selected and second that there was a tendency for the model to 'overfit'.

Optimising the convergence levels was not a major problem as the selection process was relatively short, taking on average only a few minutes for each selection run, thus it was possible to perform selection runs over a range of convergence levels in a short time period.

The problem of the model 'overfitting' the data was possibly due to all the spectra being collected over the same few days, hence to overcome this problem a second set of validation data collected on a different week was employed. Selection models that gave low SEPs with the original data were tested with the second set of validation data to identify models that gave low SEPs for the new validation data. From this the selected set of wavenumbers 2129, 2284, 3071, 3148 and 3488 cm^{-1} were found to give similar SEPs for both validation data sets.

To prove VS-MLR gave comparable results to other multivariate modelling techniques, PLS was employed. PLS is commonly employed for analysis of FTIR spectra and its effectiveness has been demonstrated in work by Griffiths *et al* [81] and Spiegelman *et al.*[83] PLS models were built for raw data, data divided by the SF_6 reference peak, data divided by a reference wavenumber (where no absorption occurred) and for absorption data. A comparison of the five latent variable models for each of the differently treated data showed that the absorption data gave the best results; however, when seven latent

variables were included the 'dividing by a reference point' treatment gave comparable results to the absorption data, thus proving that the data treatment was effective. Another conclusion drawn from the results was that there was very little difference obtained between data divided by the SF_6 reference peak and data divided by a wavenumber where no absorption occurred, thus SF_6 was no longer added to further samples.

To investigate the suitability of this approach of using selected wavelengths a rig was designed which included a gas dilution system. Different layouts for the rig were tested to determine what the most effective design was. A major component of the rig was the GC, which was incorporated as a reference method for measuring gas concentrations. This required the GC conditions to be optimised, as well as a LabVIEW program to be written to allow the GC response to be recorded. Once the conditions were optimised it was found that it was possible to obtain a RSD of <3.53% for GC peak height data from repeat injections of a certified standard gas mixture.

Generally smooth decay curves were obtained for both GC and IR responses; however, work showed that stopping and starting the dilution flow did have a detrimental effect on the smooth dilution of samples. Generally, the results obtained demonstrated that the dilution system built could be employed as a means of exponentially diluting a single sample (in this case a certified standard gas mixture) to allow investigations in response over a range of concentrations.

Good correlation was obtained between GC and IR responses (with the exception of CO for which the GC response was affected by N_2) thus the GC could act as a reference method for when the FTIR was replaced with other IR detectors.

The next stage of the work was to prove that an array of detectors measuring at the selected wavenumbers would be effective for measuring different concentrations of CH_4, CO, CO_2 and N_2O. As no QWIP detectors were available at the chosen wavenumbers, narrowband filters (with a 1% bandwidth) were employed instead. The filters blocked out light except at the chosen wavelengths hence combined with a general mid IR detector these could simulate the effect of detectors measuring at specific wavelengths. The results obtained demonstrated that it was possible to obtain good correlation between gas concentration and FTIR transmission response when different filters were in place (section 3.4).

This led to the development of a simple, robust optical set-up, independent of the FTIR, which was designed from available components that included a mid IR source, and

InSb, MCT and DTGS detectors (section 3.5). The gas cell and filter wheel (containing the different filters) were placed between the light source and one of the detectors and the detector response was measured. A lock-in amplifier and optical chopper were incorporated to maximise the detector signal, and results were recorded using LabVIEW. It was also found necessary to employ a lens to focus the light onto the detector. Once the best layout was determined (set-up 5), the set-up was tested using the dilution system. The results obtained with this optical set-up gave much lower correlations between response and concentrations than the previous work that employed the filters within an FTIR instrument (section 3.4).

Further investigations into the set-up highlighted problems with drift in response and stray light entering the system. To combat this problem a black box, that could be purged, was built to enclose the optical components. Results with the black box gave a significant improvement over the unenclosed set-up; this was particularly evident in the dilution runs where smooth response curves could be obtained after subtraction by the reference measurement. Good linearity between logs of the 4.35 μm filter response (after subtraction of reference response) at different CO_2 concentrations were obtained for concentrations less than 1,000 ppm; whereas the other filter responses gave less sensitivity to their corresponding gases.

Results for statically prepared samples (using concentration combinations determined by a five level experimental design) were much more disappointing. Poor correlations between GC and IR responses indicated equipment problems, but even a repeat of the work gave unsatisfactory results with a wide deviation in response for replicate concentrations of gases. To ensure the deviation in response was not caused by interference of the other gases, the reproducibility of CH_4 responses with and without the addition of the other gases was investigated. This determined that the problem was not due to inference between the different gases.

To further explore the effect of errors caused by the sample preparation technique, sampling loops were introduced into the system: this reduced errors arising due to variations in the amount of sample injected into the gas cell. Wide variations in response were still seen, even with the sampling valve, and the overlap between responses for different concentration levels meant that high prediction errors were obtained for prediction models built from the data.

A comparison of results collected from the box set-up over different data collection sessions shows clear differences in the magnitude of response over time. A study of

correlations between filter responses (after subtraction of the reference response) gave higher than expected correlations between 4.35 μm and 3.25 μm, and between 4.73 μm and 3.25 μm thus indicating the data treatment was not fully compensating for drift in response. A contributing factor to difference in magnitude between sessions was the occasional removal of the chopper and detector (as these were borrowed equipment and shared with other research projects) from the set-up. Work by Fructos *et al.* which employed a similar scheme for a sensor system (i.e. IR source, chopper, long path gas cell and a filter array and detector array) for the detection of CO and SO_2 gave detection limits down to a few ppm.[47] The 4.73 μm wavelength employed in their system for measuring CO was the same as that chosen in this work, thus parallels can be drawn between results, and it can be assumed that further improvements to this system can be achieved. Due to time restraints and because this system was only an intermediate stage in the development of a solid state analyser system based on QWIPs these problems were not fully resolved.

When the mid IR detector and filters are replaced with QWIP detectors the optical set-up will be more simplified, as there will be fewer components to align. In theory the QWIP detectors will be more sensitive and combined with the fact that more light will be reaching the detectors (as the filters will no longer restrict the amount of light) the system should have increased sensitivity to changes in response due to different concentrations of analyte gases and lead to improved detection limits. Possibly the lock-in amplifier and chopper may no longer be required which would further simply the set-up and signal processing, as well as reduce the size of the set-up.

The other area of investigation carried out in this research was the feasibility of developing a system for identifying and measuring different volatile organic compounds. A major part of this work was the collection of suitable spectra on which to perform the analysis. Initial problems with low concentrations of sample with high amounts of CO_2 and water vapour peaks in the spectra were solved by employing a simple injection of the sample in liquid form and allowing it to vaporise. This was particularly effective for the more volatile compounds and good correlations between concentration and response were obtained.

Cluster analysis utilising score plots between PCs from PCA and dendrograms produced from different distancing methods were used to see if different types of compound were grouped together. These resulted in clusters of: alcohols, hexane and heptane, and the aromatic compounds (toluene and xylenes); thus demonstrating the feasibility of

developing a system to distinguish between these different groups. Analysis of particular groups displayed separation between alcohols, and between different isomers of xylene, although the spectra for hexane and heptane were too similar to achieve any distinction between samples. The clustering demonstrated a high dependence on concentration and differentiation between low concentration spectra was not always achieved. Further work in this area would be to reduce the number of wavenumbers employed in the spectra while still retaining the information required to differentiate between types. A possible application for this includes devices for the identification of vapours from unknown waste solvents e.g. dumped oil drums, or odours from landfill sites.

Other applications include devices for monitoring process analysis or stack emissions where the concentrations of known gases are measured. It was decided to focus on simple four component mixtures and investigate the possibility of employing VS-MLR to select wavenumber sets tailored for analysis of known mixtures. For this purpose, mixtures of hexane, cyclohexane, ethanol and o-xylene were employed. PLS was employed to analyse the data and was found to be effective in identifying sources of error in the data (i.e. identifying a mistake in the concentration matrix, a sample with a high CO_2 absorption peak and a possible error in sample preparation). VS-MLR was performed on the data and sets consisting of four to fifteen wavenumbers were selected which gave similar levels in SEP as those obtained by the PLS models, thus demonstrating its effectiveness in reducing the number of wavenumbers to measure.

In conclusion, this project has been successful in proving that the application of chemometric techniques including, PCA, PLS and VS-MLR can play an important part in the development of analytical instrumentation. Once the problems of removing background variation was solved, it was demonstrated that the VS-MLR algorithm could be employed on both mixture and vapour mixture FTIR data to determine combinations of wavenumbers to measure in a detector array. A simple optical instrument was built which employed narrow band filters in combination with a general mid IR detector to simulate the effect of an array of QWIP detectors measuring at the selected wavelengths and look at the practical aspects of optical layout, signal processing and general construction of a measuring system.

4.2 Future work

The main areas for further work are outlined below.

- Development and testing of QWIP detectors tailored to measure at the selected wavenumbers
- Incorporation and optimisation of QWIP detectors into the purgeable box set-up
- Further investigations into vapour analysis and determination of wavenumber sets to allow characterisation of different organic vapours

Delays in the development of processed QWIP detectors measuring at the selected wavelengths meant that the performance, detection limits and suitability of the selected wavelengths could not be proved with actual QWIP detectors. The main area for further work is in the development and testing of QWIP detectors. This will lead to the incorporation of the QWIP detectors into a scaled down purgeable box or a similar, to allow detector responses to be measured for different gas mixtures.

As mentioned in the conclusion the optical set-up using the QWIPs will be simplified (the narrowband filters, and possibly the lock-in amplifier and chopper will not be required), which will reduce any errors associated with misalignment of components, however the actual optical arrangement may have to be reconsidered depending on the final size and geometry of the processed QWIP detectors and whether mirrors and lenses need to be incorporated.

Alternatively the narrow band filters could be incorporated onto the QWIP detectors to investigate whether this improves detection by increasing the signal to noise ratio.

Another area for further investigations is the vapour analysis. This can include determining wavenumber sets to measure to allow the possible identification of unknown organic vapours as well as building further calibration models for measuring concentrations of specific mixtures.

4.3 Comments about the Application of Chemometrics

A major part of this research was the application of chemometrics to aid in the development of a solid sate analyser. In this work a variety of techniques were used at different stages and the results of their application have been discussed previously. Based on these results the following recommendations for future analysis are:

Data pre-treatment

- For transmission spectrum, it would be advised to divide the data by a reference point followed by normalisation to remove baseline variations. (N.B. This scales the data so that the point of maximum transmission is 1 and minimum transmission is 0)
- For absorption spectrum, (where data is ratioed against a background spectra to remove background effects) mean centring prior to analysis is recommended.
- In this research, smoothing, reconstructing the data from PCA and autoscaling gave no significant improvement on results, therefore suggesting there is no need to apply these treatments.

Principal Component Analysis

- Examining for score plots for different PCs is a convenient way to compare/contrast samples and identify any outliers of abnormalities with samples.

Partial Least Squares

- The most important consideration for PLS models is the number of latent variables employed. This can be checked by comparing SEPs obtained for calibration data and validation data for different numbers of latent variables and employing the number that gives low SEPs for both sets of data.

Variable Selection using Multivariate Linear Regression

- The parameters that had the most influence on the results, i.e. the variables selected, were the convergence levels. The optimum values to choose depends on the data analysed, therefore it would be recommended to determine the optimum values by using a range of values with cl 2 greater or equal to cl 1.

The number maximum numbers of variables employed also had an impact on the final result, therefore it would be recommended to employ all or as many as possible.

- The VS-MLR process was found to be prone to overfitting the data, therefore additional validation data to check for this would be recommended.

5 References

1. Fox, D.L., *Anal Chem.*, **1999**, 71, R109-R119
2 Aragón, P., Atienza, J., Climent, M.D., *Crit. Reviews Anal. Chem.*, **2000**, 30, 121-151
3. Lawrence Livermore National Laboratory,7000 East Ave., Livermore, CA
4. Jongejan, A.C., Bai, Y., Veltkamp, A.C., Wyers, G.P., Slanina, J., *Int. J. Environ. Anal Chem.*, **1997**, 66, 241-245
5. Koziel J., Jia M., Khaled A., Noah J., Pawliszyn J., *Anal. Chim. Acta,* **1999**, 400, 153-162
6. Niessner R., T*rends Anal. Chem.*, **1991,** 10, 310-316
7. "Detection and Measurement of Hazardous Gases", Cullis C.F., Firth, J.G., Smith, A.F., Heinemann Educational Books Inc., **1981**
8. Sundgren, H., Lundström, I., Winquist, F., Lukkari, I., Carlsson, R., Wold, S., *Sensors Actuators B,* **1990,** 2, 115-123
9. Llobet, E., Brezmes, J., Vilanova, X., Sueiras, J.E., Correig, X., *Sensors Actuators B,* **1997,** 41, 13-21
10. Stefan, R., Staden, J.F., Aboul-Enein, H.Y., *Crit. Rev. Anal. Chem.*, **1999**, 29, 133-153
11. Park, C.O., Akbar, S.A., Hwang, J., *Material Chem. and Phys.*, **2002,** 75, 56-60
12. Jahnke, SA., Niemann, J., Kautek, W., Bischoff, R., Pfeiffer, C., Kossmehl, G., *Int. J. Environ Anal Chem.*, **1997,** 67, 223-236
13. Althainz P., Goschnick J, Ehrmann S., Ache H.J., *Sensors Actuators B,* **1996,** 33, 72-76
14. Jacquinot, P., Hodgson, A.W.E., Muller, B., Hauser, P.C., *Analyst*, **1999,** 124, 871-876
15. Hodgson, A.W.E., Jacquinot, P., Jordon, L.R., Hauser, P.C, *Anal. Chim. Acta,* **1999,** 393, 43-48
16. Fung, Y.S., Wong, C.C.W., *Anal. Chim. Acta*, **2002**, 456, 227-239
17. Abbas, M.N., Moustafa, G.A., Mitrovics, J., Gopel,W. *Anal. Chim. Acta* , **1999**, 393, 67-76
18. Cai, Q-Y, Park J., Heldsinger D., Hsich M-D, Zellers E.T, *Sensors Actuators B*, **2001,** 60, 121-130
19. Groves W.A., Zellers E.T., *Ann. Occup. Hyg.*, **2001,** 45, 609-623
20. Gardner, J.W., Bartlett, P.N., *Sensors Actuators B*, **1994**, 18, 210-211
21. Gostalow, P., Parsons, S.A., Stuetz, R.M., *Water Res.*, **2001**, 35, 579-597
22. Di Natale C., Zude-Sasse M., Macagnamo A., Paolesse R., Herold R., D'Amico A., *Anal. Chim. Acta,* **2002,** 459, 107-117

23. Maekawa, T.,Suzuki, K, Takada, T.Kobayashi, T., Egashira, M., *Sensors Actuators B*, **2001**, 80, 51-58
24. Yang, Y-M, Yang, P-Y, Wang, X-R, *Sensors Actuators B*, **2000**, 66, 167-170
25. Barisci, J.N., Wallace, G.G, Andrews, M.K., Partridge, A.C., Harris, P.D, *Sensors Actuators B*, 54, **2002**, 54, 252-257
26. Moeckli, M.A., Hilbes, C., Sigrist, M.W., *Appl. Phys. B*, 67, 449-458
27. Sneider, J., Bozoki, Z., Miklos, A., Bor, Z.S., Szabo, G., *Int. J. Environ. Anal. Chem.*, **1997**, 67, 253-260
28. Miklós, A., Hess, P., *Anal. Chem*, **2000**, 72, 30A-37A
29. "The Infrared Handbook", Wolfe, W.L., Zissis, G.J, Office of Naval Research, Department of the Navy, Washington D.C., 1978
30. "Gas Analysis Manual for Analytical Chemists, Infrared Measurement of Gases and Vapors", P.L. Hanst, S.T. Hanst, *Infrared Analysis, Inc.*, 1990
31. Esler, M.B., Griffith, D.W.T., Wilson, S.R., Steele, L.P., *Anal. Chem.*, **2000**, 72, 206-215
32. Kumer, P.C., Wehrmeyer, J.A., *Appl. Spectrosc.*, **1997**, 51, 849-855
33. Wehling, B., Hill W., Klockow, D., *Int. J. Environ. Anal. Chem.*, **1999**, 73, 223-236
34. Hodges, J.Kunsch, *Spectrosc. Europe*, 11, **1999**, 18-23
35. Werle P., Semr F., Maurer K., Kormam R., Müke R., Jänker B., *Optics & Lasers Eng.*, **2002** ,37, 101-114
36. Fisher, H., Hoor, P., Wiehold, Wagner, V., Konigstedt, R., Harris, G.W., *Appl. Phys.* B, 67, 411-417
37. G.Toci, Mazzinghi P., Mielke B., Stefanutti L., *Optics & Lasers in Eng.*, **2002**, 37, 429-480
38. Seiter, M., Sigrist, M., *Appl. Opt.*, **1999**, 38, 4691-4693
39. Tredicucci, A., Gmachl, C., Capasso, F., Sivco, D.L., Hutchinson, A.L., Cho, A.Y., *Nature*, **1998**, 396, 350-396
40. Krier, A., Fisher, M., *IEE Proc-optoelec.*, **1997**, 144, 287-293
41. Krier, A., Mao, Y., *Infrared Phys.Technol.*, **1997**, 38, 397-403 0
42. Smith, S.D.,Vass, A., Bramley, P., Crowder, J.G., Wang, C.H., *IEE Proc. Optoelectron.*, **1997**, 144, 266-27
43. Ahonon, I., Riipinen, H., Roos, A., *Analyst*, **1996**, 121, 1253-1255
44. Chen, T., *Sens. Actuators, B*, **1993**, 13-14, 284-287

45. Meléndez, J., Castro, A.J., López, F., Meneses, J., *Sens. Actuaotrs, A* , **1995**, 46-47, 417-421

46. Muñoz, A., .Meléndez, J., Torquemada, M.C., Rodrigo, M.T., Cebrián, J., Castro, A.J., Meneses, J., Ugarte, M., López, F., Vergara, G., Hernández, J.L, Mertín, J.M., Adell,L., Montojo, M.T., *Thin Solid Films*, **1998**, 317, 425-428

47. Frutos, J., Rodríguez, J.M., López, F., Castro, A.J., Meléndez, J., Meneses, J., *Sens. Actuators, B*, **1994**, 18-19, 682-686

48. Alause, H., Grasdepot, F., Malzac, J.P., Knap W., Hermann J., *Sens. Actuators, B*, **1997**, 43, 18-23

49. Myrick, M.L., Soyemi, O.O., Schiza M.V., Farr, J.R., Haibach F.G., Greer, A.E., Li, H., Priore, R.J., *Proc SPIE,* **2002**, 4574, 208-215

50. Haibach F.G., Greer, A.E., Schiza, M.V., Priore, R.J., Soyemi, O.O., Myrick M.L., *Applied Optics,* **2003**, 42, 1833- 1838

51. Soyemi, O.O., Zhang, L., Eastwood, D., Li, H., Gemerline, P.J., Myrick, M.L., *Proc. SPIE,* **2001**, 4284, 17-28

52. Goldthorp M.D., Lambert P., *J. Hazardous Materials*, **2001**, 83, 135-152

53. Levine B.F., *J Appl.Phys*, **1993**, 74, R1-R81

54. Fromherz, T., Nützel, J.F., Hertle H., Helm, M., Bauer G., Abstreiter G., *Vibrational Spectroscopy,* **1995**, 8, 109-119

55. Krier, A., Gao, H.H., Mao, Y., *Semiconductor Sci. Technol.*,**1998**, 13, 950-956

56. A.Krier, Y.Mao, *Infrared Phys.Technol*, **1997**, 38, 397-403

57. "QuantumWell Intersubband Infrared Photodetectors (QWIPs) for Gas Sensing Applications", Lai, K.T., MSc. Thesis, University of Hull, 2001

58. Lai K.T., Haywood S.K., Gupta R., Missous M., *Electronics Letters,* **2002**, 38 (11), 529-530

59. Wolfbeis, O.S., *Anal Chem,* **2000**,72, 81R-89R

60. Albert, K.J., Walt, D.R., Gill, D.S., Pearce, T.C., *Anal chem.*, **2001**, 73, 2501-2508

61. Eklov, T., Lundstrom, I., *Anal. Chem.*,**1999**, 71, 3544-3550

62. Xu Y., Jones N.B., Fothergill J.C., Hanning C.D, *Optics and Lasers in Engineering*, **2001**, 36, 607-615

63. "Gas Mixtures: preparation and control", Nelson, Gary O., Lewis Publishers, 1992

