Measurements of Atmospheric Trace Gases Using Open Path Differential UV Absorption Spectroscopy for Urban Pollution Monitoring

Philip Nathaniel Martin
Imperial College of Science, Technology and Medicine
London SW7 2BW

August 2003

A Thesis submitted for the degree of
Doctor of Philosophy
of the University of London
and the
Diploma of Imperial College

Abstract

Urban air pollution is widely recognised as being one of the greatest environmental concerns to those living or working in urban centres. Key in dealing with any problem is understanding, and key in understanding is accurate measurement, without which the scale of the problem, trends, effectiveness of legislation and the accuracy of models cannot be evaluated. Measurement systems do exist; however, data are currently collected over timescales of the order of minutes or more and are often retrieved from samples removed from the point of measurement, making correlations with pollutant sources difficult. In order to facilitate more rapid timescale measurements of *in situ* samples, a short path length, photodiode-based, differential ultraviolet absorption spectrometer (DUVAS) unit was designed, constructed, characterized, calibrated and tested in an urban environment. This work covers the characterisation, calibration and first use of the measurement unit termed Generic UV Sensor Technologies and Observations (GUSTO). The emphasis of this work is on high throughput measurements of atmospheric trace gases, in particular sulphur dioxide, using open path differential UV absorption spectroscopy for the monitoring of pollution in an urban environment specifically by the roadside. We conclude that the unit is able to detect sulphur dioxide to parts per billion levels outdoors; measurements of benzene were also performed in the laboratory. The next steps necessary to achieve the long-term aims of the project are also discussed in detail. The author played an active role in all experiments both characterisation/calibration and generic data collection from which data presented in this text is derived. The author is also solely responsible for the software used to analyse these said data and played a major role in the development of the algorithms employed therein.

Acknowledgements

I would like to thank both the Atmospheric Physics and High Energy Physics groups for their respective contributions to the facilitation of this work and also my supervisors, Dr John Hassard and Dr Ralf Toumi for their support and encouragement. As part of the GUSTO project I acknowledge NERC for funding this research, and the other contributors to the GUSTO project for their financial support.

I would like also to gratefully acknowledge the invaluable contributions and support offered by Richard McArthy as the other GUSTO postgraduate student, Ray Wrigley as the engineering 'wizard' and Mark Richards as the GUSTO project manager, without whose respective contributions none of this work would have been possible.

I am grateful to my fiancée Lisa for her invaluable help and encouragement during the "thesis" and to my parents for all their support. I would also wish to thank my brother Matthew for telling me that there was no way I could possibly not overrun by less than six months; once he told me this I made it my aim to prove him wrong.

Finally I would like to thank my colleagues and friends in the HEP group for creating the distinctive environment which I have enjoyed being a part of during the course of this work.

Significance of Each Chapter

Chapter 1 – Introduces the thesis and gives a general introduction to urban air production followed by a more detailed discussion of the trace gases that are covered by current legislation pertaining to such pollution. The basic concepts applied by current legislation are also outlined to further allow the reader to see the wider context in which GUSTO-like systems can be applied. A summary of current measurement techniques is provided as is the need for, and aims of, the GUSTO project.

Chapter 2 – Describes the theory and methods underpinning the operation of the unit. Emphasis is placed on general DOAS and DUVAS systems (applicable to any system working within the DOAS/DUVAS framework) as well as the specific implementation relating to the GUSTO unit.

Chapter 3 – Outlines the experimental hardware that forms the current GUSTO prototype and sets the physical basis from which the characterisation of the system follows.

Chapter 4 – Explores the characterisation and calibration of the instrument, starting with the optics before moving on to the photodiode array and data acquisition hardware.

Chapter 5 – Explains the methods applied to the data retrieval, their implementation, with respect to qualitative and quantitative measurement together with a discussion of errors. The chapter ends by presenting the results of our findings.

Chapter 6 – Summarises the achievements with the current work and explores further areas of design, testing and research that can be applied with the GUSTO system.