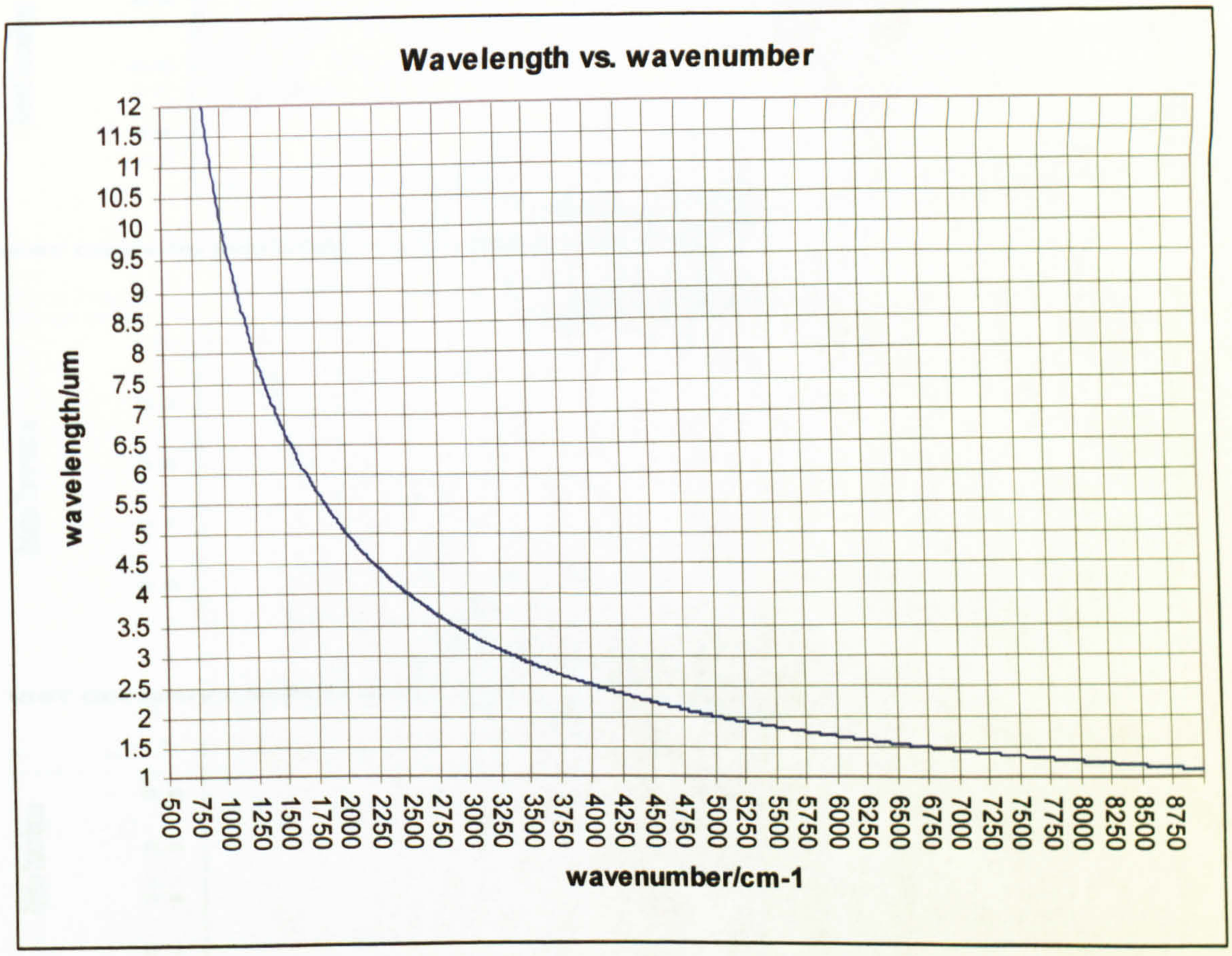
64. Chu P.M., Guether F.R., Rhoderick G.C., Lafferty W.J., *J. Research National Institute Std. Technol.*, **1999**, 104, 59-81

65. Griffiths, P.R., Richardson, R.L., *Applied Spec.,* **1998,** 52 (1), 143-153
66. "Gas Chromatography", Lovelock, J.E., R.R.W. Scott, Ed., Butterworths Inc., Washington D.C., 1960
67. Mitchell G.D., *Separation and Purification Methods*, **2000,** 29(1), 119-128
68. "Chemometrics: Experimental Design", Morgan E., John Wiley & Sons Ltd., 1995
69. "Chemometrics", Sharaf, M.A., Illman, D.L., Kowalski, B.R., John Wiley & Sons Ltd., 1986
70. Savitsky A., Golay, M.J.E, *Anal. Chem*, **1964**, 36, (8), 1627-1639
71. Estienne F., Massart D.L., Zanier-Szydlowski N., Marteau P.H., *Anal. Chim. Acta*, **2000**, 424, 185-201
72. Koehler F.W, Small G.W., Combs R.J., Knapp R.B., Kroutil R.T, *Anal. Chem.* **2000**, 72, 1690-1698
73. F.W, Small G.W., Combs R.J., Knapp R.B., Kroutil R.T, *Vibrational Spec.*, **2001**, 27, 97-107
74. Vogt F., Rebstock K., Tacke M., *Chem. Intell. Lab. Sys.*, **2000**,50, 175-178
75. "Analytical Chemistry", Kellner, K., Mermet, J.-M., Otto, M., Widmer, H.M., Wiley VCH, Weinhim, 1998
76. Hong-kui X., Levine, S.P., D'Arcy, J.B., *Anal. Chem.,* **1989**, 61, 2708-2714
77. Li-shi, Y., Levine, S.P., *Anal. chem.,* **1989**, 61, 677-683
78. Gether, J.J., Thrane, K.E., *Anal. Chim. Acta,* **1986**, 184, 99-106
79. Birring, K.S., Haswell, S.J., *Anal. Proc. Anal. Communication,* **1994,** 31, 201-204
80. Perkins J.H., Hasenoehrl E.J., Griffiths P.R., *Anal Chem.*, **1991**, 63, 1738-1747
81. "Chemometrics: A textbook", Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman, L., Elsevier Science Publishers B.V., 1988
82. Maesschalek, R., Jouan-Rimbaud, D., Massart, D.L., *Chem Intell. Lab. Sys.*, **2000,** 50, (1), 1-18
83. Walmsley, A.D., Haswell, S.J., Metcalfe, E., *Anal. Chim. Acta,* **1991,** 242, 31-36
84. Geladi, P., *Chem. Intell. Lab. Sys.,* **2002,** 60, 211-224
85. Szczurek A., Szecówka P.M., Licznerski B.W., *Sensors Actuators B,* **1999**, 58, 427-432
86. Jaakkola, P.T., Vahlman, T.A., Roos, A.A, Saatinen, P.E., Kauppinen, J.K., *Water, Air Soil Pollution*, **1998,** 101, 79-92
87. Wold, S., Sjöström, M., Eriksson, L., *Chem. Intell. Lab. Sys.*, **2001,** 58, 109-130

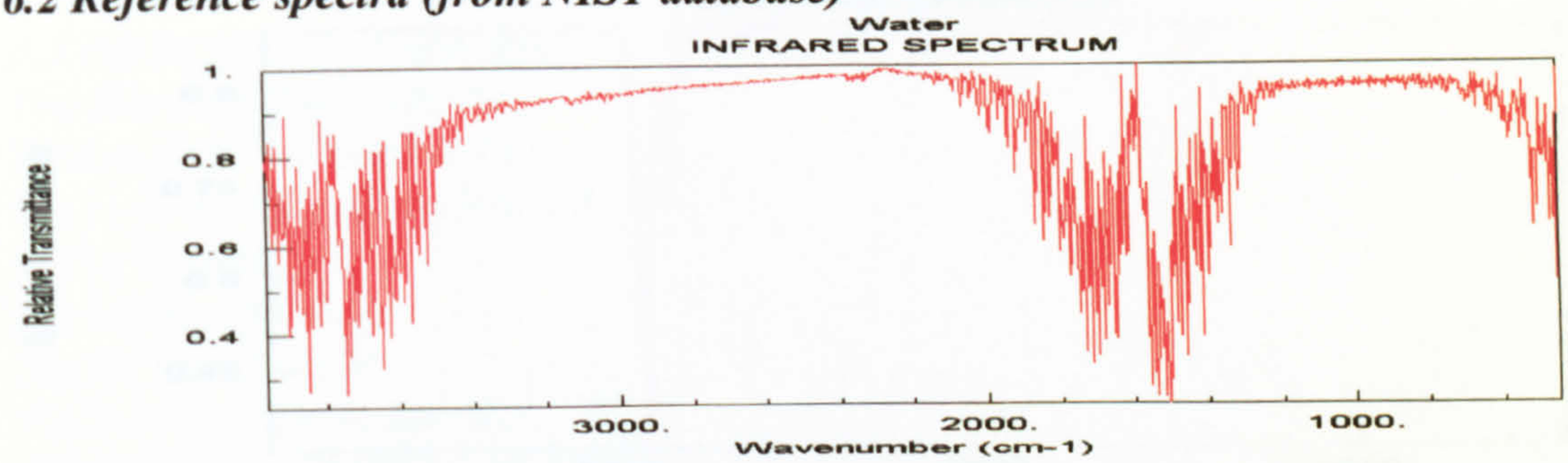
88. Johansen, I-R, Line, G.T., Honne, A., Midtgaard, T, *Applied Spect.*, **1997**, 51, 1540- 1546
89. Griffiths P.R., Hart B.K., Yang H., Berry R.J, *Talenta*, **2000**, 53, 223-231
90. Xu L., Zhang W-J, *Anal. Chim. Acta*, **2001**, 446, 477-483
91. Speigelman C.H, McShane M.J., Goetz M.J., Motamedi M., Yue L., Coté G.L., *Anal. Chem.*, **1998**, 70, 35-44.
92. Jouan-Rimbaud D., Massart D.L.,Nooad O.E., *Chem. Intell. Lab. Sys.*, **1996**, 35, 213-220
93. Broadhurst, D., Goodacre, R., Jones, A., Rowland, J.J., Kell, D.B., *Anal. Chim. Acta*, **1997**, 348, 71-86
94. Walmsley, A.D., *Anal. Chim. Acta.*, **1997**, 354, 225- 232
95. Emery, S., University of Hull, 1999
96 Kaldy, B., Technical Support, Alltech UK. Ltd., 1998
97. "The Aldrich library of FT-IR Spectra", Behnke, J., Aldrich Chemical, 1997
98. Rothman, L.S., Gamache, R.R., Goldman, A., Brown, L.R., Toth, R.A., Pickett, H.M, Poynter, R.L., Flaud, J.-M., Camy-Peyret, C., Barbe, A., Husson, N., Rinsland, C.P., Smith, A.H., *Applied Optics*, **1987**, 26, 4058-4095

6 Appendices

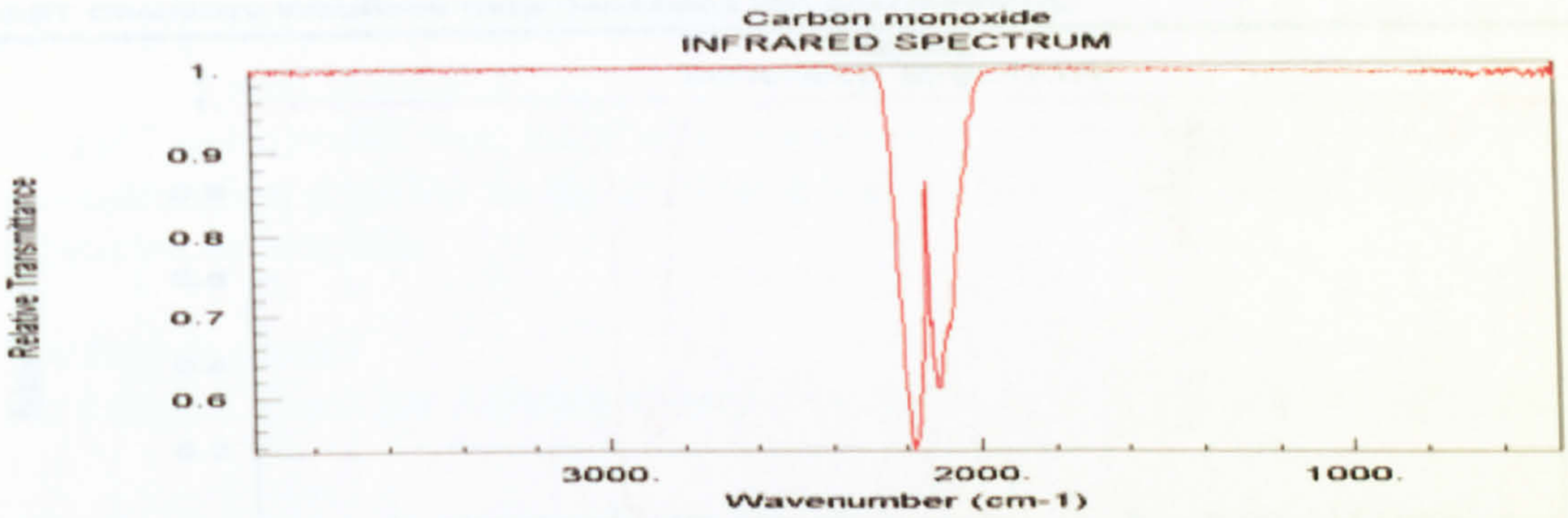
6.1 Plot of wavenumbers versus wavelengths



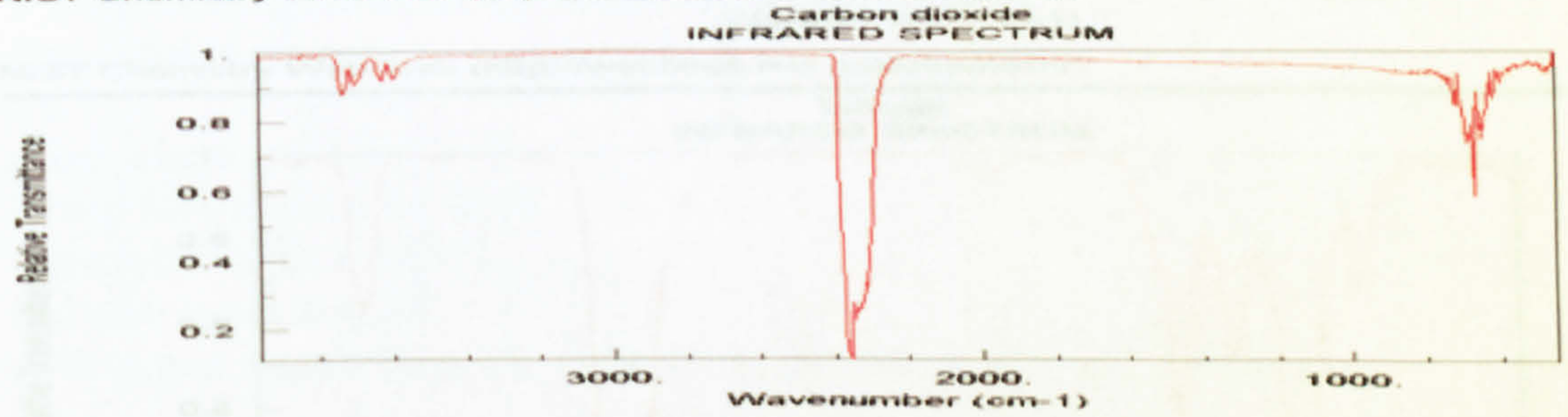
6.2 Reference spectra (from NIST database)



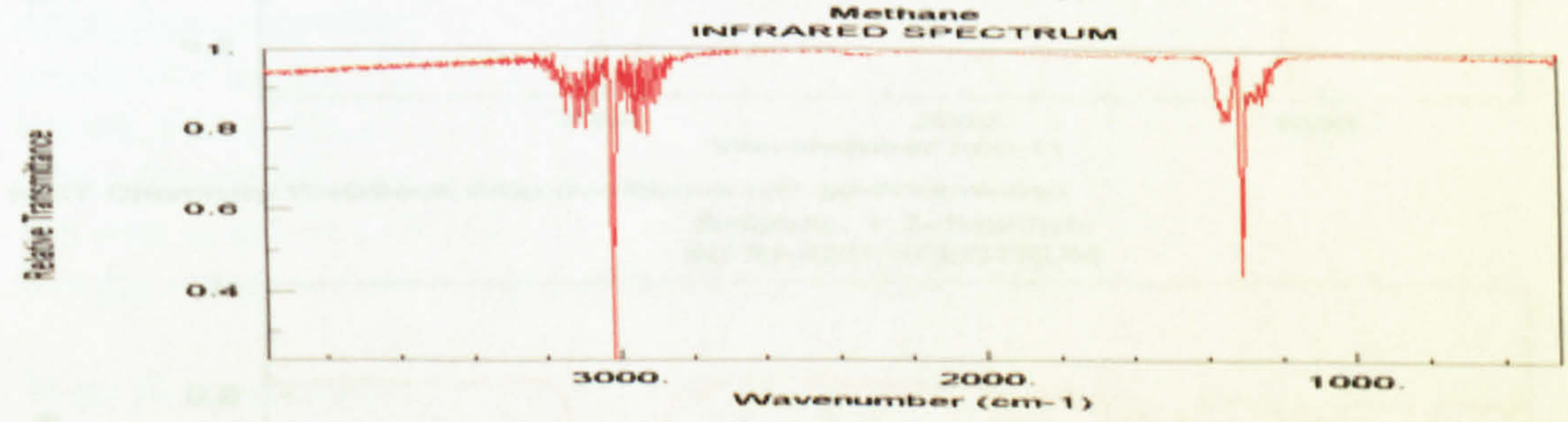
NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)



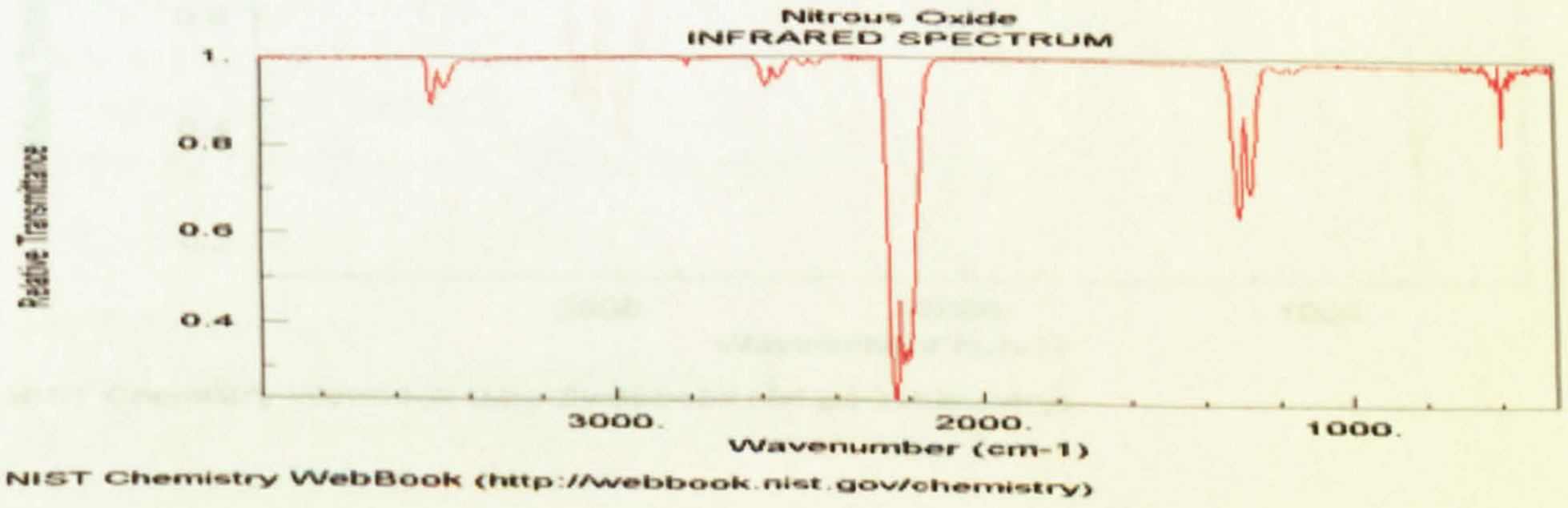
NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)



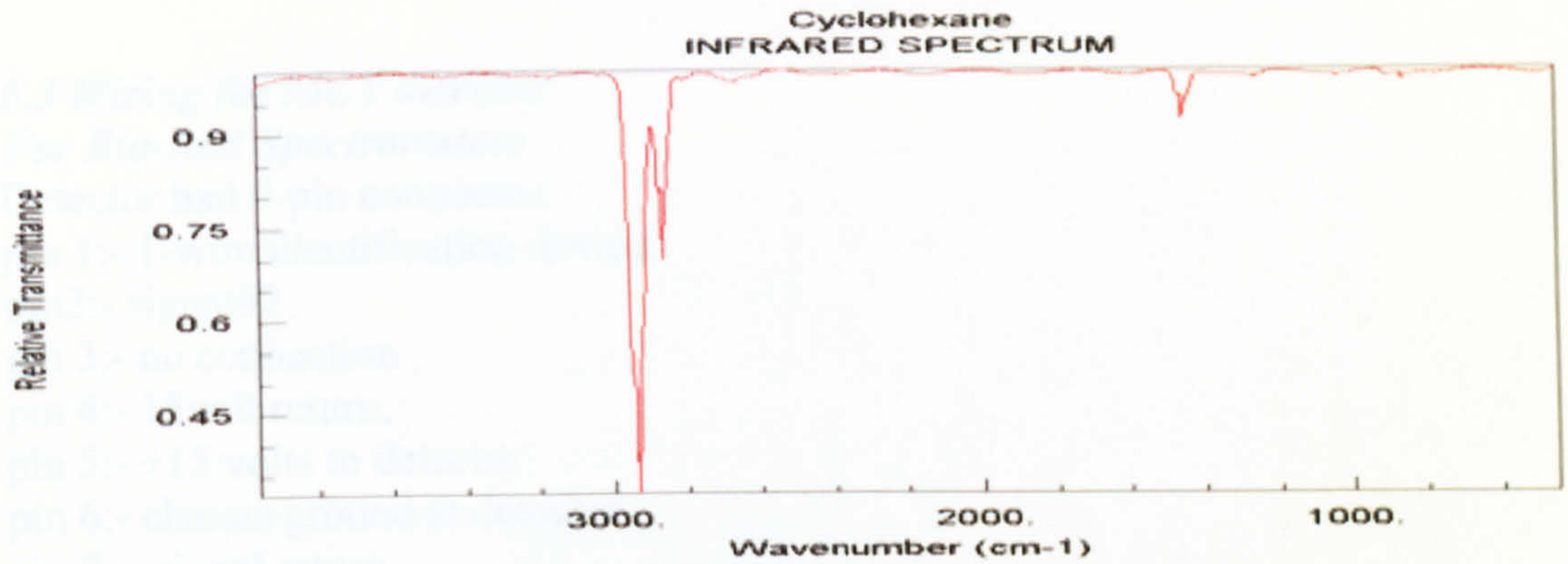
NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)