Appendix – Detail of the role of government with respect to air pollution control together with the current and future legislative framework.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1 Overview

The environment in which we live is affected by pollution. This takes many forms and understanding these effects is key in developing useful strategies to minimize harm to the environmental system in which we all live. The first stage in understanding is measurement, for without accurate and representative measurements, existing theories cannot be tested and strategies cannot be evaluated. Also, as new data become available, new models will need to be developed as phenomena are observed.

Obtaining such data requires good instrumentation designed to collect accurate information, ideally in real time.

Many[1] pollutants are monitored across the UK at automated sites; there are currently 112 of these [1] (80% in urban areas) with the distribution being around 80% in England with the remaining 20% evenly spread across Scotland, Wales and Northern Ireland. The effects on human health that directly result from exposure to air pollution are difficult to measure, however the Government Committee on the Medical Effects of Air Pollutants (COMEAP) [2] estimates that around twenty thousand[2] deaths each year in the UK are directly linked to air pollution with a further five hundred thousand hospital admissions for respiratory related complaints aggravated by direct exposure to air pollution. The response of the human body to such exposure to air pollution is nonlinear [3] and therefore accurate understanding of potential health hazards is not fully facilitated by long time averages such as those that form the basis of measurements from the said monitoring networks. Conclusions drawn from a discussion on similar instruments covered in section 1.5 together with an introduction to generic trace gas monitoring techniques in section 1.4 leads to there being a lack of data as outlined in section 1.5.

The control of air pollution is entirely legislation driven (see further discussion in section 1.3). The passing of new legislation can only specify compounds that may be monitored, thus unless techniques exist with the required accuracy or integration times, legislation may not be drafted. The scope of a new instrument to take interesting new atmospheric data as well as act as a method for monitoring pollution levels leads directly to the Generic UV Sensor Technologies and Observations (GUSTO) instrument programme, an instrument uniquely suited to such applications and described in detail in sections 1.5 and chapter 3. This project is part of a general technology transfer programme with Atmospheric Physics in which High Energy Physics techniques and technology are applied in a more general sense. Further detail including the short and long-term aims is given in section 1.5. The methodology and strategy of the

[1] Currently monitored are O_3, NO_2, NO, CO, SO_2, PM_{10} and 25 hydrocarbon species, generally less than 6 types at each site

[2] Compare to figure of 430,000 total estimated deaths in urban areas of UK per year [2]

programme has been to construct and characterise a prototype instrument capable of measurements as outlined in section 1.5.2 with the ultimate long-term ambition of implementing some or all of the long-term aims also outlined in section 1.5.2. The role of government and past/current/new legislation is not strictly atmospheric physics and thus beyond the scope of this text, however it is discussed in detail in the appendix and outlined in the main text in section 1.3.

The conclusions, as detailed in chapter 6 of this text, are that the present GUSTO system works within the specified margins given by the original aims for the project (see section 1.5). Measurements have been taken with the unit in the environment within which it will be deployed (see chapter 5, section 2). Further work, including the necessary framework for achieving the longer-term aims is discussed in detail in chapter 6.

We will return to the subject of measurement and the techniques employed to look at the constituents of our atmosphere later in this chapter and then in more detail in following chapters. The remainder of this chapter deals with the question of what atmospheric pollution, and more specifically, urban air pollution, actually is and the ways in which it may be considered.

1.1 Urban air pollution

Human beings breathe in and out around once every four seconds, which equates to over eight million times a year. As a consequence each of us processes about four million litres (4,000m^3) of oxygen-containing air from the earth's surface, every year. Urban air pollution is one of the most important environmental problems that may be considered because of its impact on human health. It results from human activities and has diverse causes and sources. "Stationary sources," such as factories, power plants, and smelters; "area sources," which are smaller sources such as dry cleaners and degreasing operations; "mobile sources," such as cars, buses, planes, trucks, and trains;

and "natural sources," such as windblown dust and wildfires, all contribute to air pollution.

Air is made up of nitrogen (78%), oxygen (21%), argon, carbon dioxide and other gases in smaller quantities (1%). Therefore atmospheric pollution may be defined as the emission of quantities of substances that disturbs the physical and chemical properties of air.