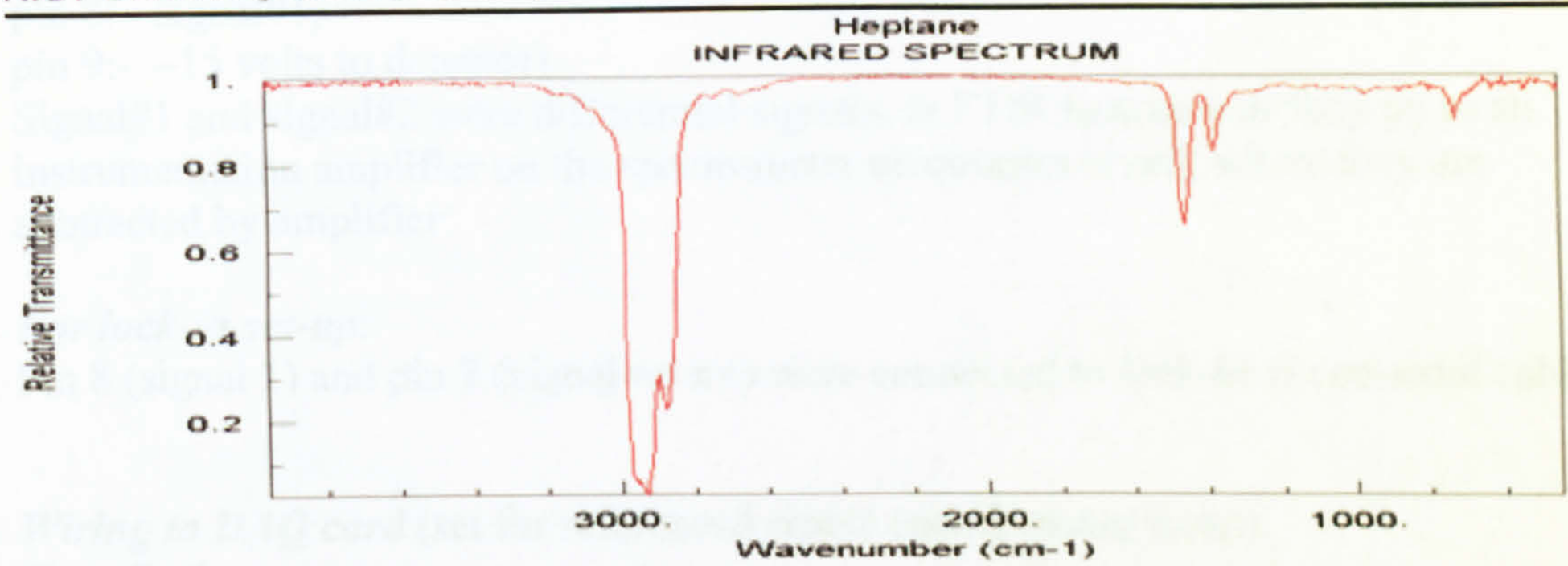
NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)



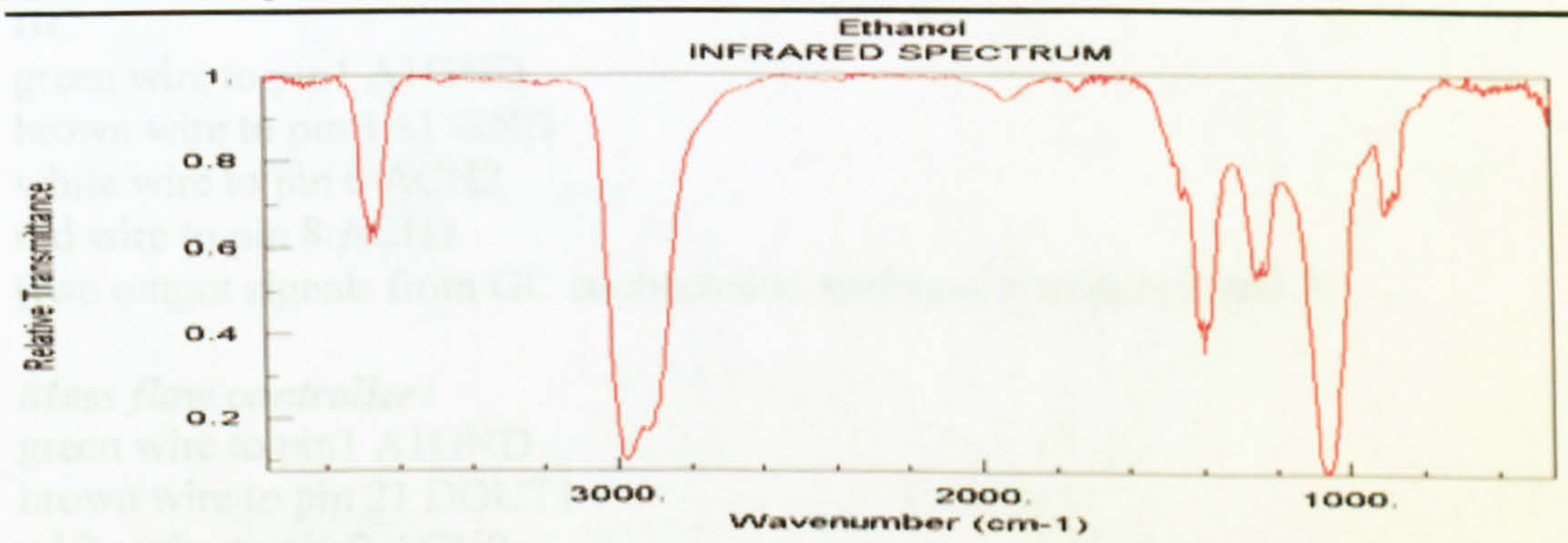
NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)



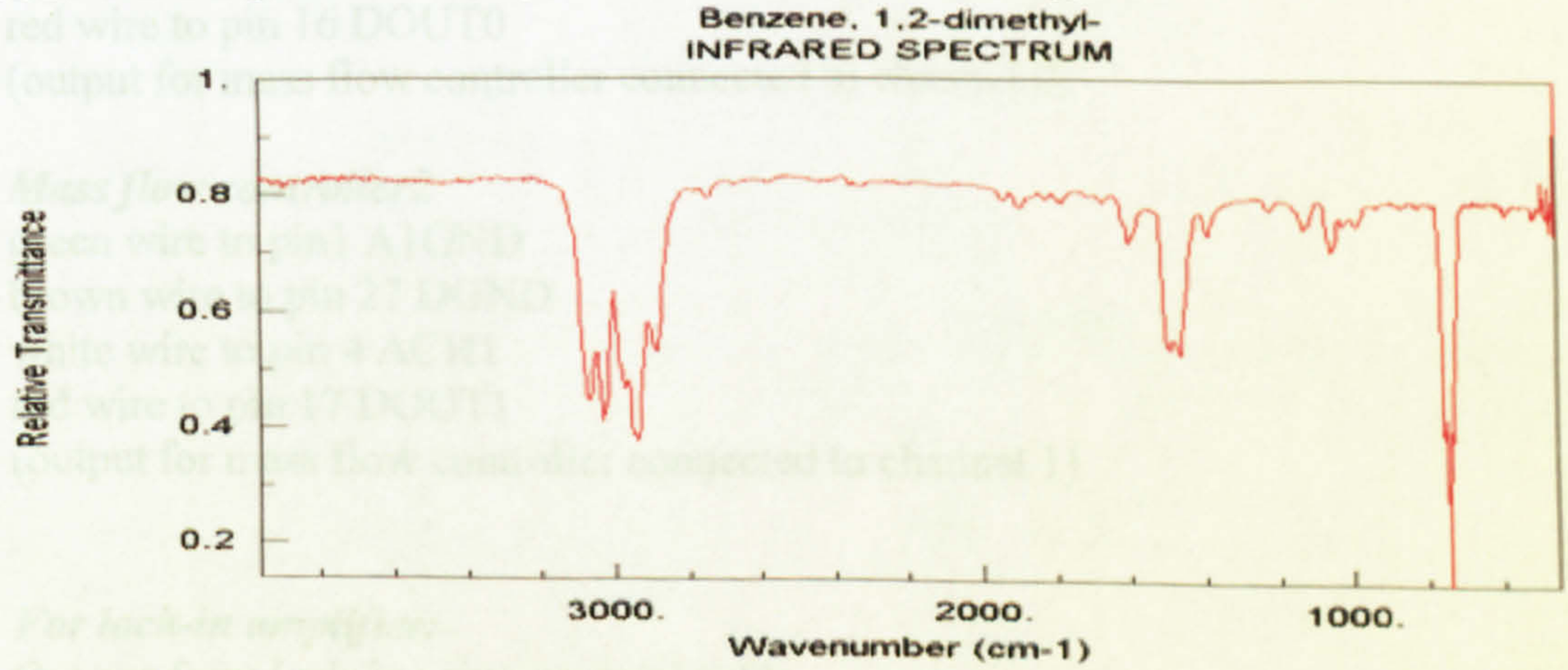
NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)



NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)



NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)



NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)

6.3 Wiring for MCT detector

For Bio-Rad Spectrometer:

Detector had 9-pin connector
pin 1:- 1-wire identification device,
pin2:- signal#2
pin 3:- no connection ,
pin 4:- 15volt return,
pin 5:- +15 volts to detector,
pin 6:- chassis ground at detector,
pin 7:- signal return,
pin 8:- signal#1,
pin 9:- –15 volts to detector)
Signal#1 and signal#2 were differential signals, in FTIR instrument- they go to an instrumentation amplifier on the spectrometer electronics board where they are subtracted by amplifier

For lock-in set-up:
Pin 8 (signal 1) and pin 7 (signal return) were connected to lock-in via co-axial cable

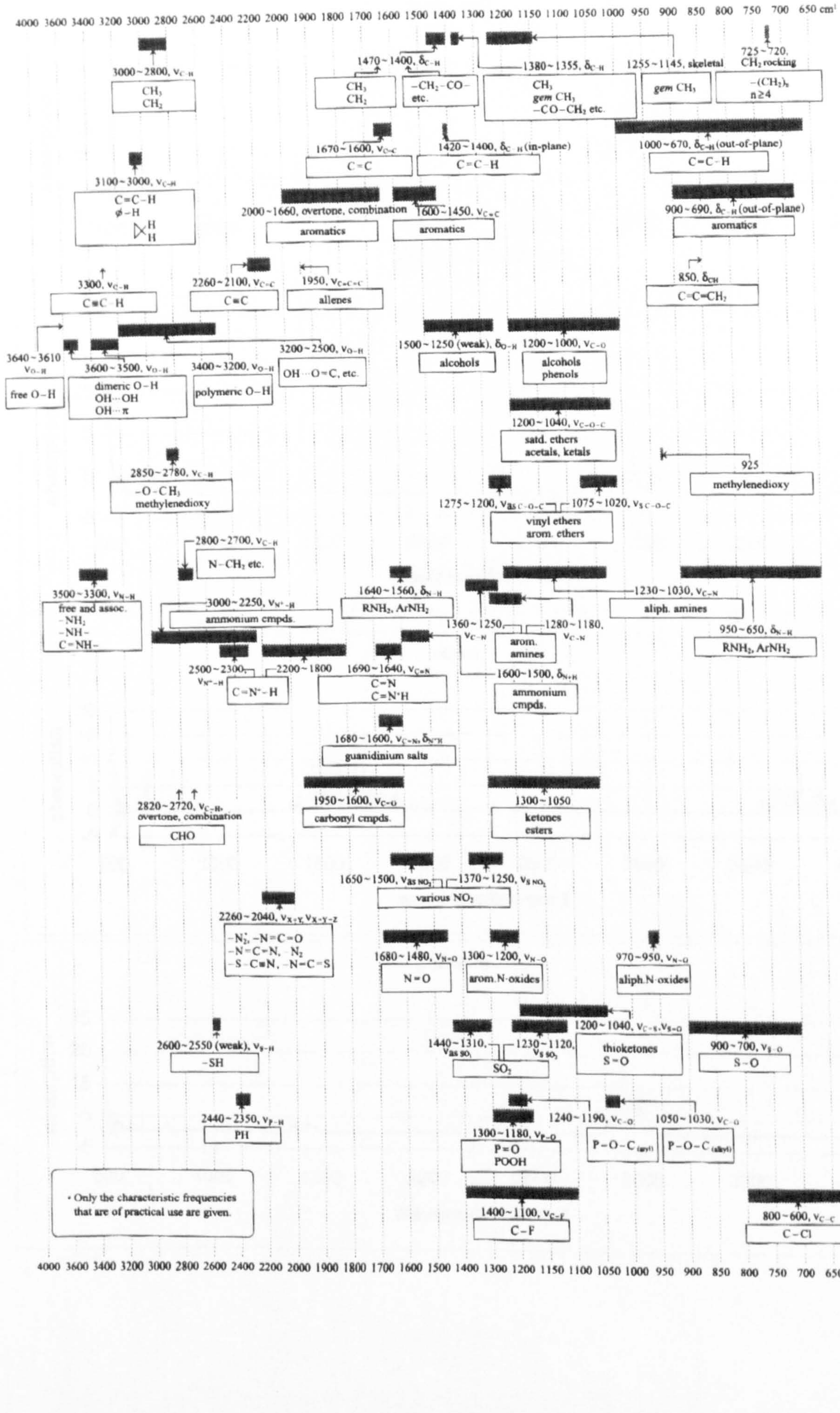
Wiring to DAQ card (set for referenced single ended analog input)
For dilution system:
GC
green wire to pin1 A1GND
brown wire to pin 1A1 GND
white wire to pin 6 ACH2
red wire to pin 8 ACH3
(two output signals from GC connected to analogue channels 2 and 3)

Mass flow controller1
green wire to pin1 A1GND
brown wire to pin 21 DOUT1
white wire to pin 2 ACH0
red wire to pin 16 DOUT0
(output for mass flow controller connected to channel 0)

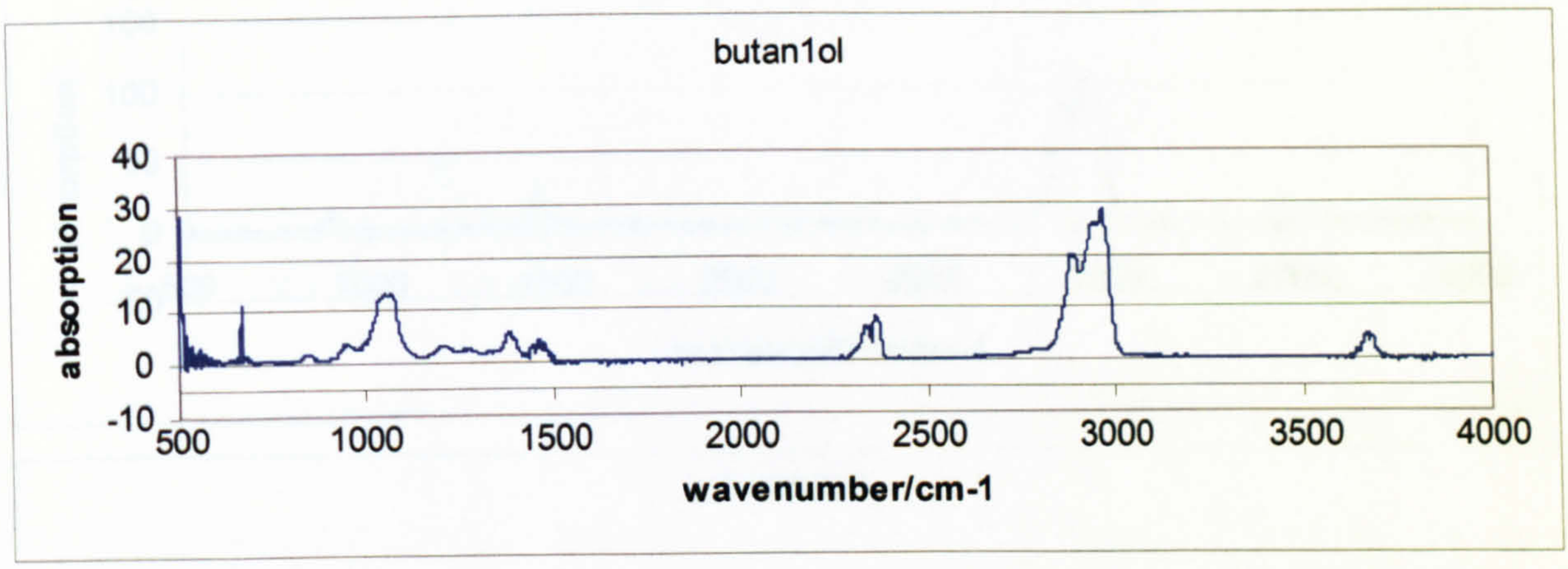
Mass flow controller2
green wire to pin1 A1GND
brown wire to pin 27 DGND
white wire to pin 4 ACH1
red wire to pin 17 DOUT1
(output for mass flow controller connected to channel 1)

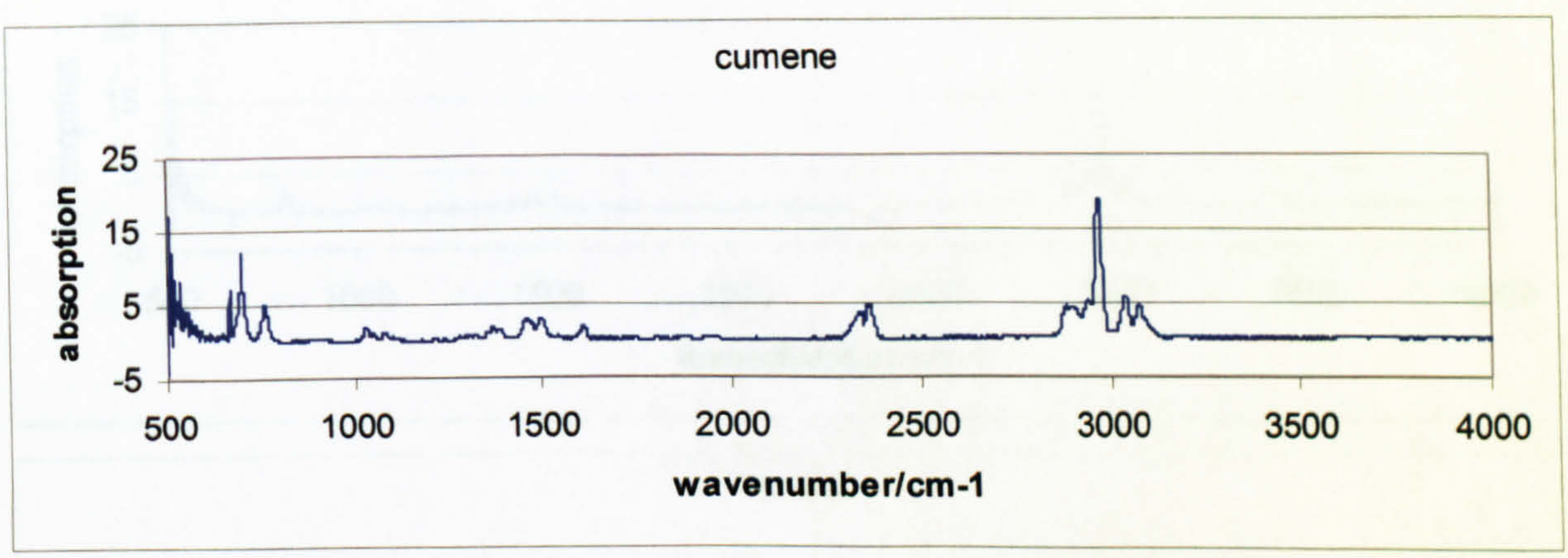
For lock-in amplifier:-
Out put from lock-in using co-axial cable,
shielding wired to ground (pin1 A1 GND)
core wired to pin 2 ACH0
(output connected to analogue channel 0)