Urban air pollution predominantly dates back to the industrial revolution in Europe and the US although control of emissions into the air may be traced back further. The following table 1.1 gives some brief historical aspects of air pollution.

Date	Description of Action/Event	Notes
1306	Royal proclamation issued forbidding the use of coal in London	Followed by a commission to punish miscreants 'for the first offence with great fines and ransoms, and upon the second offence to destroy their furnaces' [71]
1909	Polluted air in event in Glasgow, Scotland	1000 deaths. Word 'Smog' popularised after report on these deaths was presented by Dr. H.A. Des Voeux in Manchester (1911). This portmanteau was coined by him in 1905 [6]
1930	First causal relationships identified between high pollution levels and health hazards, Meuse Valley, Belgium	60 deaths. Belgian Commission of Investigation blamed SO_2, subsequent analysis points toward fluorine [7]. Both originated from stationary sources (factories) in the Meuse Valley
1948	Smog over Donora-Webster, Pennsylvania	48 deaths. Stationary source, local zinc factory (sulphur, carbon monoxide and heavy metal dusts) [8]
1952	London smog resulting in up to 4000 premature deaths	See [6], [9], [10], [11]and more detailed discussion below
1970's-1980's	Realisation that long range transport of pollutants from urban to rural areas could result in acid deposition and climate change	e.g. Acid rain first discovered by Robert Angus Smith, England 1872. Rediscovered by Eville Gorham, Canada, 1950. Rediscovered again by Svante Oden, Sweden, 1962

Table 1.1: Historical aspects of air pollution

An obvious question would be: Why do we see localised events similar to the ones outlined in the above table 1.1? The answer lies in the fact that the pollution emitted to the atmosphere by an industrial centre will always cause a build-up of pollutants unless the atmospheric conditions are such as to dissipate the pollution to surrounding areas. The pollutants may be transported either up into the atmosphere by rising buoyant air, carried along at a given altitude by low level winds or taken to the ground by deposition, or washed out by rain or reactions with other atmospheric constituents. In the case of a ground level temperature inversion (ground and surrounding air cools faster than higher altitude air, e.g. clear sky night time conditions) the situation is more complex. Pollutants released under the level of the inversion layer before the inversion is cleared (by heating of the ground etc) will be unable to escape as rising currents of cool air lose their buoyancy and are thereby inhibited from rising further when they reach the warmer, less dense air in the upper layers of a temperature inversion. Because high-pressure systems often combine temperature inversion conditions and low wind speeds, their long residency over an industrial area usually results in episodes of severe smog. Thus if the geography of an industrial area leads to low ground level wind speeds, as is the case if the industrial centre is located in a natural basin and if the atmospheric conditions dictate a high probability of a long lasting temperature inversion, pollution will stay trapped around its source and inevitably lead to an inflated and potentially dangerous build-up.

Take the London smog, or 'peasouper,' of 1952 as an example: In December 1952 the night time temperature inversion trapped enough pollutants such that the accumulation of smoke close to the ground was so great that the sun never broke through, and the air stayed cold and static. The dense yellow-black sulphurous *peasouper*, named after its colour as well as its density, persisted from the 5^{th} to the 10^{th} of December, an unprecedented five days, before it was taken down the Thames Estuary and out into the North Sea by winds from the west. At one point, the smog extended over a radius of 30 km around London, reducing visibility to between one and five metres. Suspended near ground level in the atmosphere were thousands of tonnes of black soot, particles of tar and gaseous sulphur dioxide, all of which came mostly from the domestic burning of coal. *The water from the fog condensed around the soot and particles, and the sulphur*

dioxide reacted inside these foggy, sooty droplets to form dilute sulphuric acid creating, in effect, a very intense form of acid rain [12]. The health effects of the original pollutants that formed the smog (soot and sulphur dioxide) was not as severe as the effect of breathing this acid aerosol, which causes irritation of the bronchial tubes, causing them to become inflamed and resulting in the production of large amounts of mucus [13]. This then directly caused people with bronchitis or asthma to choke on the mucus, resulting in death through suffocation or heart attack. Though the majority were old, the death rate among young children doubled during the week of the smog and it trebled for those aged between 45 and 64 [13]. It was only later when death records could be examined that the full extent of the event was discovered, the facts are summarised in figure 1.1 [14].