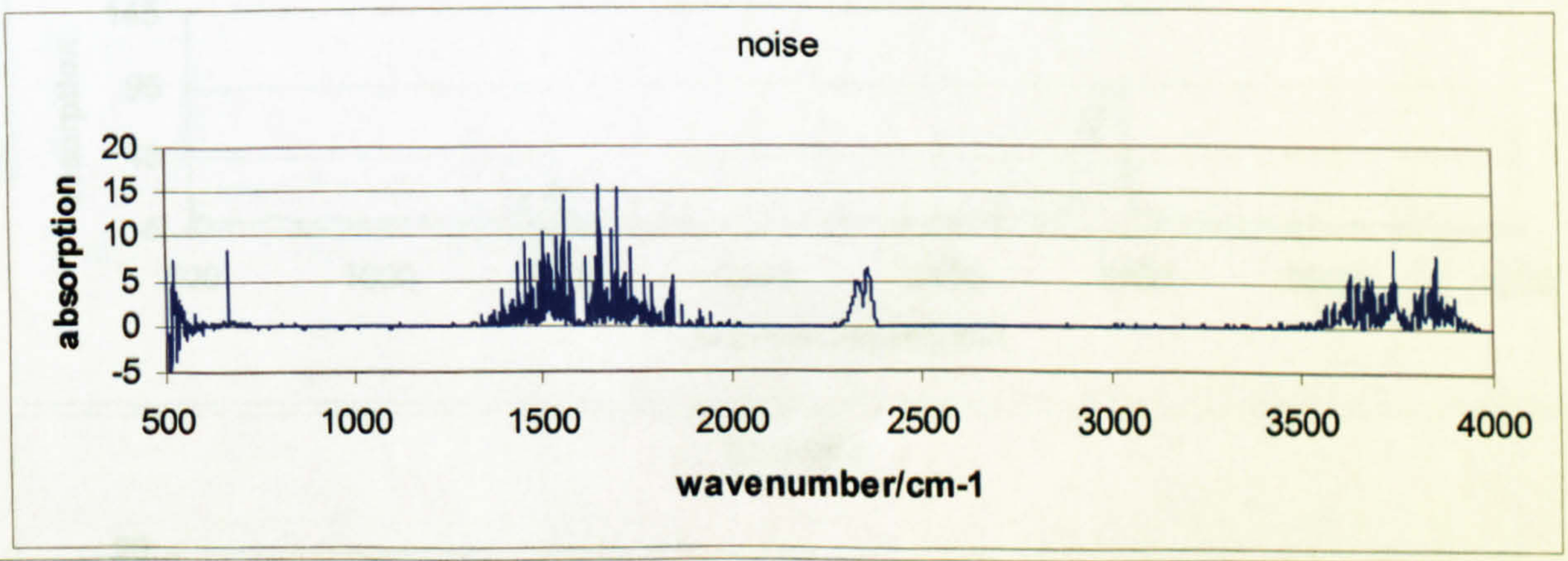
6.4 Characteristic IR absorptions

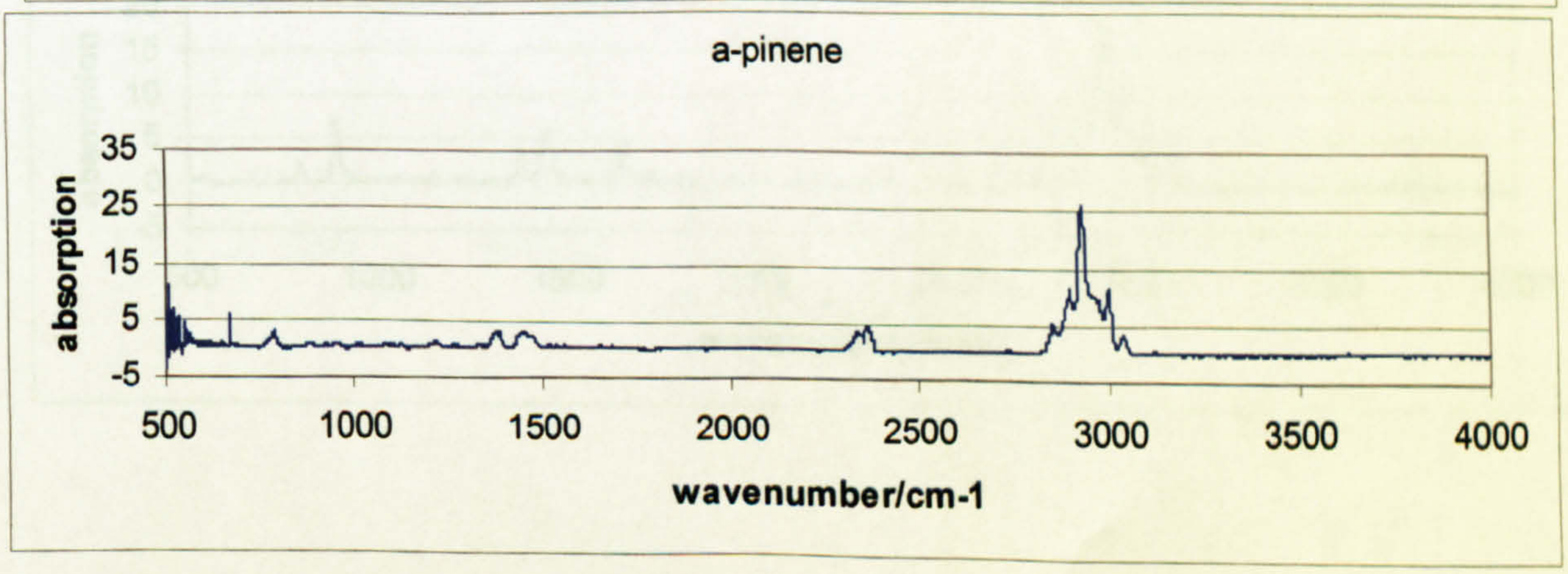


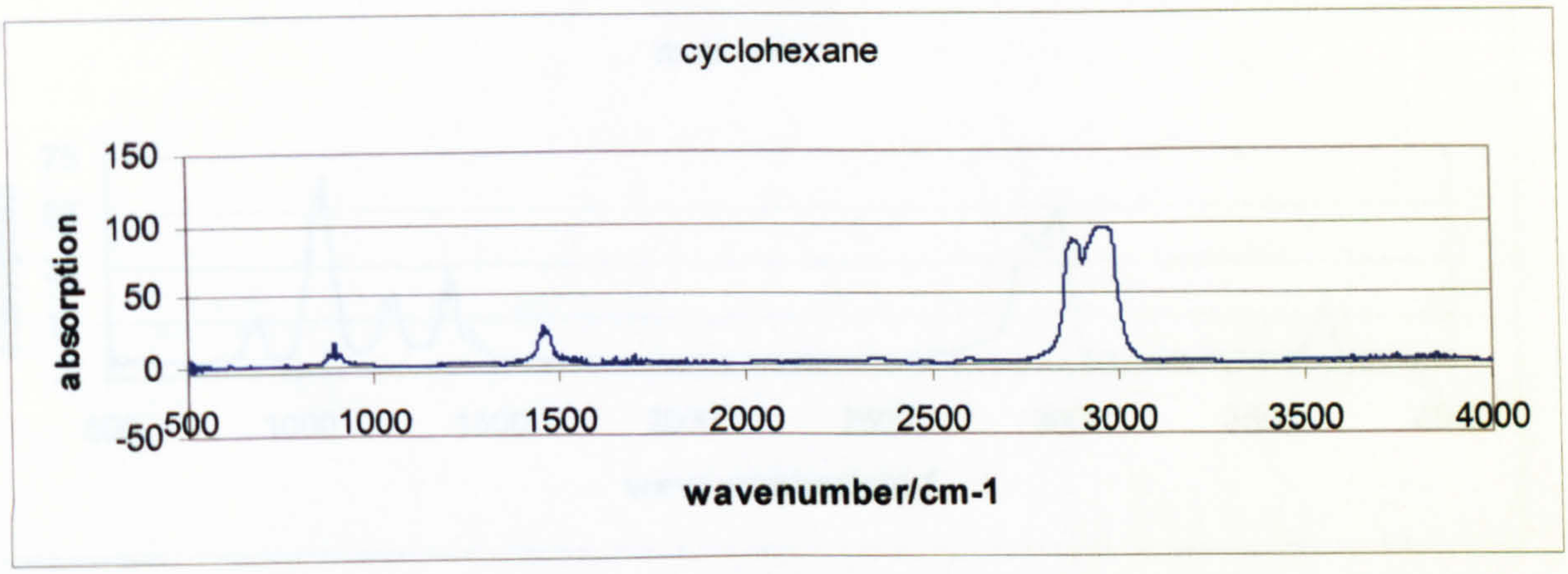
6.5 Vapour Spectra

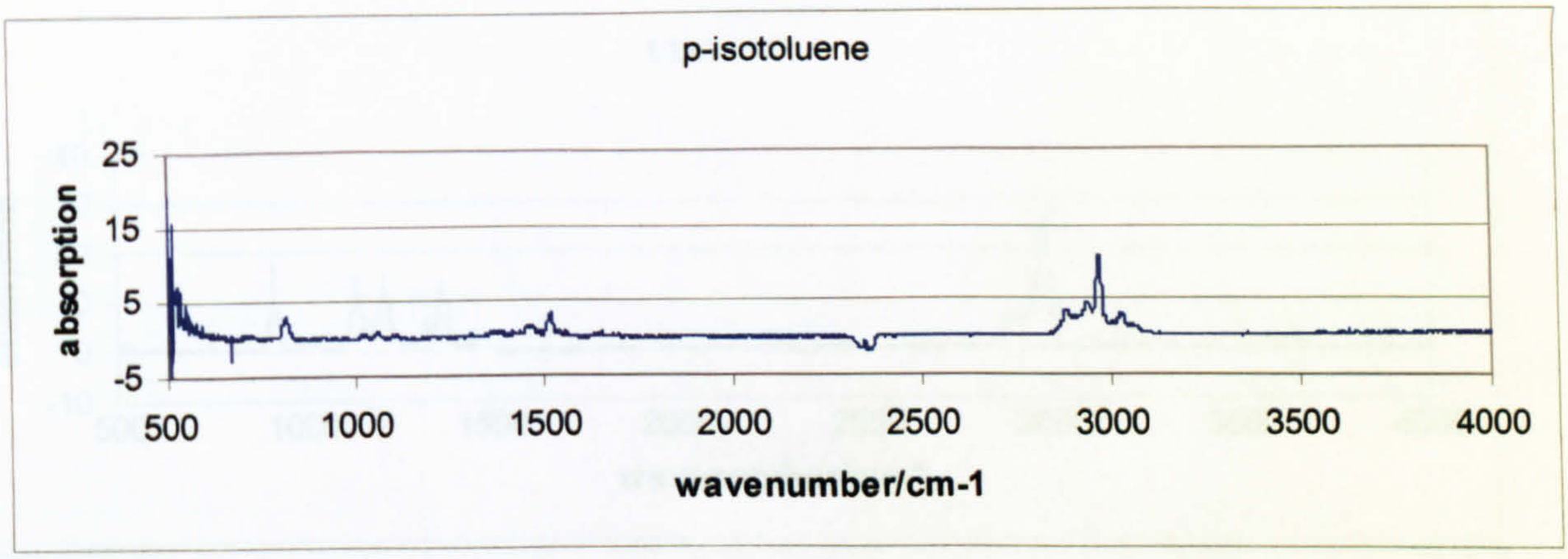


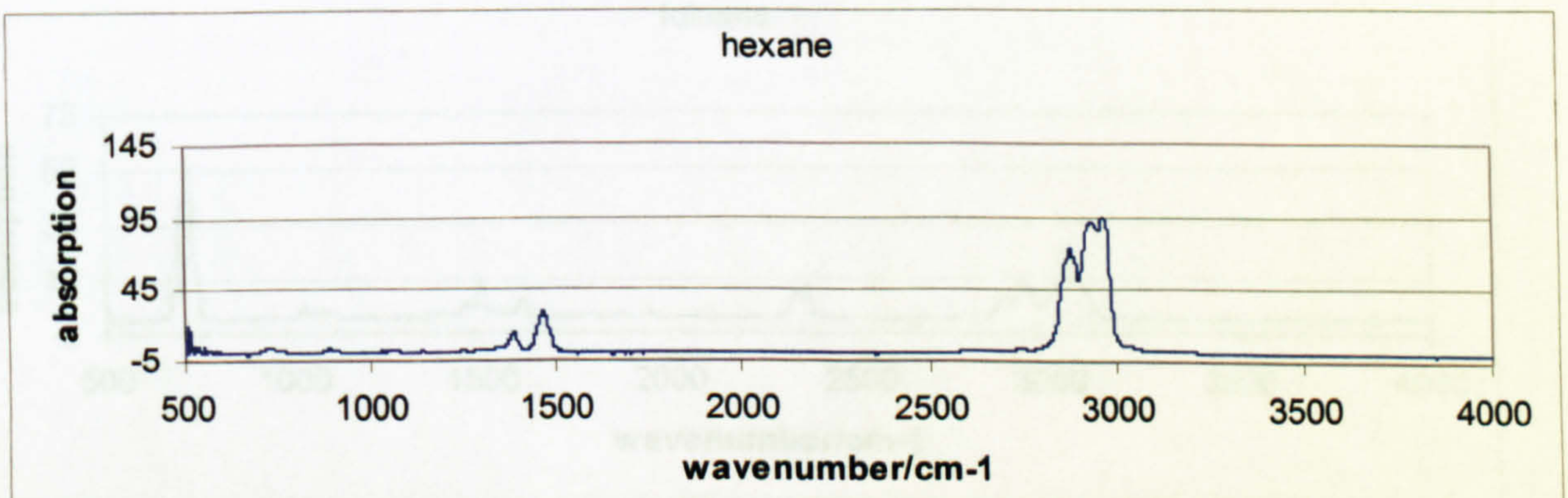


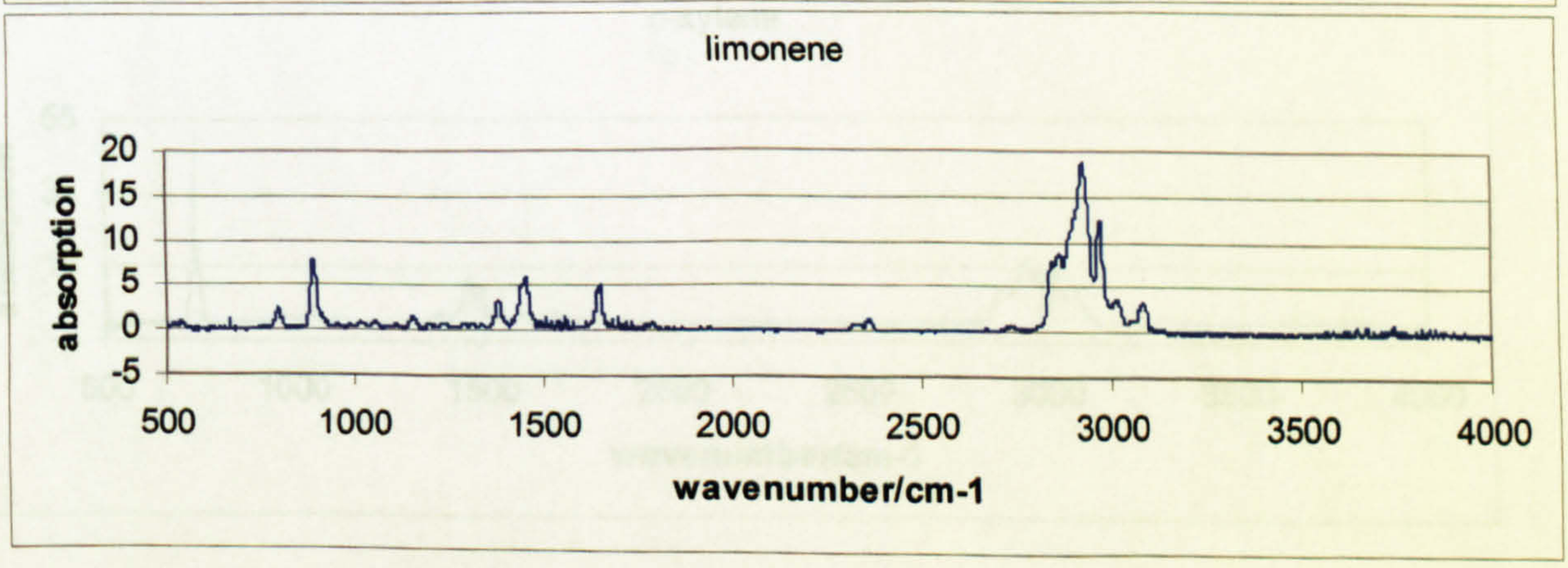


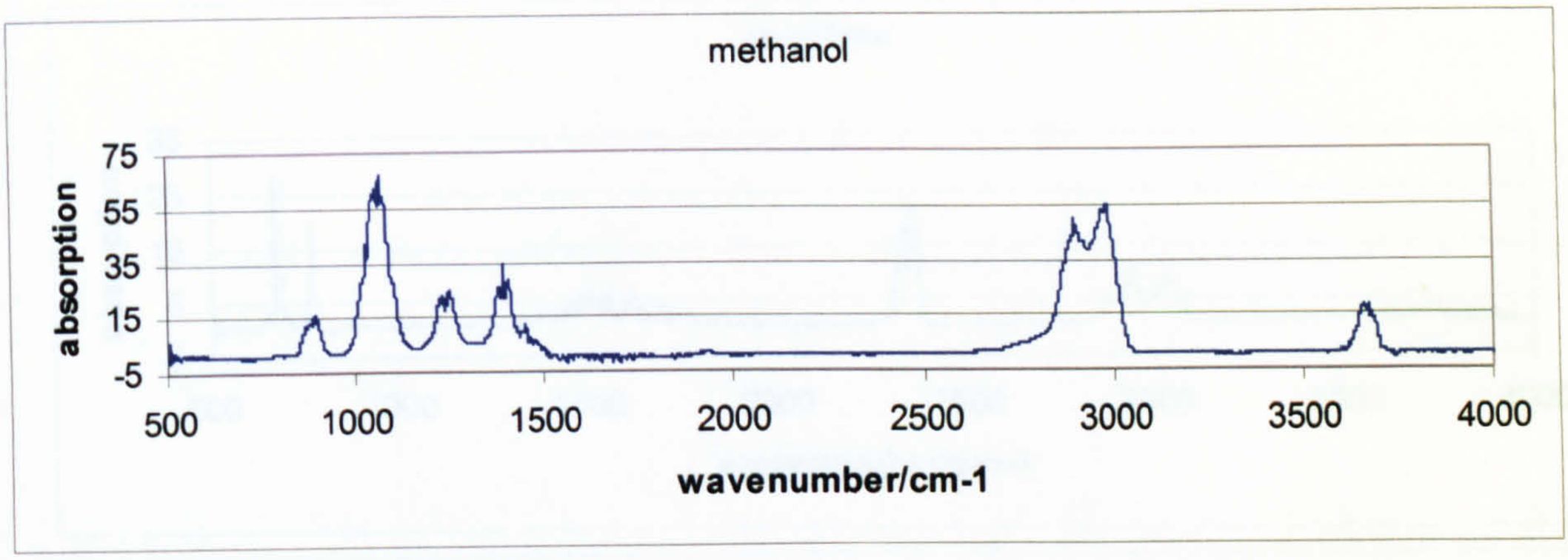


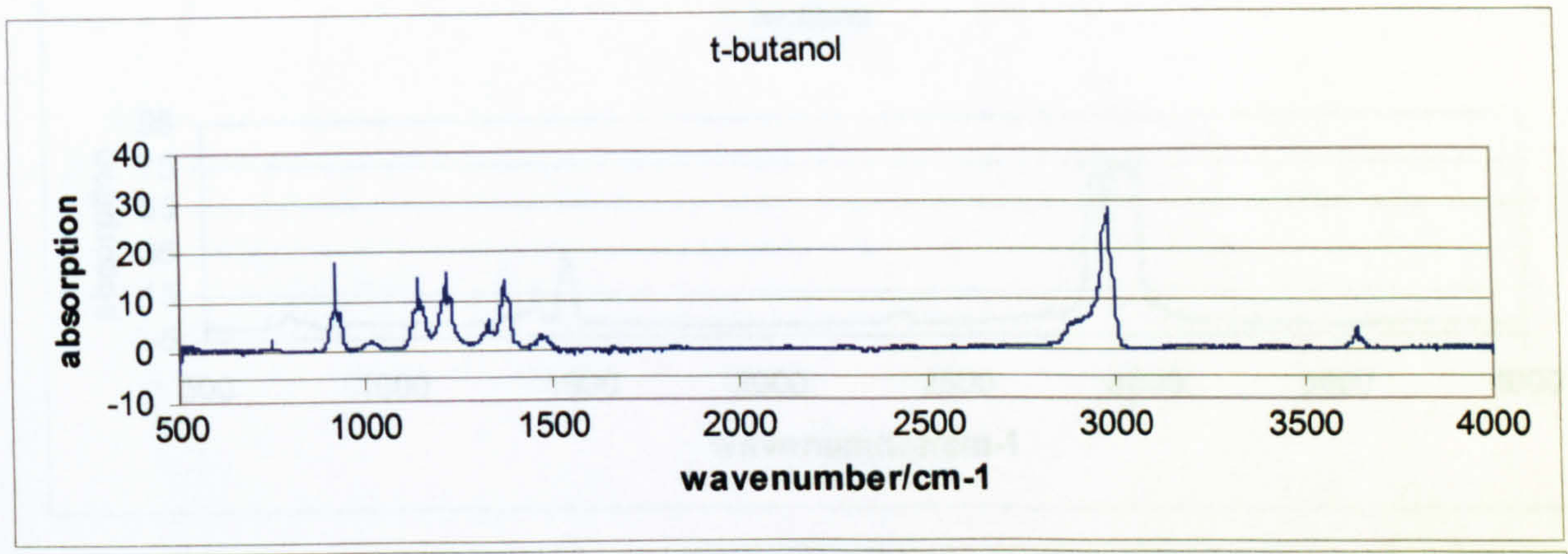

cyclohexane
absorption
150
100
50
0
-50
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

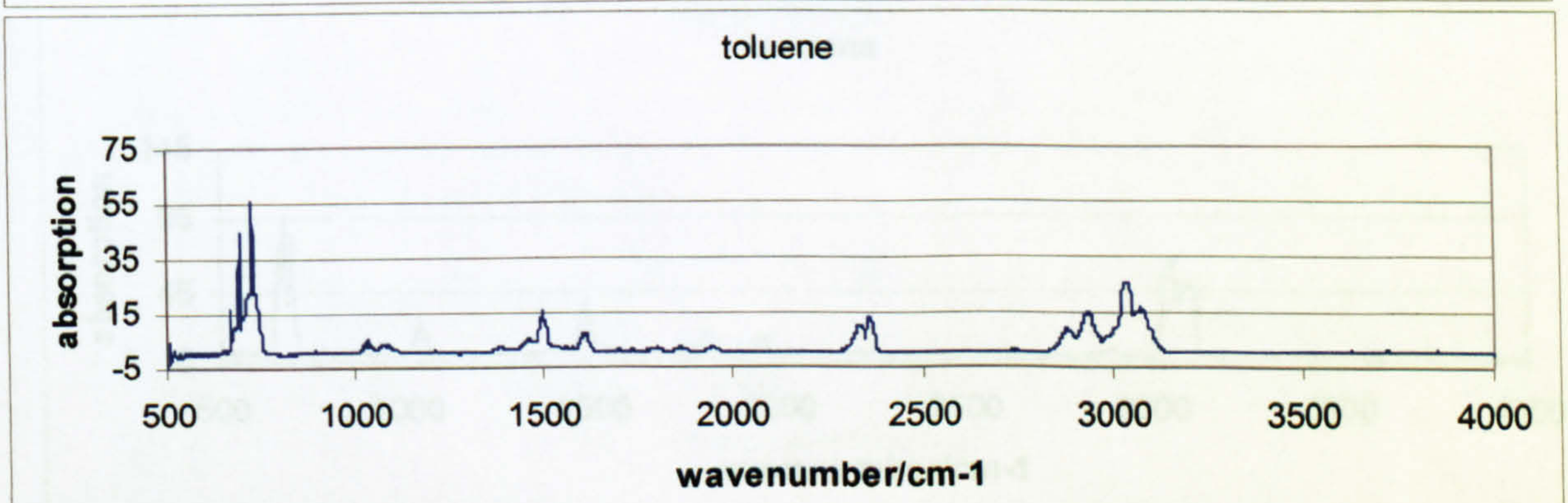

p-isotoluene
absorption
25
15
5
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