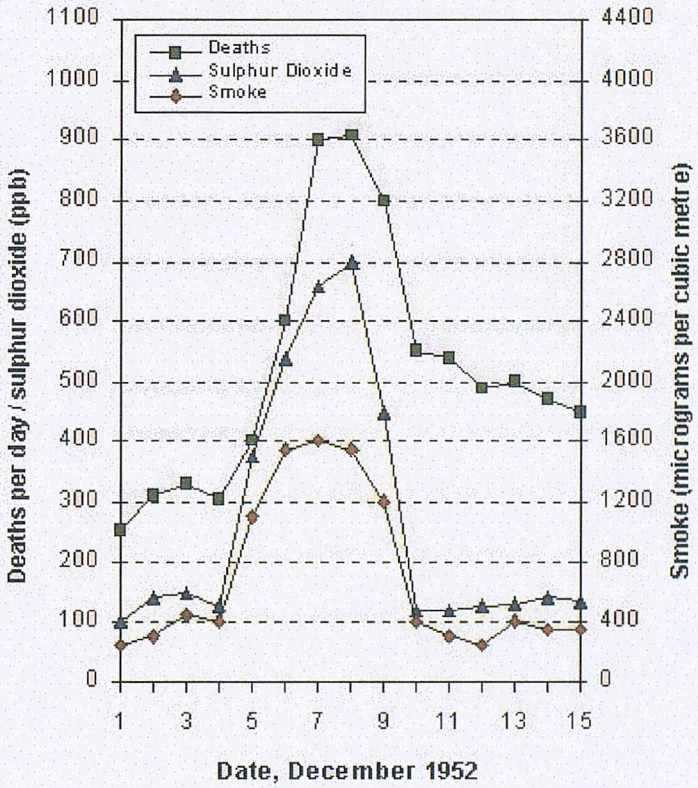


Figure 1.1: Daily mean smoke mixing ratio and daily mean deaths during London smog event 1952 [14]

As a result of the investigations after the 1952 event the UK government passed the Clean Air Act (1956). This authorised local councils to set up smokeless zones and

make grants to householders to convert their homes from traditional coal fires to heaters fuelled by gas, oil, smokeless coal or electricity. This may however have not had as much of an effect as at first seems: '*because already there was a shift away from the coal fire. Many people wanted all-electric houses. The legislation reinforced a change that was already under way*' [12]. Since this early legislation aimed at reducing pollution, much progress has been made, (daily averages for sulphur dioxide in London are now around 5ppb[3] and were closer to 1ppm in 1952 [16]).

Another technique to reduce the problem of pollution becoming trapped in a temperature inversion above an industrial centre is to utilise tall chimneys and thus emit the polluting material above the top of the inversion layer. While this approach disperses the plume, the pollution inevitably ends up somewhere. For example, polluting the air above the North Sea is preferable to the air above London (at least to Londoners). Many rural areas around the world receive pollution that has travelled far from its original source. The subject of pollutants travelling from point to point will be returned to later in this chapter.

The worst polluted cities/areas in the world include Sheyang and Xian in China and Mexico City in Mexico together with parts of Eastern and Central Europe – particularly Romania, Poland, Czech Republic and former states of the Soviet Union.

Taking Mexico as an example, atmospheric pollution is limited to high demographic or industrial density zones. Annual emissions surpass sixteen million tons with sixty five percent originating from transportation. Mexico City accounts for 23.6% of these emissions, with Guadalajara at 3.5% and Monterrey 3%, the remaining 70% coming from other industrial centres (2001 figures) [15]. Mexico City is a special case as not only is it one of the world's most populous cities with 21 million inhabitants[4] and 3 million vehicles, but the atmospheric conditions often lead to temperature inversions and at 2000 m above sea level oxygen pressure is 20% less than it is at sea level so fuels burn less efficiently.

[3] Compare to Mexico City daily mean of 0.13ppm [15] see section 1.2.1
[4] Only cities with more population in 2002, Tokyo 21 Million and New York 30 Million [17] figures are for urban area as opposed to city centre.

1.2 Pollutant species

Today emissions of air pollutants are covered in current legislation under emission categories, for example the US Environment Protection Agency produces a list applicable in the US 'National Emission Standards for Hazardous Air Pollutants' latest revision [18]. The main pollutants are classified into primary and secondary groups.

Primary pollutants: These stay in the atmosphere in the same state as they were emitted by their original source. For evaluation purposes they are usually considered to be sulphur oxides, carbon monoxide, nitrogen oxides, hydrocarbons and particulates.

Secondary pollutants: Those that experience chemical changes as a result of a chemical reaction between two or more atmospheric constituents such as photochemical oxidants and some radicals like ozone. A brief summary of the properties, sources, lifetimes and current efforts to limit/reduce emissions of the main members of both the primary and secondary groups will now follow.

1.2.1 Oxides of Sulphur

Sulphur oxide gases (SO and SO_2) dissolve in water. The primary member of this group is sulphur dioxide or SO_2, which is a colourless gas with a strong odour. Sulphur is prevalent in most industrial raw materials, including crude oil, coal, and ore that contains common metals like aluminium, copper, zinc, lead, and iron.

SOx gases are produced when fuel-containing sulphur, such as oil and especially coal, is burned, during mining and industrial processes e.g. when petrol is extracted from crude oil and naturally from volcanic eruptions. Many vehicles tend not to be significant sources because the sulphur content of the fuel is low, especially petrol (around 0.04%

by mass) however diesel contains larger amounts of sulphur (around 0.2% by mass) and as such there are some significant emissions from these vehicles.

Figures for the US [19] show that around 65% of SO_2 released to the air (more than 13 million tons per year), comes from electric utilities (power stations), especially those that burn coal and that other sources are industrial facilities that process sulphur containing ore/materials e.g. petroleum refineries. This is more than ten times the tonnage of the UK equivalent, which is closer to 1 million [1]. This difference is almost certainly due to the fact that the US electricity market is sourced by coal accounting for nearly 60% of total supply (2000 figures [20]) whilst coal fired generation of electricity in the UK accounts for only 28% of total supply[5].

The estimated breakdown of sources for SO_2 within the UK are illustrated below in figure 1.2 [1]

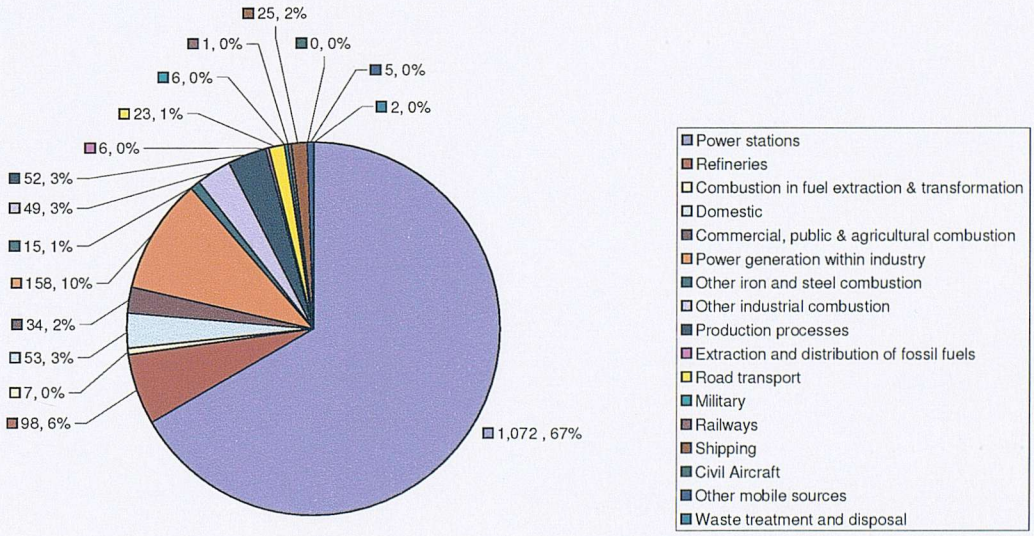


Figure 1.2: Sulphur dioxide emissions in the UK in 1998. Total 1608 thousand tonnes [1]

[5] 1999 figures [68], representing a fall of 58% since 1994.