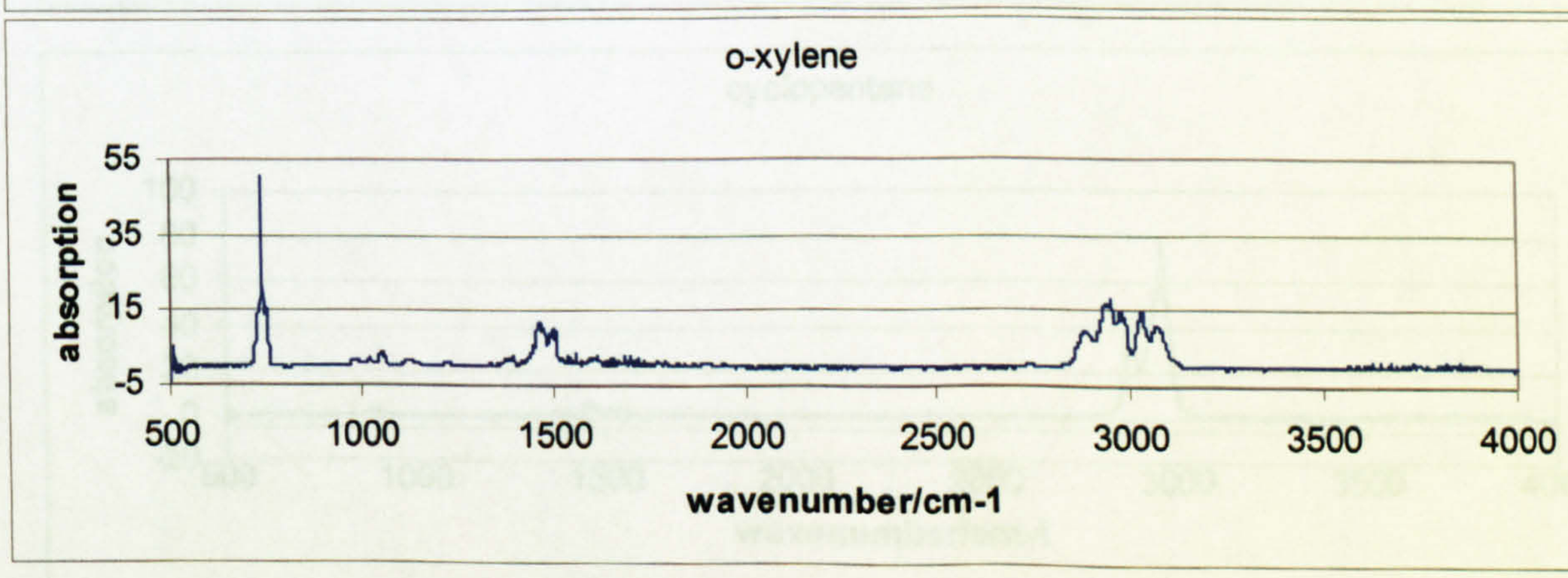

hexane
absorption
145
95
45
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

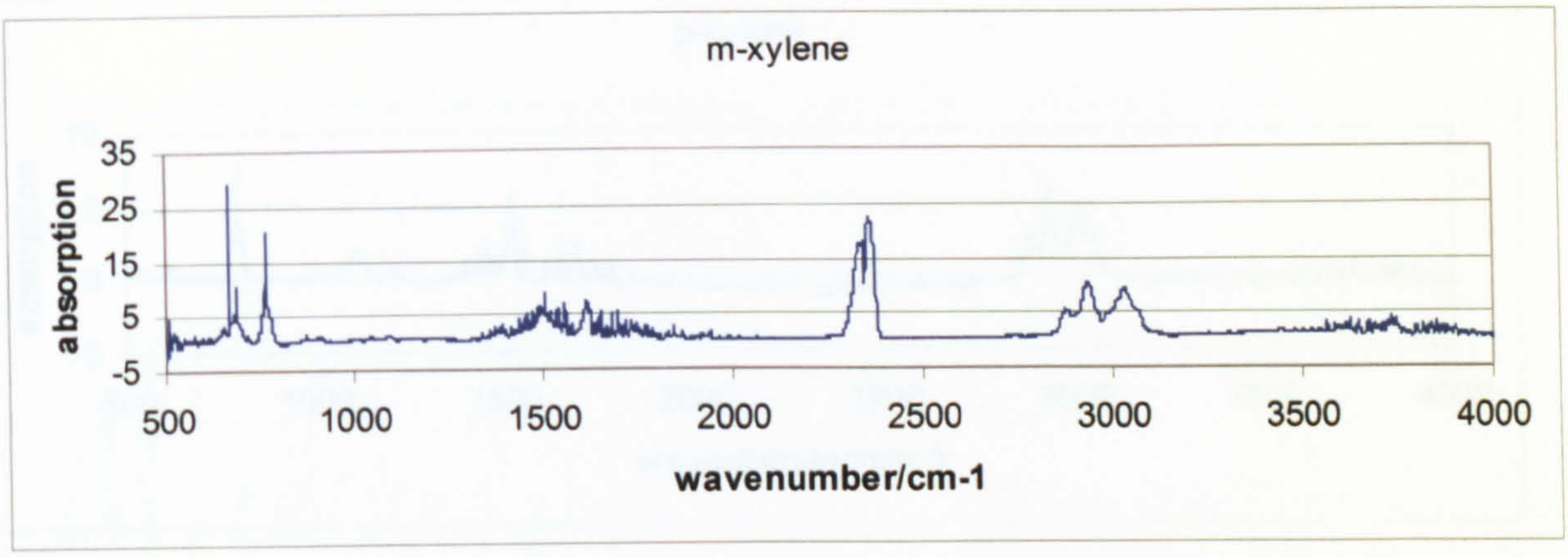

limonene
absorption
20
15
10
5
0
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

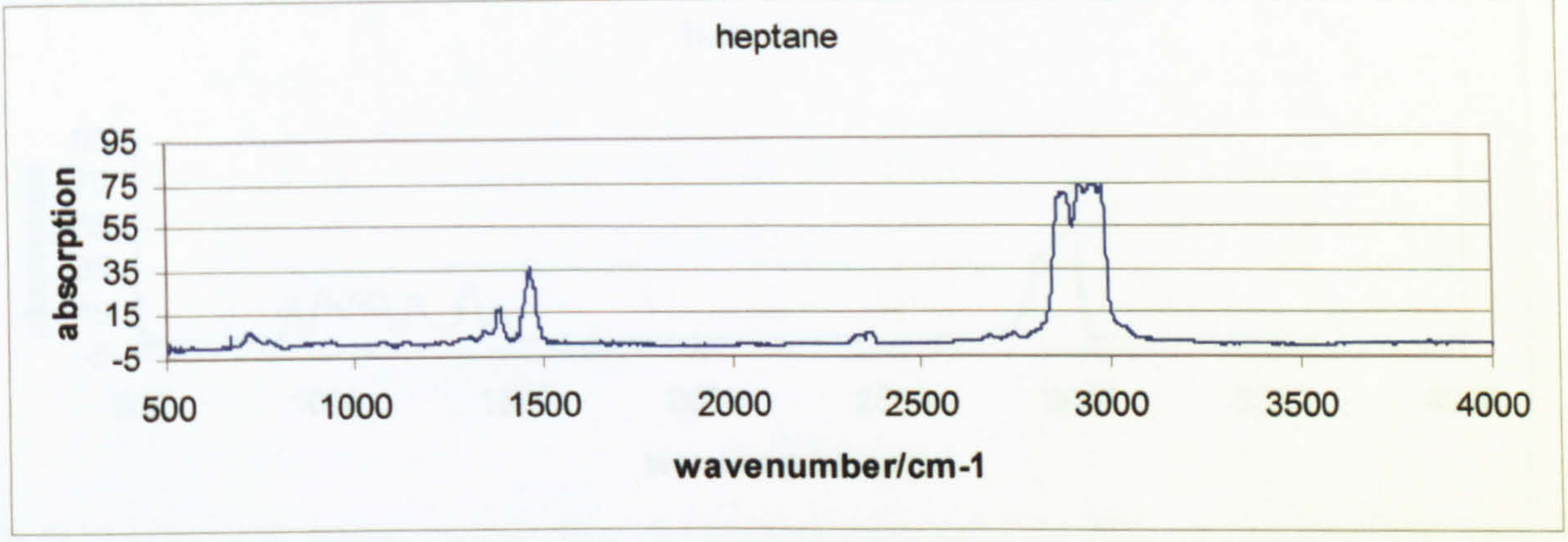

methanol
absorption
75
55
35
15
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

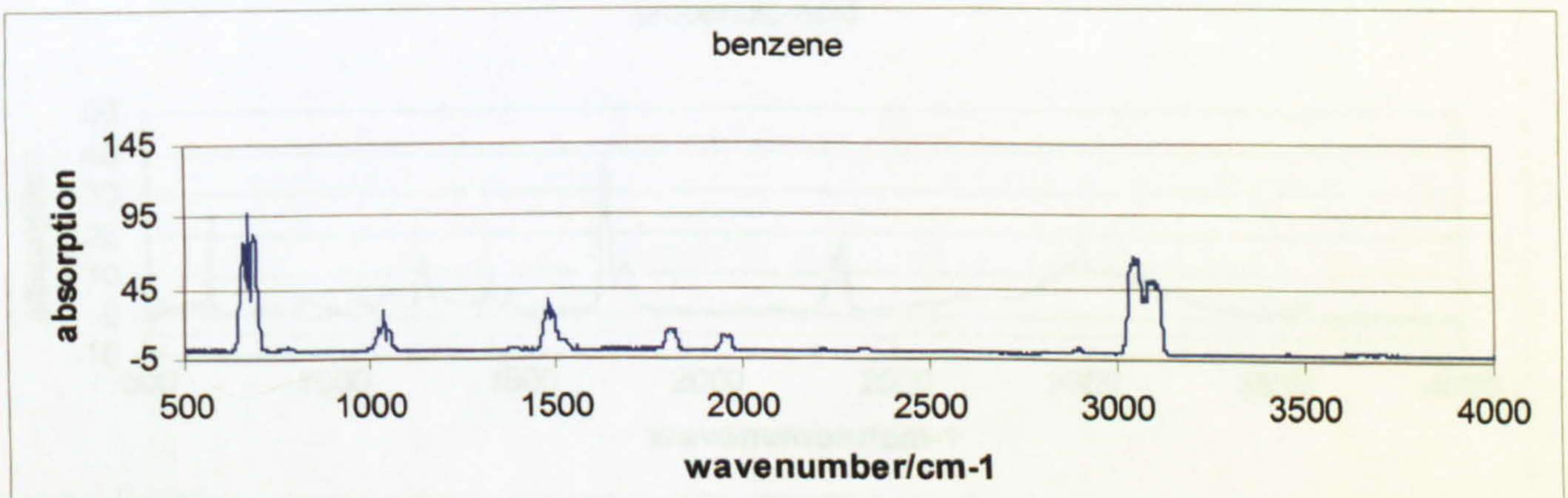

t-butanol
absorption
40
30
20
10
0
-10
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

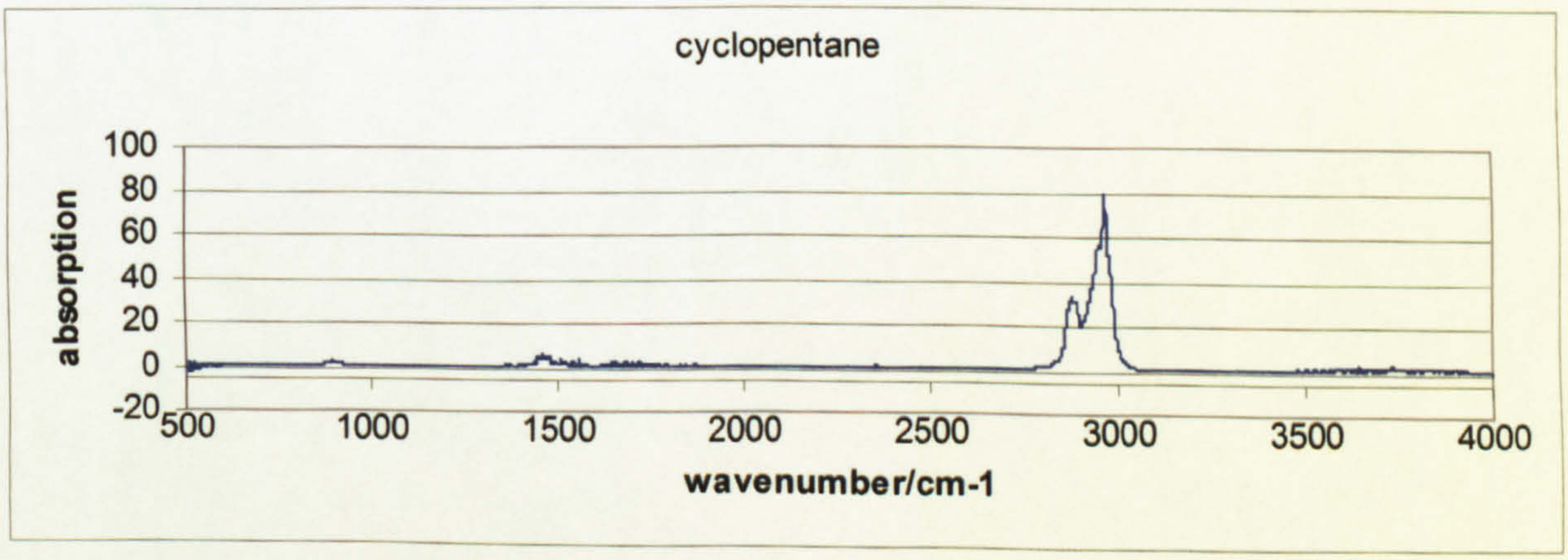

toluene
absorption
75
55
35
15
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

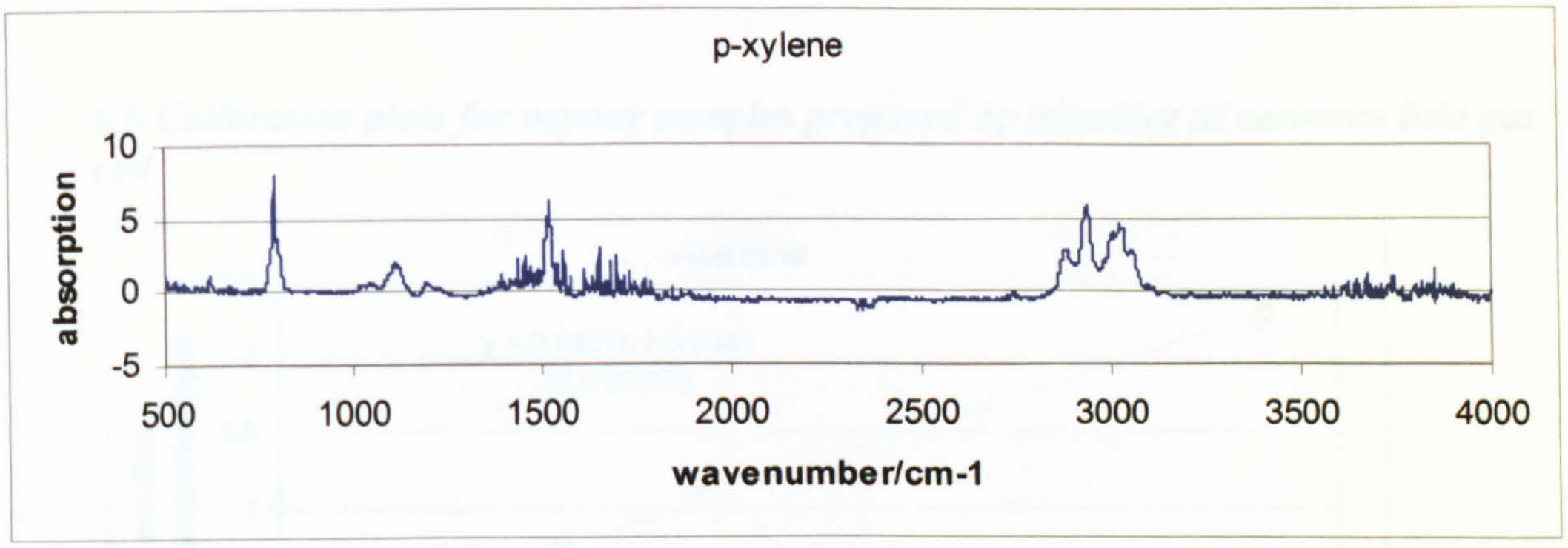

o-xylene
absorption
55
35
15
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

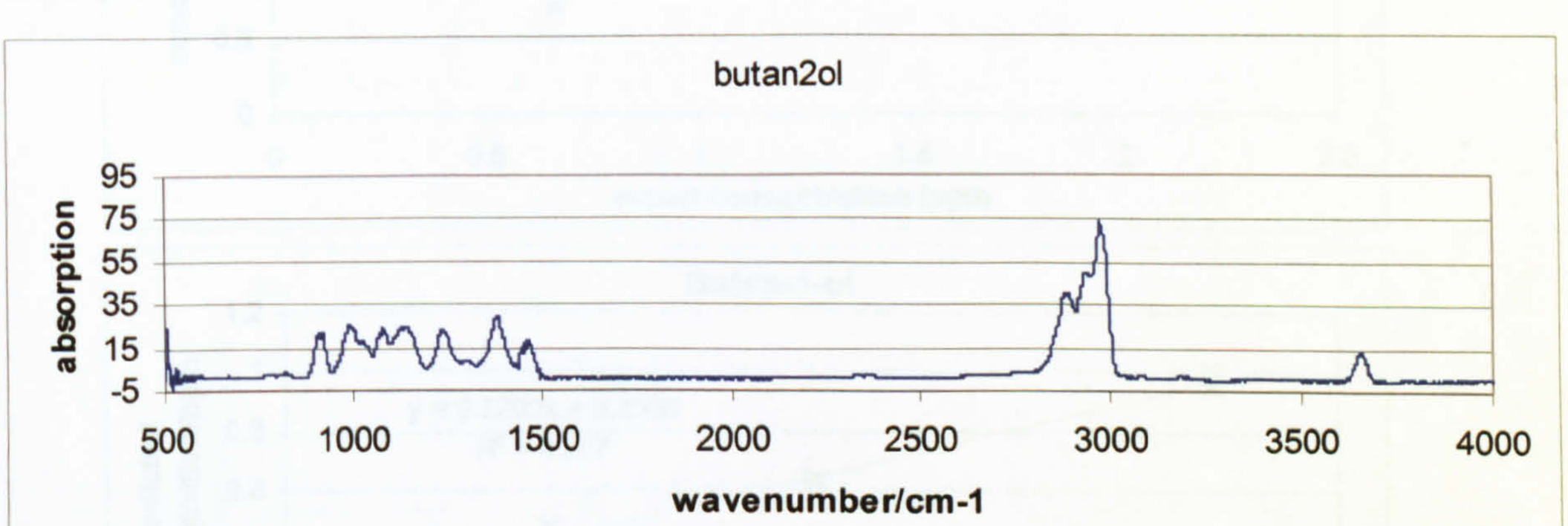

m-xylene
absorption
35
25
15
5
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

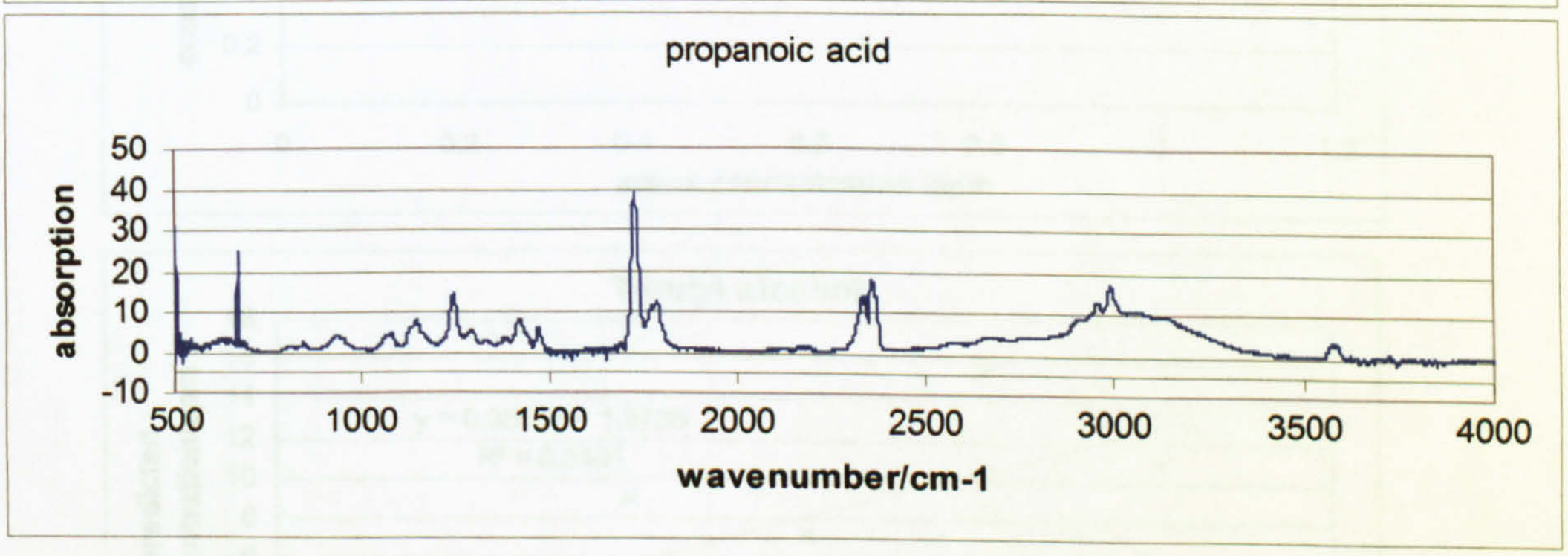

heptane
absorption
95
75
55
35
15
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1


benzene
absorption
145
95
45
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