Concentrations tend to reflect the extent to which coal and oil are burned and changes in such activity have greatly reduced emissions over the past 30 years. The annual UK SO_2 mean was at 68ppb in 1962 and had dropped to below 10ppb by 1992 [16], comparing starkly with the 0.13ppm daily mean in Mexico City [15], 13 times larger.

Sulphur dioxide is a natural component of the earth's atmosphere with natural emissions accounting for around 50-70 million tons per year, total anthropogenic emissions far exceed this however at between 150 and 200 million tons per year.

The lifetime of sulphur dioxide molecules in the troposphere is a few days. Conversion to aerosol takes around 1.5 days, wet deposition in 7/3 days for gas/aqueous, dry deposition in 2/7 days for gas/aqueous [80]. It is removed from the troposphere by

- Oxidation to H_2SO_4:
  - $2SO_2 + O_2 \rightarrow 2SO_3$
  - $SO_3 + H_2O + M \rightarrow H_2SO_4 + M$. (1a)

- Reaction with hydroxyl radical [80], which is the most important sink process, where M is an energy absorbing chemical species such as O_3, N_2:
  - $SO_2 + OH\cdot + M \rightarrow HO\cdot SO_2 + M$
  - $HO\cdot SO_2 + O_2 + M \rightarrow HO_2\cdot + SO_3 + M$
  - $SO_3 + H_2O \rightarrow H_2SO_4$. (1b)

- Dissolving in fog, rain drops to form a dilute solution of sulphurous acid [80]:
  - $SO_2(g) \leftrightarrow SO_2(aq)$
  - $SO_2(aq) + H_2O \leftrightarrow SO_2\cdot H_2O(aq)$
  - $SO_2\cdot H_2O(aq) \rightarrow HSO_3^-(aq) + H^+(aq)$
  - $HSO_3^-(aq) + H_2O_2(aq) + H^+(aq) \rightarrow SO_4^{2-}(aq) + 2H^+(aq) + H_2O$. (1c)

- Directly, by way of an uptake on aerosols and clouds, which leads to dry and wet acid depositions.

The lifetime of sulphur dioxide in the stratosphere is several weeks, during which it produces sulphate aerosols. Clean continental air contains less than 1 ppb of sulphur dioxide [16].

Health effects of SO_2 gas are irritation to the eyes and respiratory system, reduced pulmonary functions and aggravation to respiratory diseases such as asthma, chronic bronchitis and emphysema. Exposure to extremely high concentrations will cause permanent damage to the respiratory system as well as extreme irritation to the eyes (due to production of dilute sulphuric acid around the eyes). When SO_2 reacts with other chemicals in the air to form tiny sulphate particles, these may also be inhaled in which case they gather in the lungs and are associated with increased respiratory symptoms and disease, difficulty in breathing, and premature death [19].

The acidification of clouds and rain (SOx converts to SO_4 in 1.5 days, lifetime of SO_4 is around 3 days) affects the environment by such means as metallic corrosion, deterioration of electrical installations, paper, textiles, paints, construction materials and historical monuments. There may also be substantial effect on plant life due to leaf injuries and reduction of photosynthesis. Water based eco systems such as lakes and rivers are affected if their acidic content is changed. Sulphate particles in the air also impair visibility [19].

Governments have introduced legislative measures to control SO_2 emissions by targeting such areas as the installation of pollution control equipment at coal-fired power plants, reducing pollution from industrial processing facilities, reducing the average sulphur content of fuels burned, and using cleaner fuels like natural gas for residential and commercial heat. The plan in the US began in 1995 by targeting the highest emitting power plants and moved into its second phase of setting tighter controls on other emitting installations in 2000, the aim being to reduce the emissions to half the 1980 level between 1980 and 2010 [19].

Sulphur dioxide is one of the primary target species for the GUSTO unit as its absorption is large within the target range (see figure 2.3.6, range is 200-260nm) and it is one of the most common trace gases to be measured with the existing (5 minute and upward) averages.

1.2.2 Nitrogen Oxides

Nitrogen Oxides or NOx (in the case of pollution NO, NO_2 and NO_3) are a group of highly reactive gases, all of which contain nitrogen and oxygen in varying amounts. Many of the nitrogen oxides are colourless and odourless. However, nitrogen dioxide is a brownish gas with a strong odour.

The main source of NOx is high temperature combustion in industries and vehicles (i.e. $N_2 + O_2 \rightarrow 2NO$ which then leads to the production of NO_2 by direct oxidation of NO and through a variety of other reactions [80]), though some is produced in electrical storms. The following figure 1.3 shows the breakdown of sources for the UK [1]:

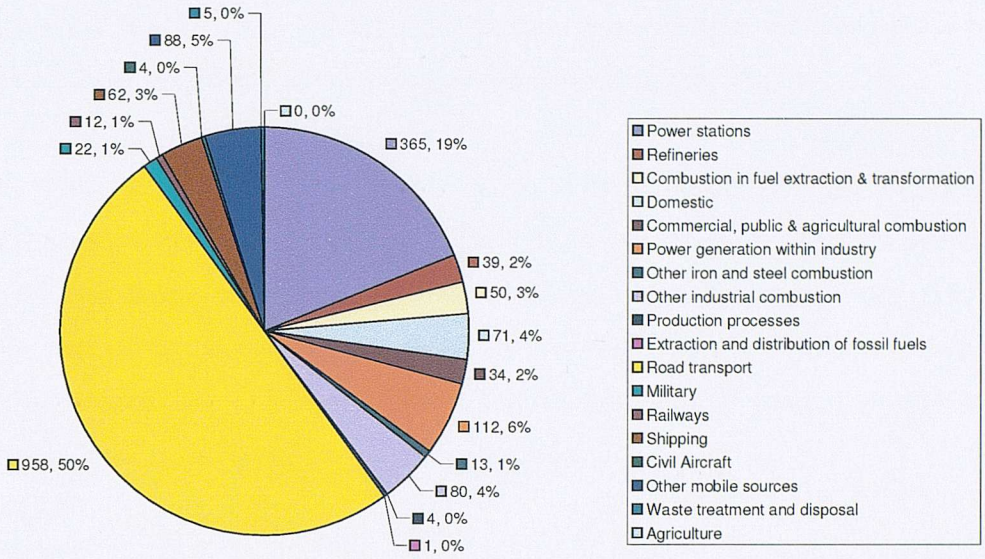


Figure 1.3: Nitrogen oxides emissions by source for the UK in 1998. Total 1918 thousand tonnes (NO_2 and NO) [1]

Natural background levels of NOx (in this case NO and NO_2) within rural UK districts are between 1 to 4 ppb [21]. Urban areas of the UK have roughly averaged concentrations of NO_2 of between around 20ppb since 1976 [21]. The generally accepted reason for the apparent lack of change in concentrations is that while there has been a reduction in nitrogen dioxide emissions from industrial sources, there has been a rise in emissions from road transport [21]. NOx emissions in the US have increased by around 10% since 1970 due mainly to increased emissions from motor vehicles [19]. The maximum hourly concentration in Mexico City is around 0.21ppm [15] a clear order of magnitude above the maximum levels observed in the UK or US.

Nitrogen oxides are also a natural component of the earth's atmosphere, with natural emissions accounting for around 10-20 million tons per year and total anthropogenic emissions of around 30-50 million tons per year. The average residence time of nitrogen oxides in the atmosphere is approximately days. Conversion to aerosol takes around 1.5 days, wet deposition in 7/3 days for gas/aqueous, dry deposition in 3/7 days for gas/aqueous [80]. Conversion to nitric acid is by reaction of NO_2 with OH radicals,

by nighttime reactions involving O_3 or by hydrogen abstraction reactions NO_3, HCHO, and hydrocarbon radicals [80]. Nitric acid is in turn removed from the atmosphere by direct deposition to the ground, or transfer to water droplets (clouds/rain).