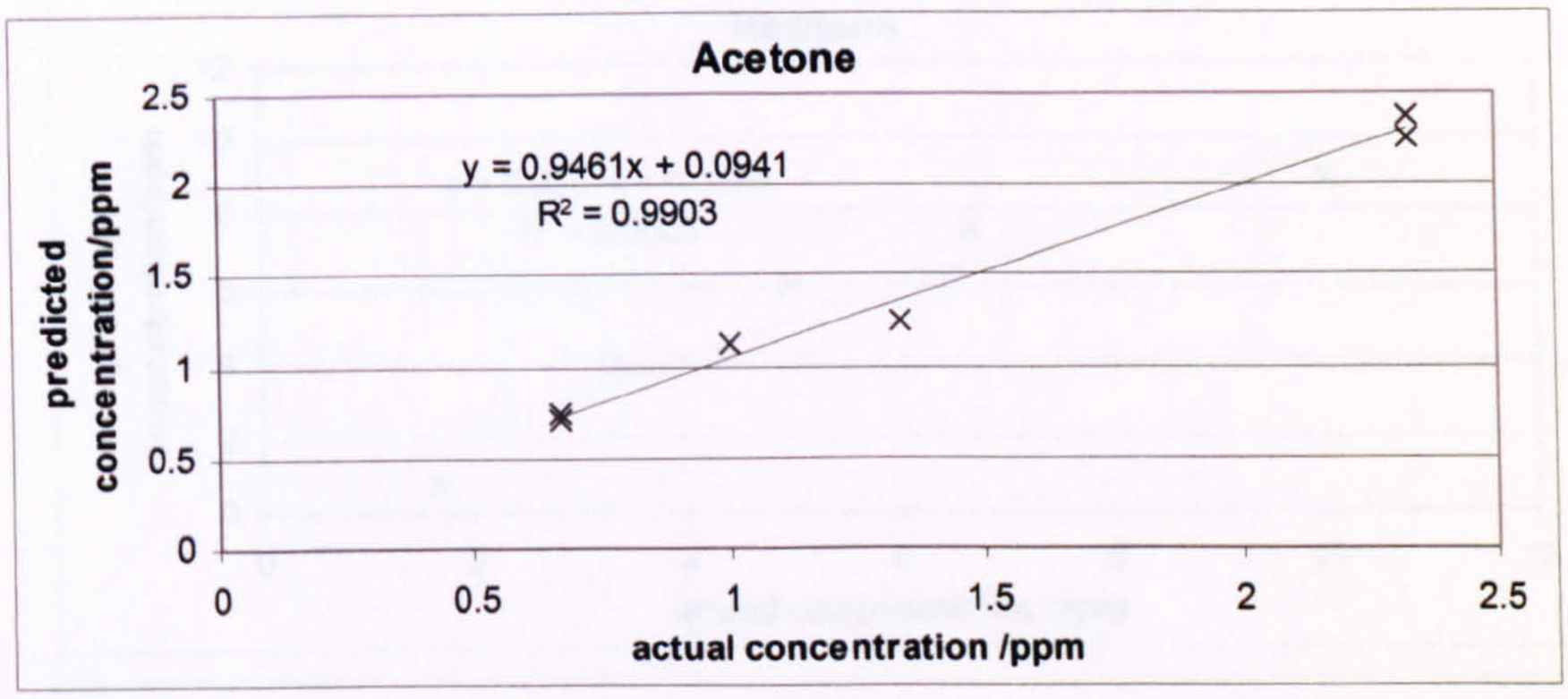

cyclopentane
absorption
100
80
60
40
20
0
-20
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

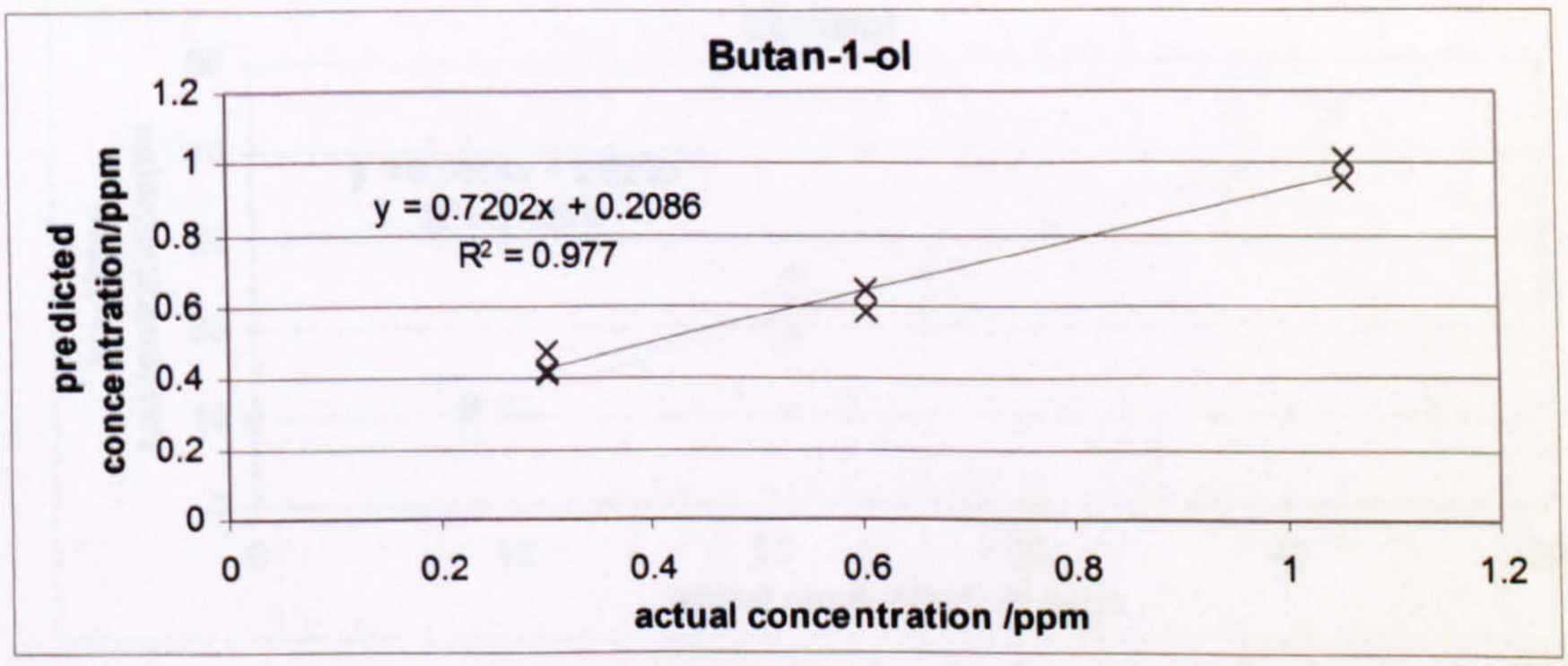

p-xylene
absorption
10
5
0
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

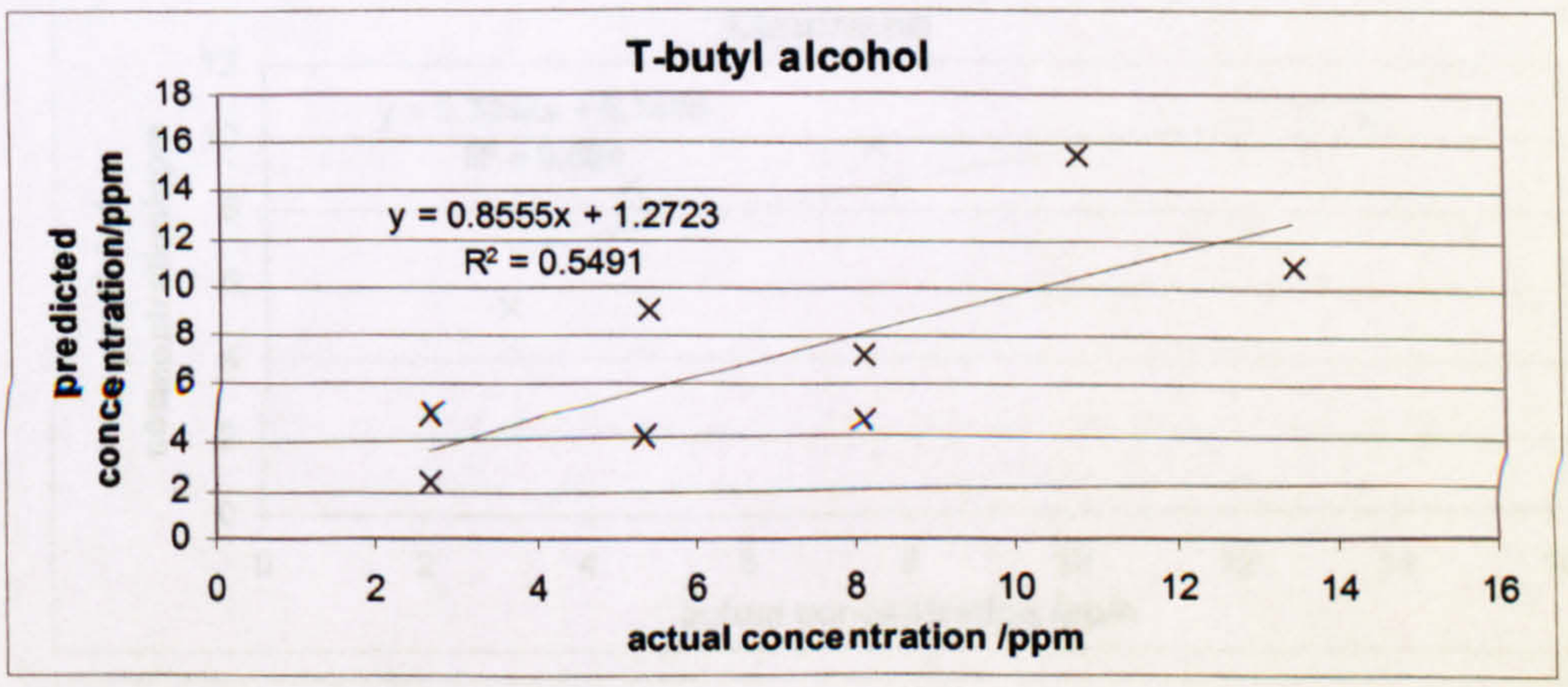

butan2ol
absorption
95
75
55
35
15
-5
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

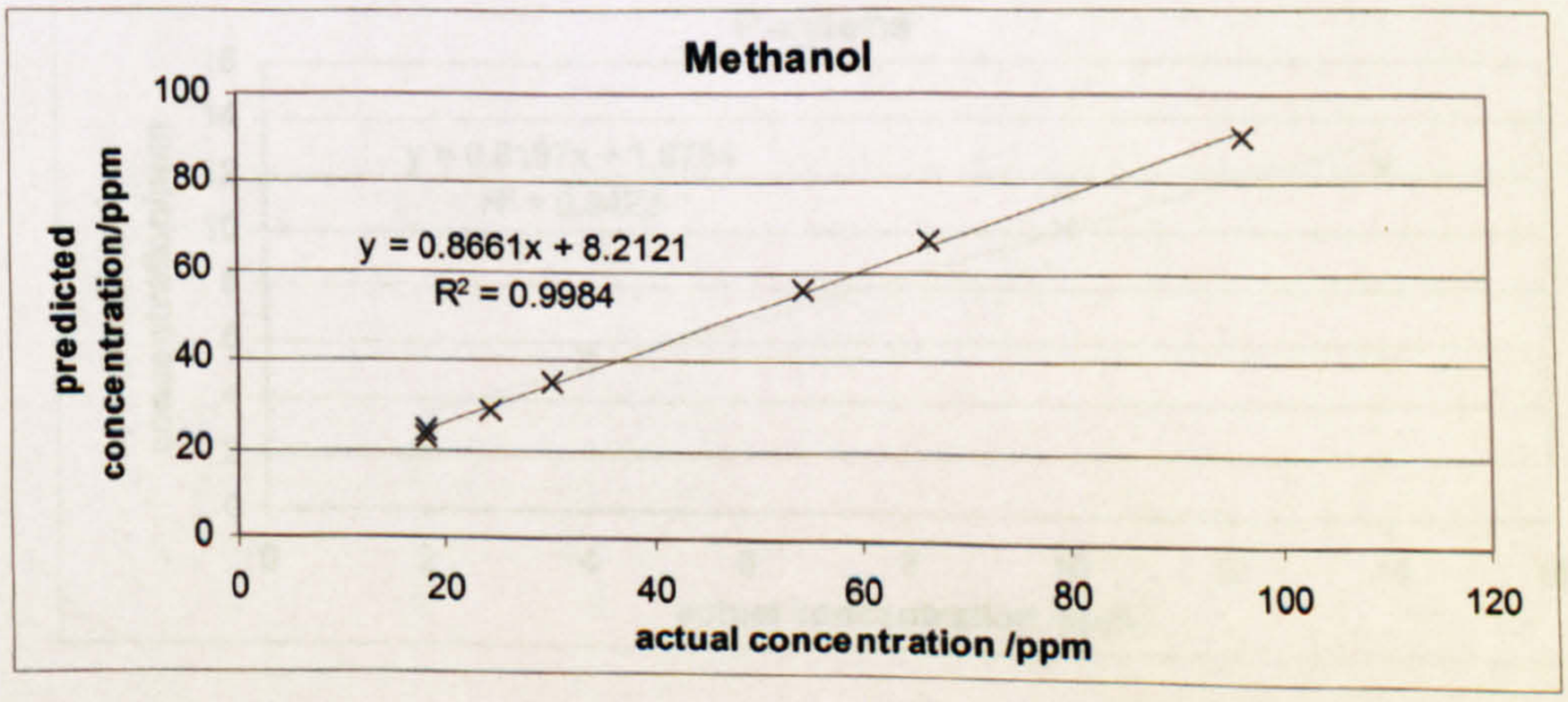

propanoic acid
absorption
50
40
30
20
10
0
-10
500
1000
1500
2000
2500
3000
3500
4000
wavenumber/cm-1

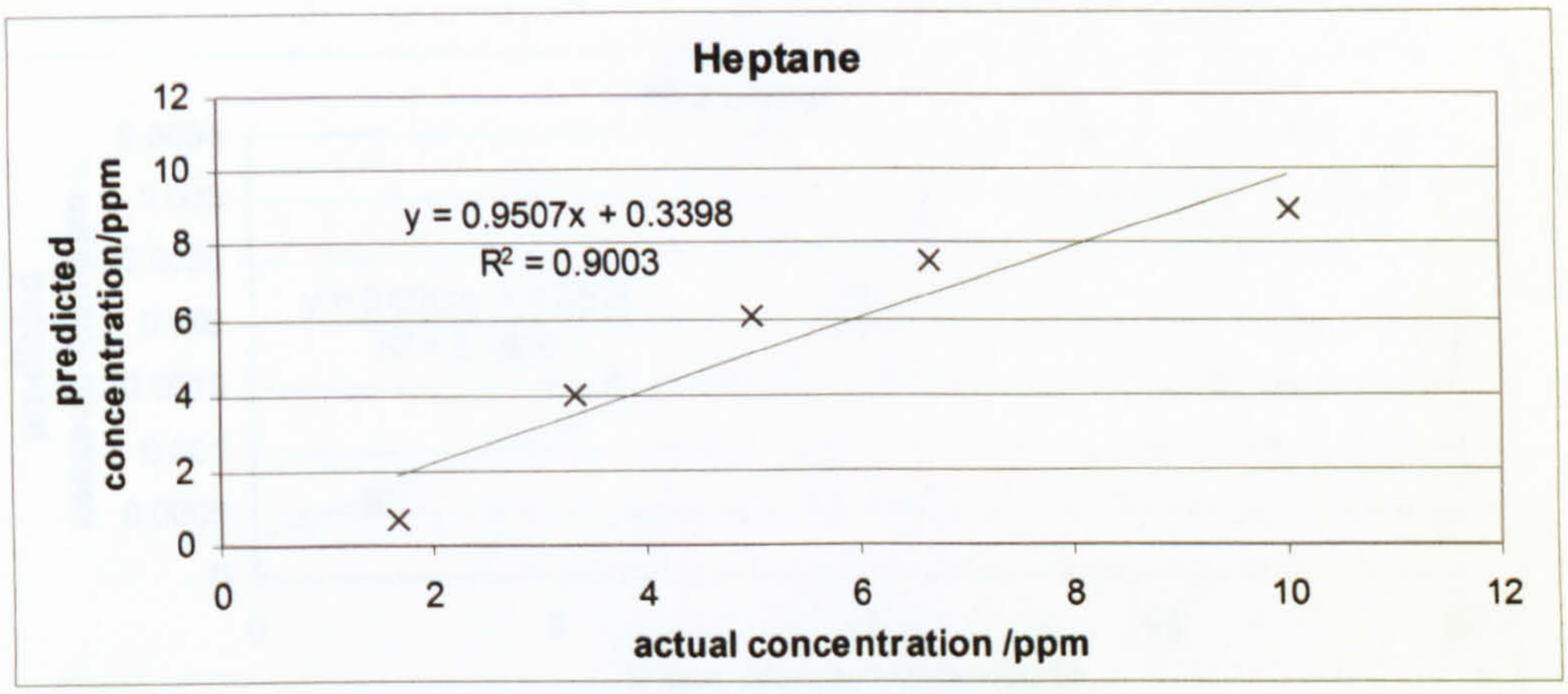
6.6 Calibration plots for vapour samples prepared by injecting µl amounts into gas cell

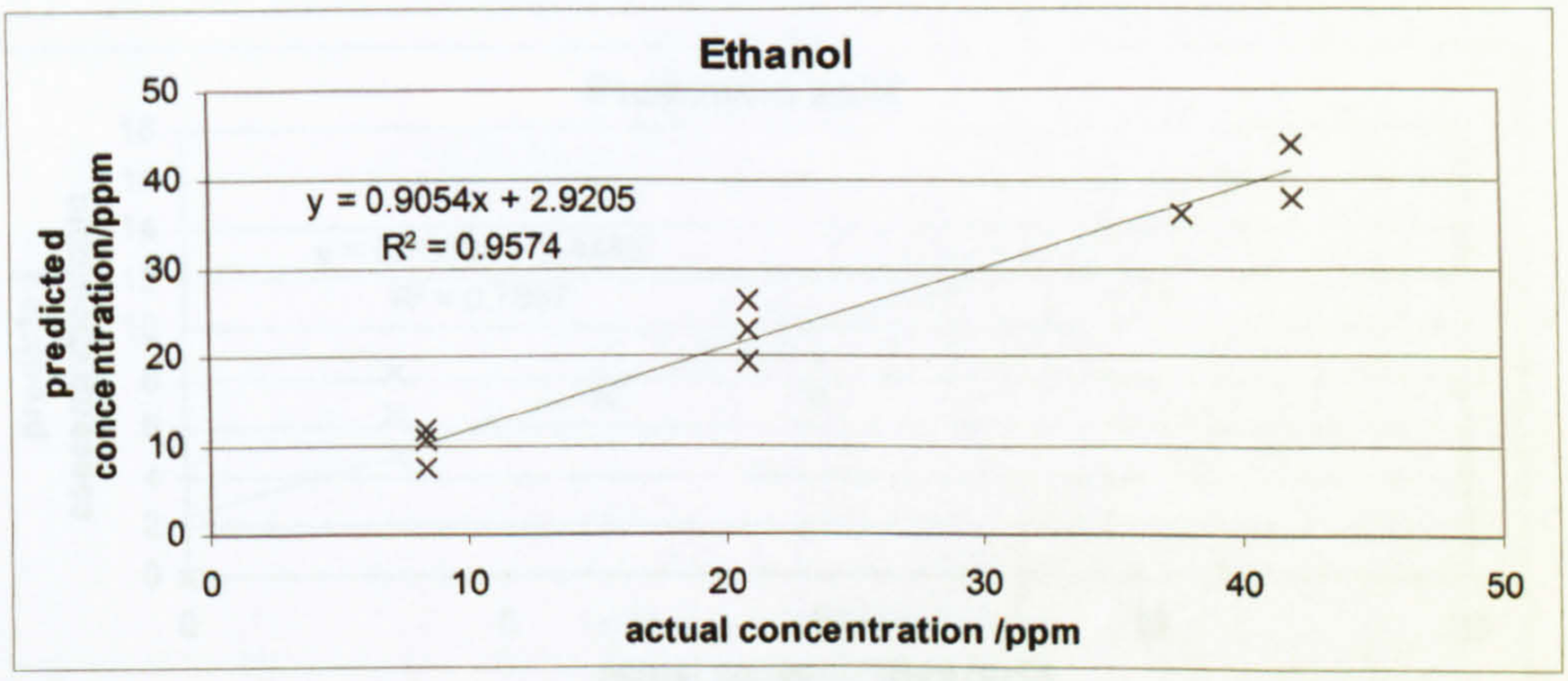


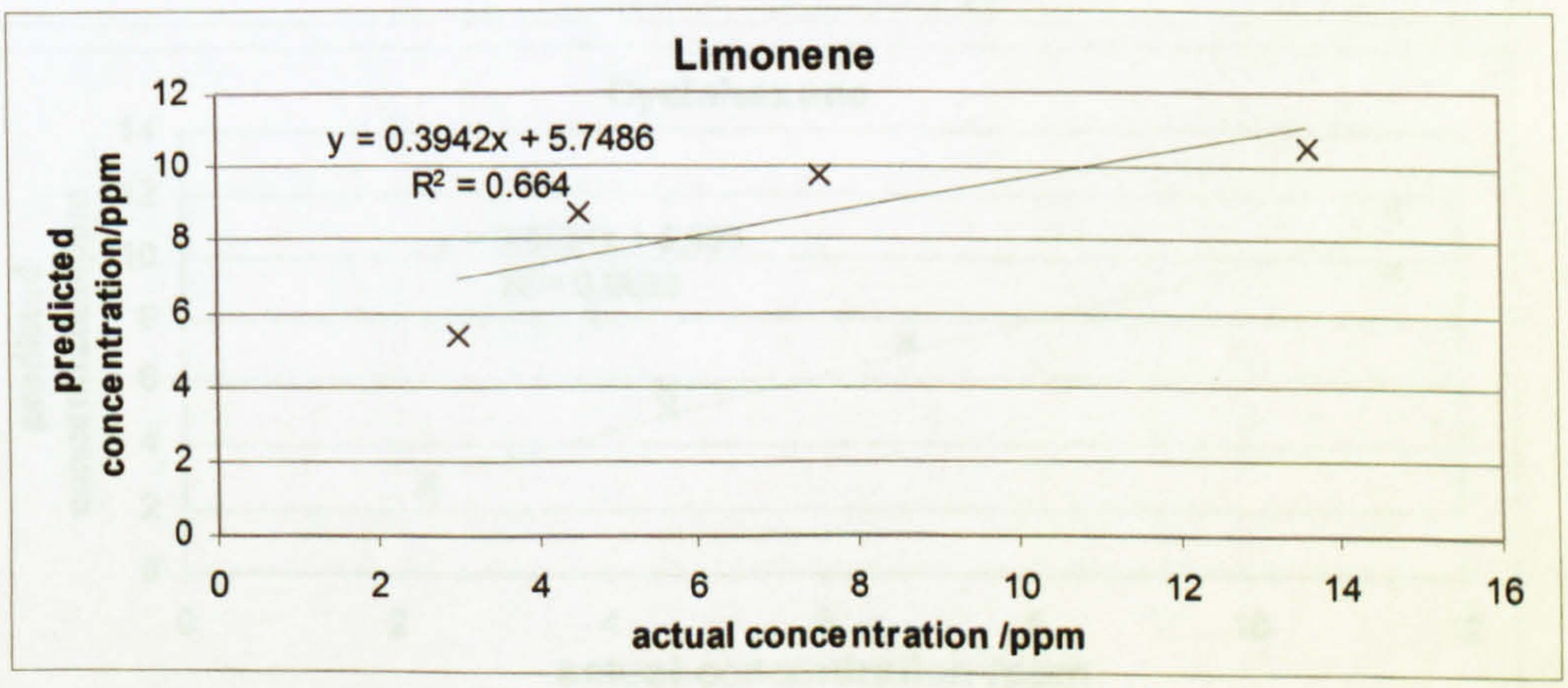


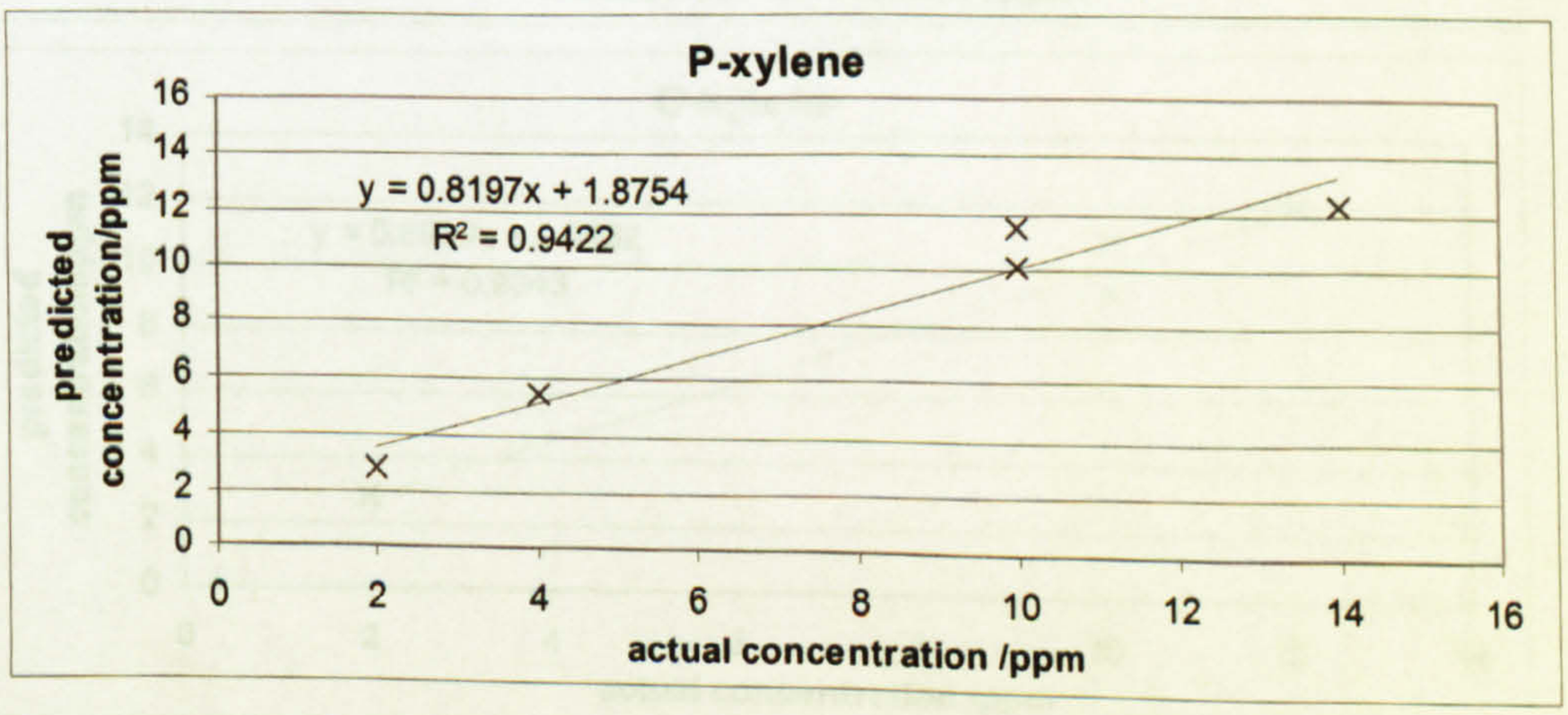


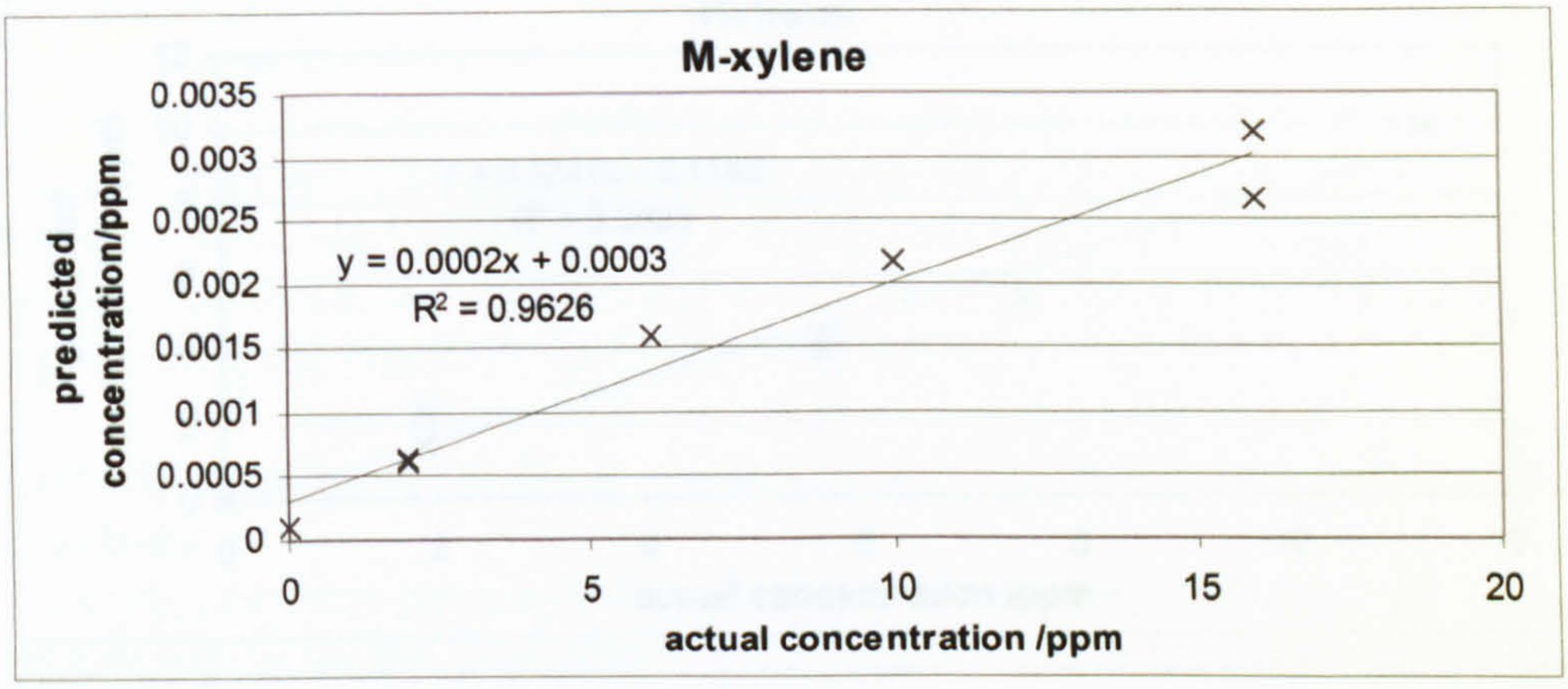


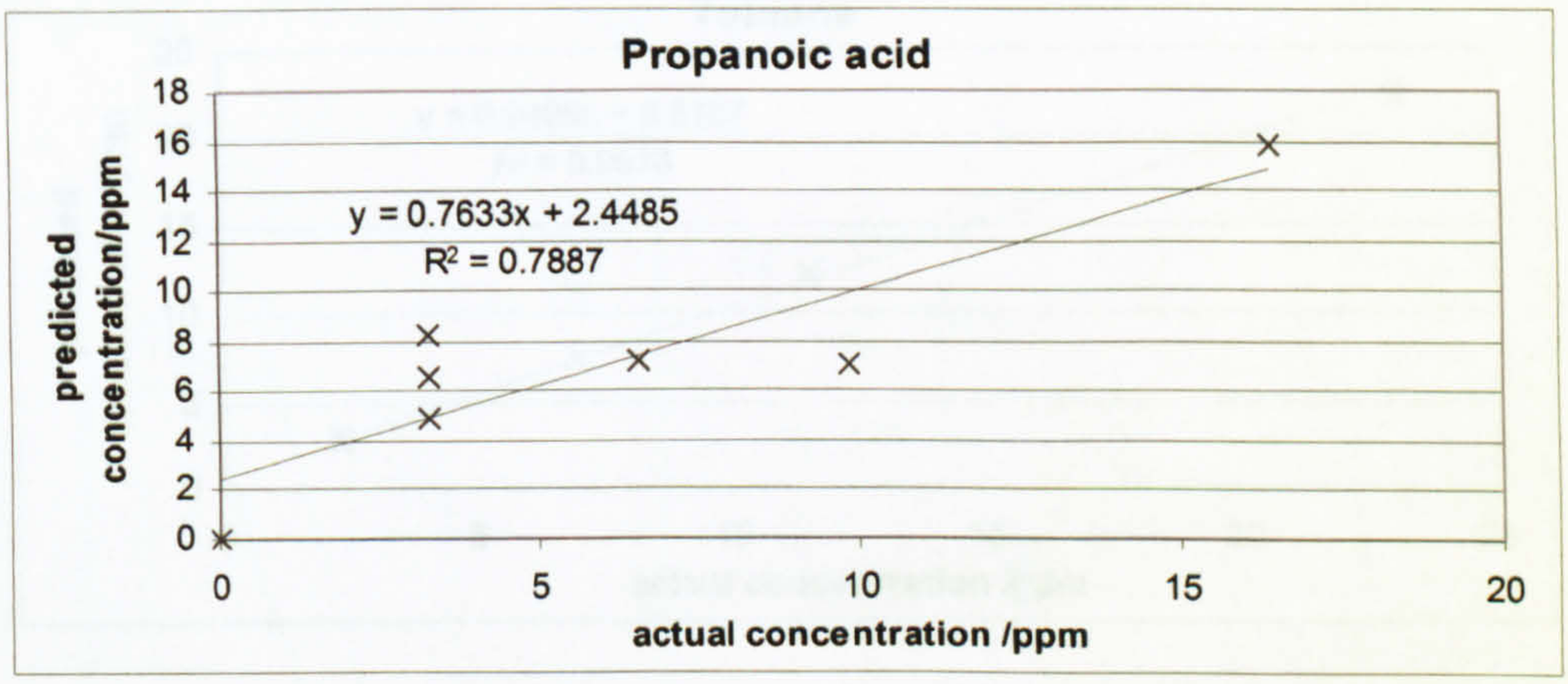

Heptane
predicted concentration/ppm
y = 0.9507x + 0.3398
R² = 0.9003
0
2
4
6
8
10
12
actual concentration /ppm

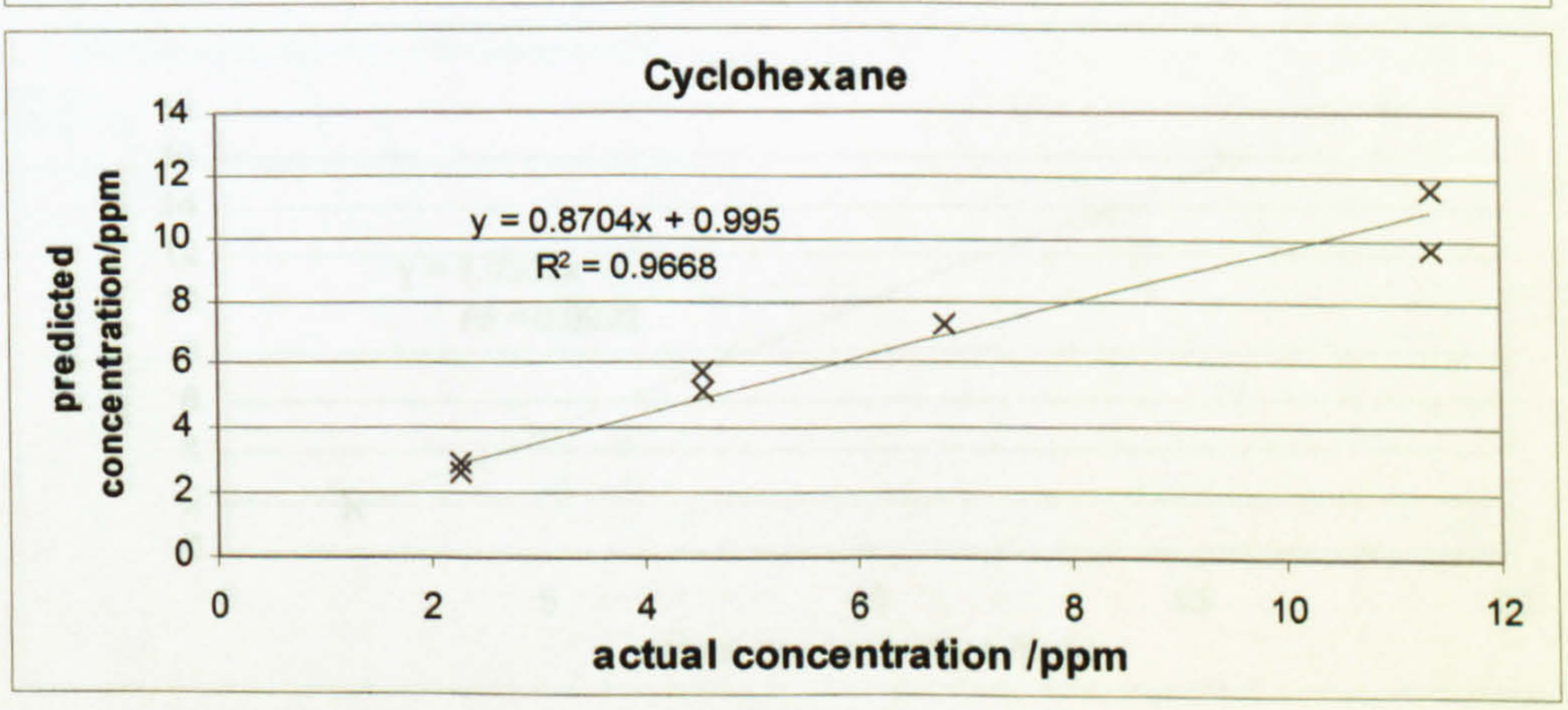

Ethanol
predicted concentration/ppm
y = 0.9054x + 2.9205
R² = 0.9574
0
10
20
30
40
50
actual concentration /ppm

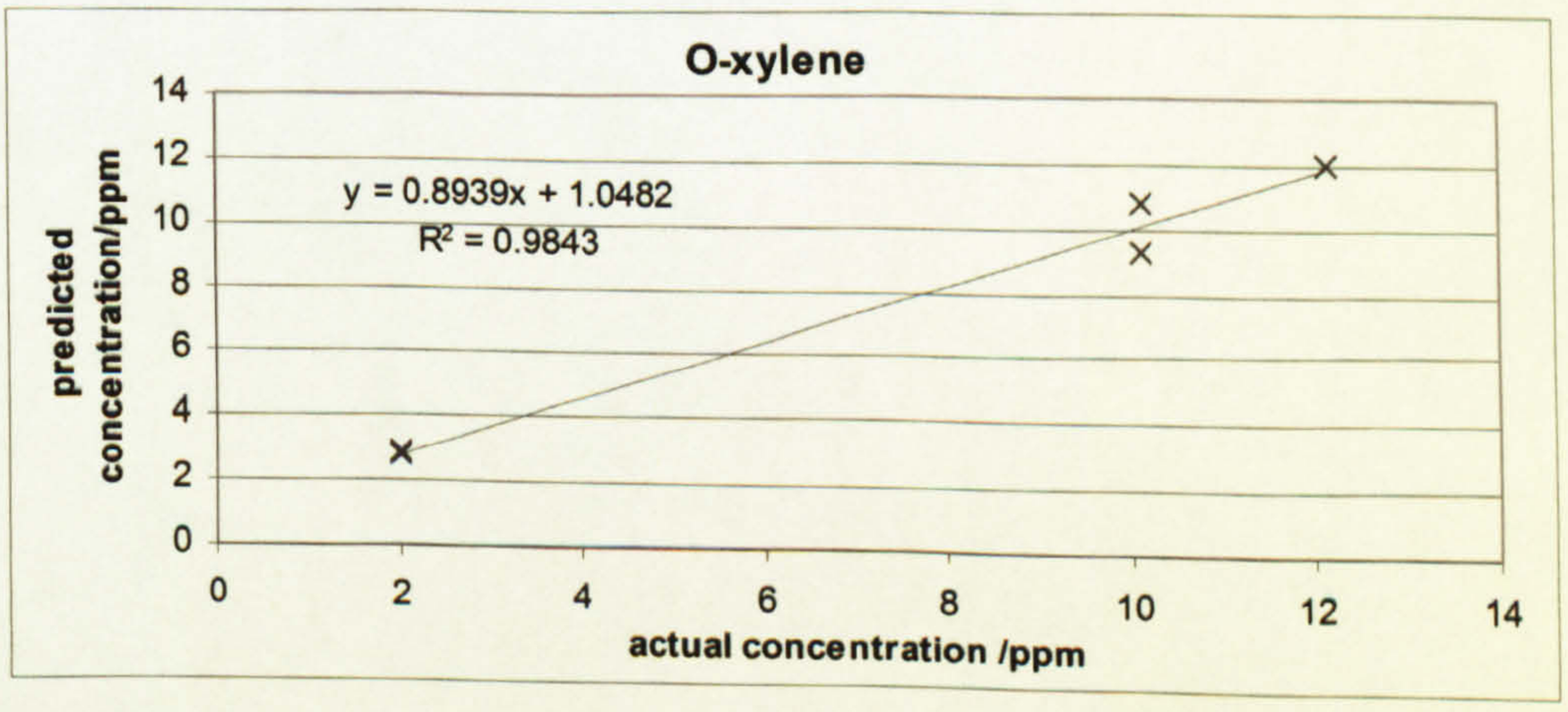

Limonene
predicted concentration/ppm
y = 0.3942x + 5.7486
R² = 0.664
0
2
4
6
8
10
12
14
16
actual concentration /ppm

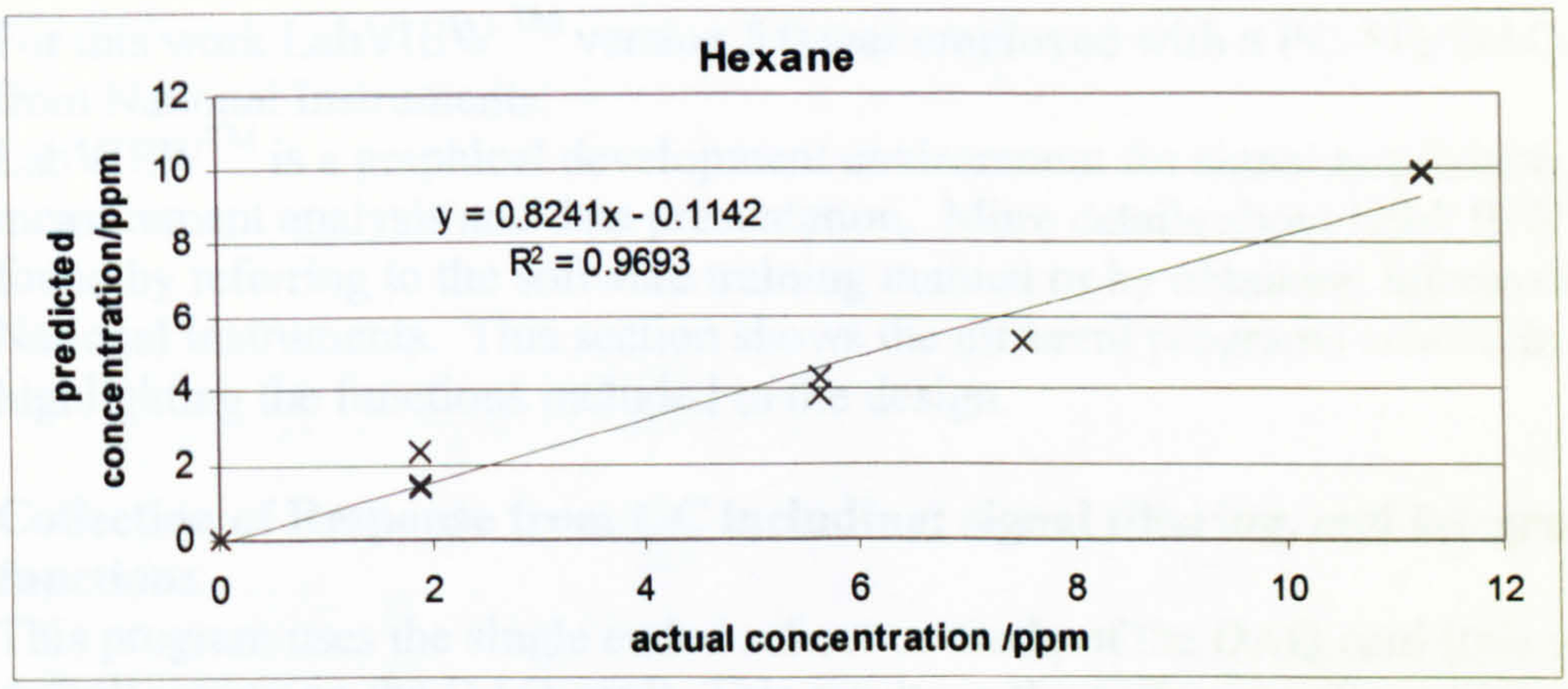

P-xylene
predicted concentration/ppm
y = 0.8197x + 1.8754
R² = 0.9422
0
2
4
6
8
10
12
14
16
actual concentration /ppm

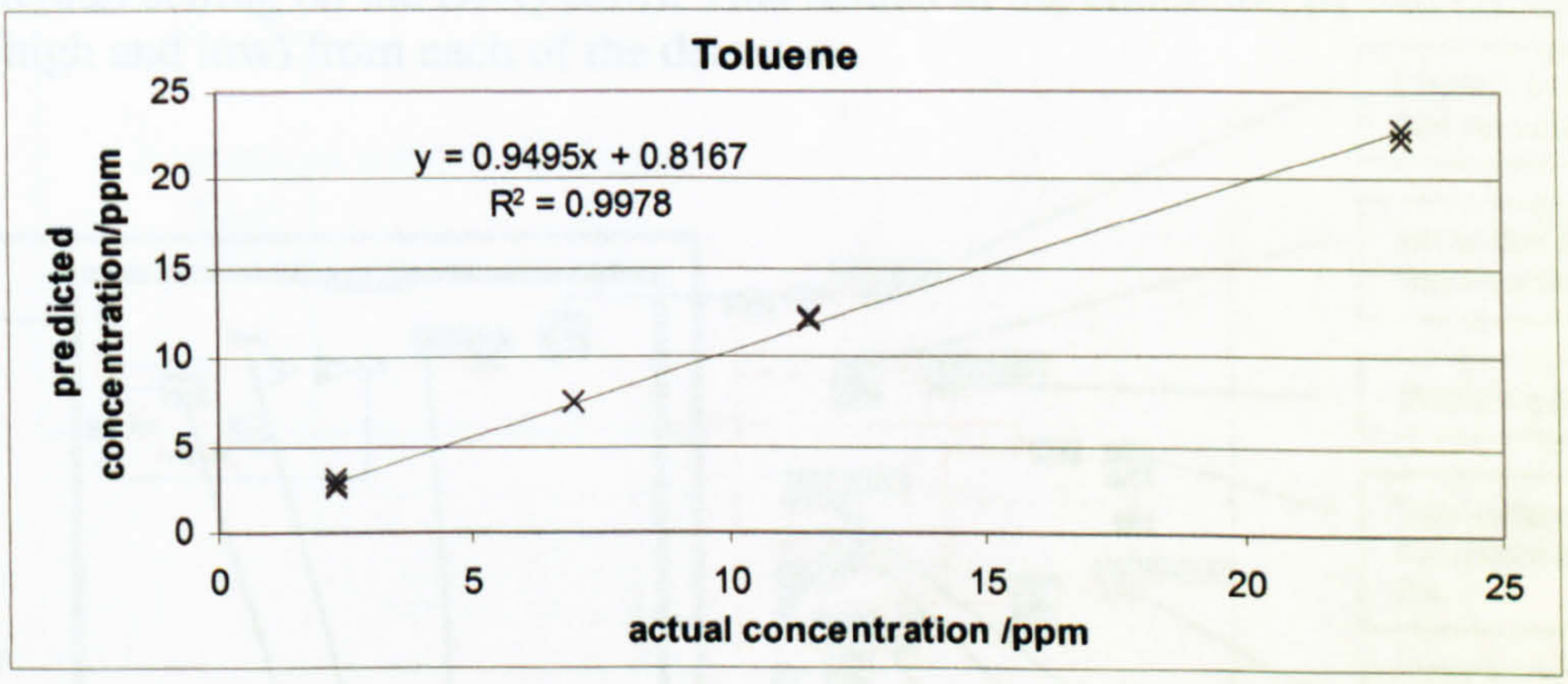

M-xylene
predicted concentration/ppm
0.0035
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
y = 0.0002x + 0.0003
R² = 0.9626
0
5
10
15
20
actual concentration /ppm

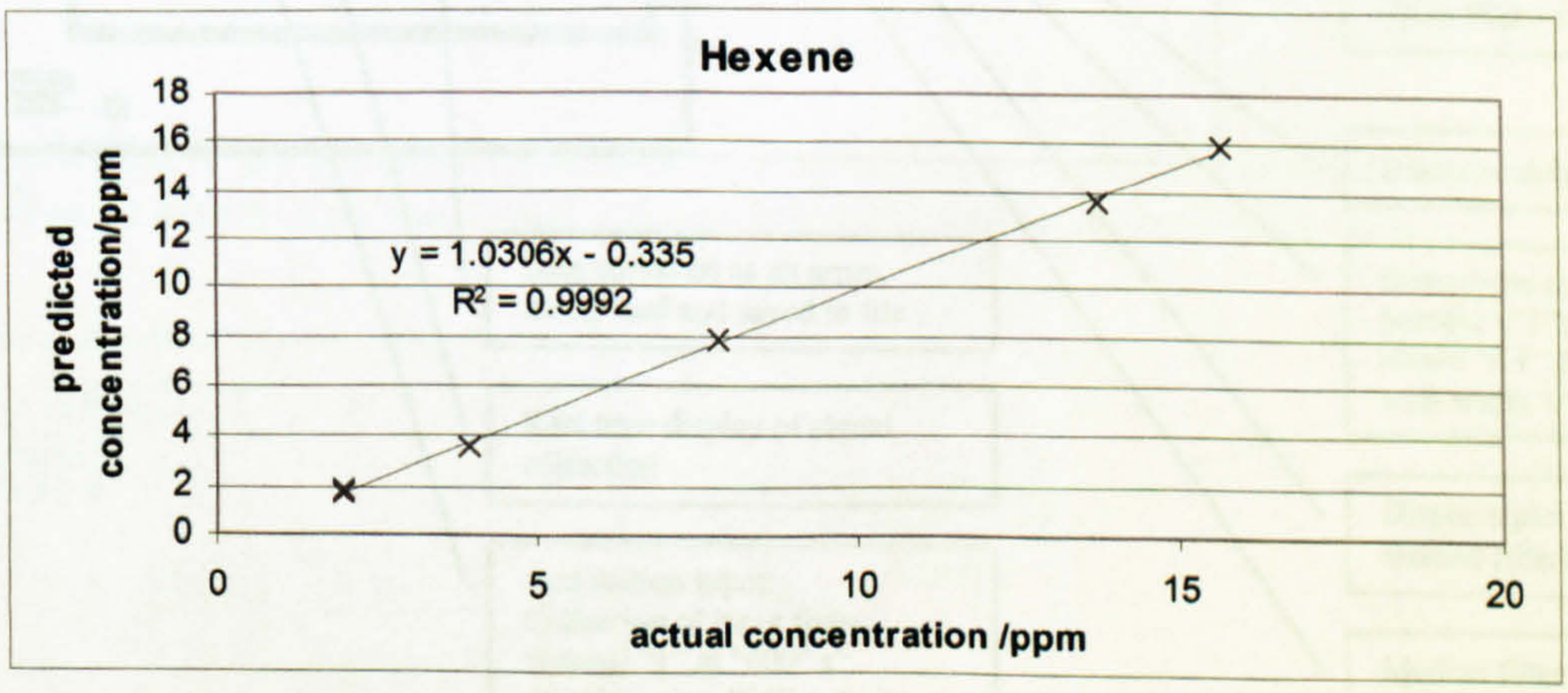

Propanoic acid
predicted concentration/ppm
18
16
14
12
10
8
6
4
2
0
y = 0.7633x + 2.4485
R² = 0.7887
0
5
10
15
20
actual concentration /ppm


Cyclohexane
predicted concentration/ppm
14
12
10
8
6
4
2
0
y = 0.8704x + 0.995
R² = 0.9668
0
2
4
6
8
10
12
actual concentration /ppm


O-xylene
predicted concentration/ppm
14
12
10
8
6
4
2
0
y = 0.8939x + 1.0482
R² = 0.9843
0
2
4
6
8
10
12
14
actual concentration /ppm


Hexane
predicted concentration/ppm
y = 0.8241x - 0.1142
R² = 0.9693
0
2
4
6
8
10
12
actual concentration /ppm


Toluene
predicted concentration/ppm
y = 0.9495x + 0.8167
R² = 0.9978
0
5
10
15
20
25
actual concentration /ppm

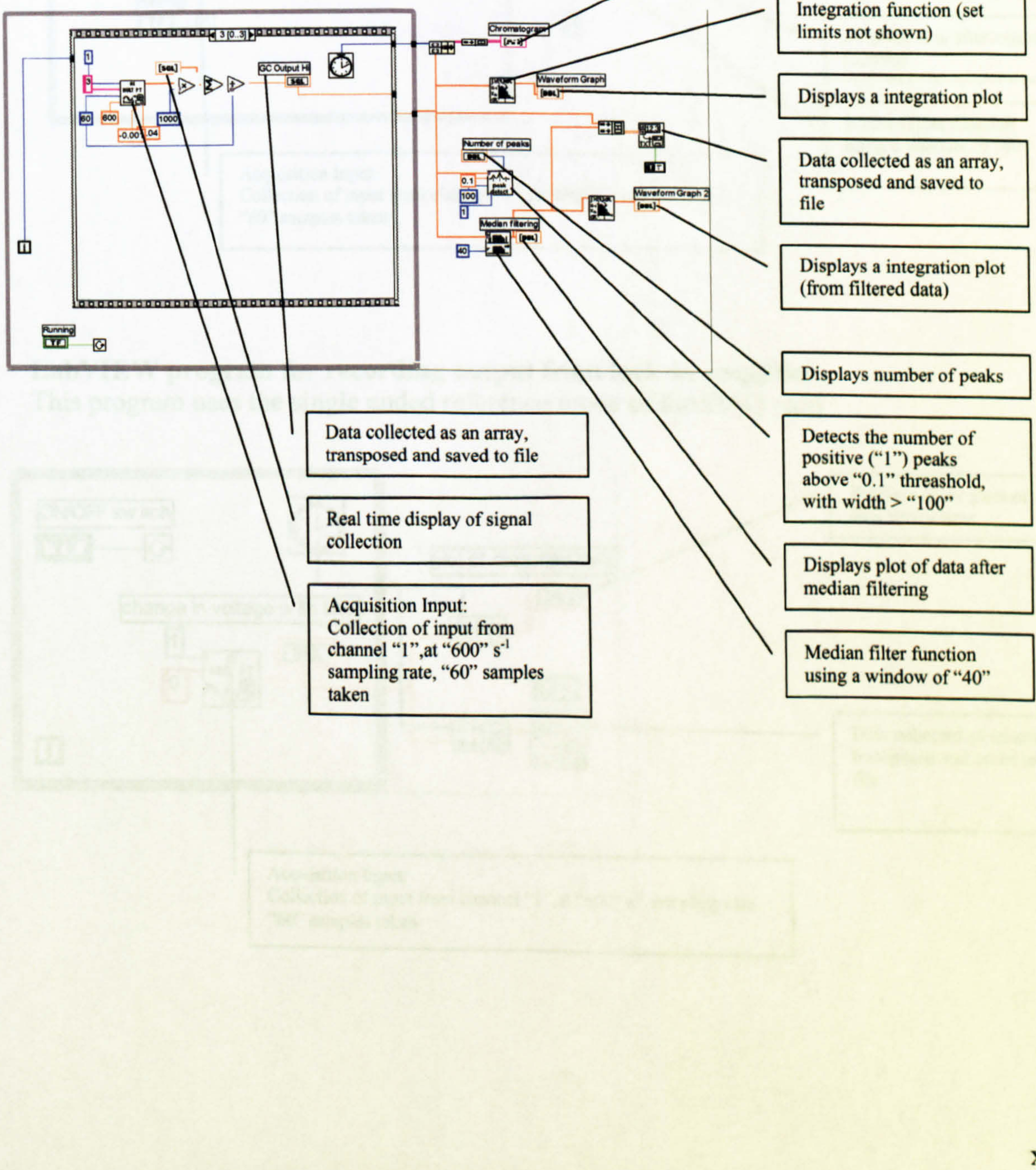

Hexene
predicted concentration/ppm
y = 1.0306x - 0.335
R² = 0.9992
0
2
4
6
8
10
12
14
16
18
0
5
10
15
20
actual concentration /ppm

6.7 LabVIEW™

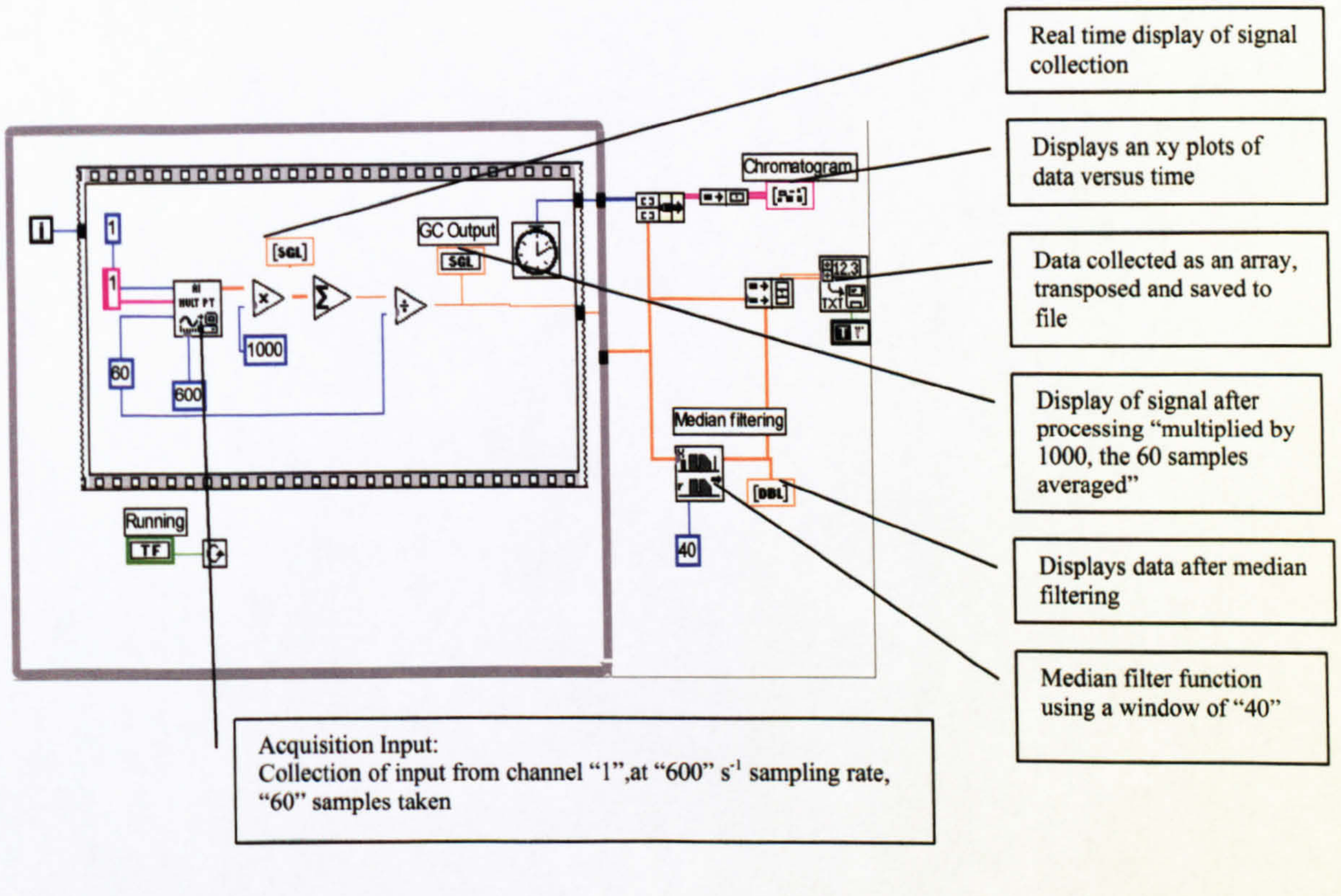
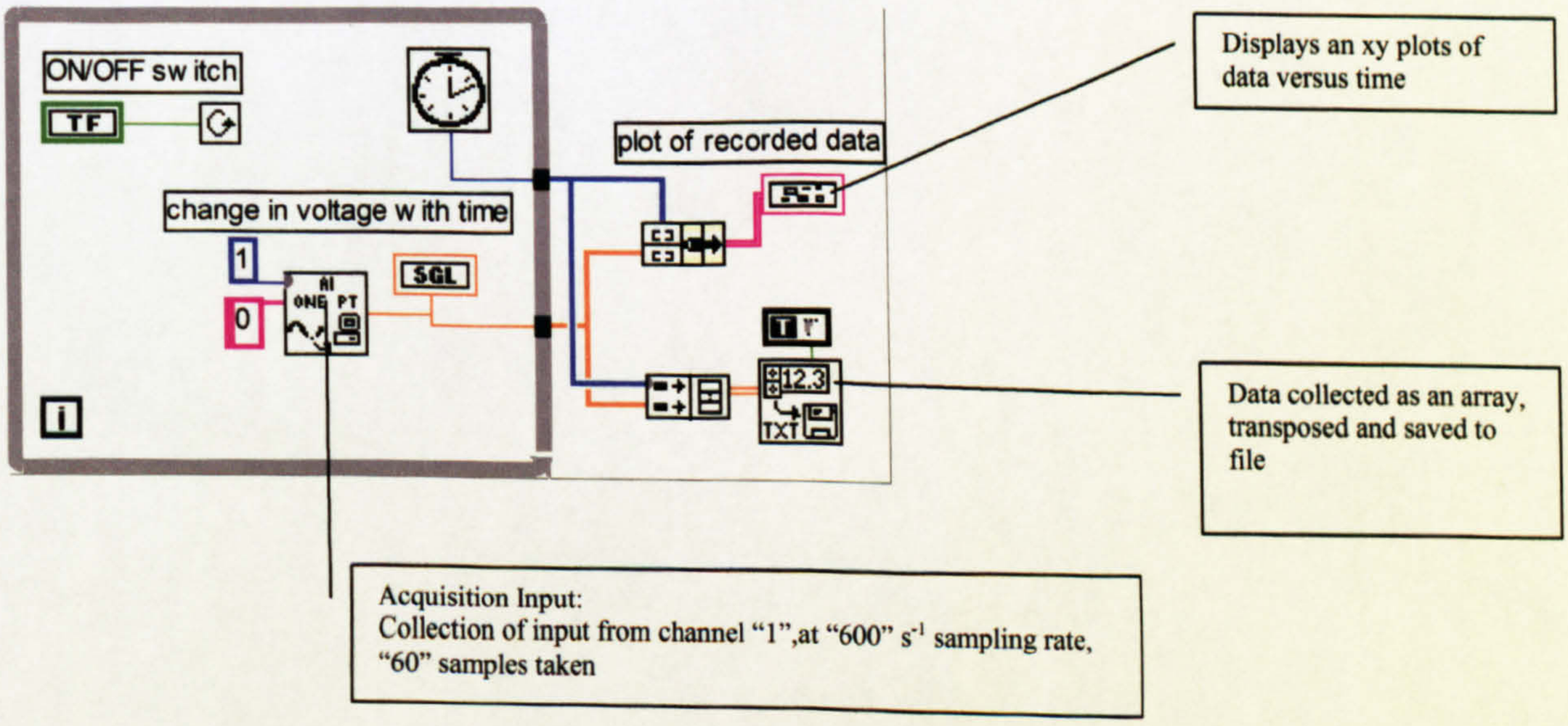
For this work LabVIEW ™ version 5.0 was employed with a PC-516/DAQ card ™ from National Instruments.
LabVIEW™ is a graphical development environment for signal acquisition, measurement analysis and data presentation. More details about LabVIEW™ can be found by referring to the software training manual or by obtaining information from National Instruments. This section shows the different programs created in LabVIEW highlighting the functions included in the design.

Collection of Response from GC including; signal filtering, and integration functions

This program uses the single ended reference mode of the DAQ card (this was the default setting on the DAQ card). This results in the collection of two GC responses (high and low) from each of the detectors.



Collection of Response from GC (using differential mode on DAQ)
This program uses the differential mode of the DAQ card. This results in the collection of a differential response between the two detectors in the TCD on the GC (i.e. a differential signal from the difference in conductance between the sample and reference detector)



LabVIEW program for recording output from lock-in amplifier
This program uses the single ended reference mode of the DAQ card