The health effects of nitrogen dioxide are irritation to the lungs and lowered resistance to respiratory infections such as influenza. The effects of short-term exposure are still unclear, however frequent exposure to concentrations higher than rural background levels may increase incidences of acute respiratory disease in children [19]. Prolonged exposure to nitrogen dioxide can also lead to the decolouration of painted objects and premature leaf loss and inhibition of growth in plant life.

Nitrogen oxides are one of the main ingredients involved in the formation of ground-level ozone, which can trigger serious respiratory problems. It contributes to formation of acid rain and nutrient overload that deteriorates water quality [19]. NOx can also react readily with common organic chemicals and even ozone, to form a variety of toxic products (causing for example, biological mutations), e.g. the nitrate radical, nitroarenes, and nitrosamines [22].

Legislative measures aimed at controlling NOx emissions run along the same lines as those for SO_2 outlined above as both are important in the formation of acid rain. However levels are yet to begin to noticeably drop in the UK [21] and have in fact increased in the US [19].

Oxides of nitrogen do absorb in the GUSTO wavelength range, however, while NO has absorption that is comparable to the other absorbers (though still lower than SO_2, see figure 2.3.6), NO_2 has lower absorption cross sections, meaning that the accuracy of the instrument must be improved beyond that required to meet the targets set for other species. For example if an accuracy of ± 1ppb were obtained for SO_2 then the accuracy with the same original signal height and retrieval algorithm would be around ± 10 ppb for NO and closer to ± 50 ppb for NO_2.

1.2.3 Carbon Monoxide

Carbon monoxide or CO is a colourless, odourless gas that is formed when carbon in fuel is not burned completely. The primary source is from road vehicle emissions though other sources include industrial processes (e.g. chemical manufacturing) and natural sources such as forest fires. Figure 1.4 shows the breakdown of sources for the UK [1]

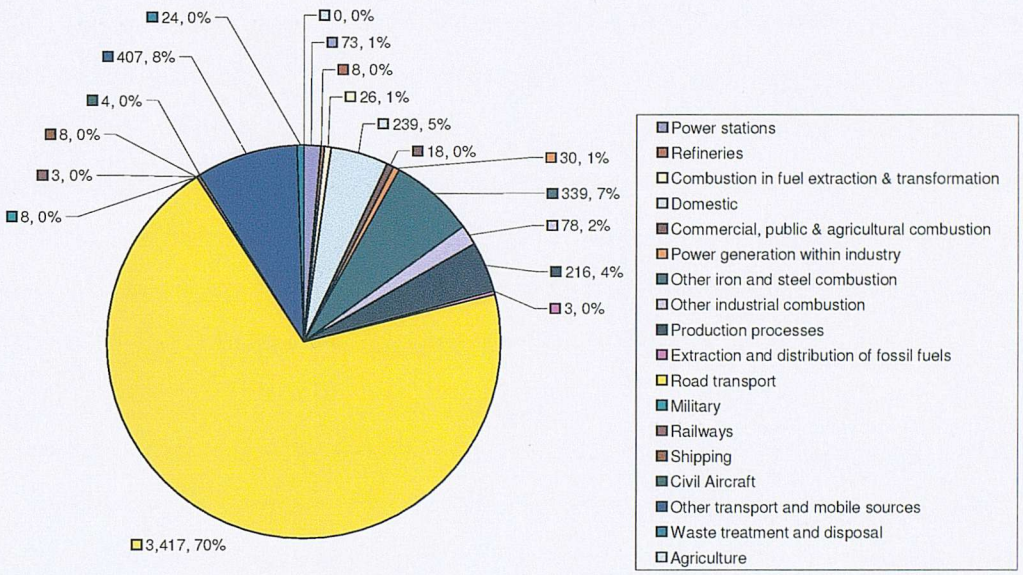


Figure 1.4: Carbon monoxide emission distribution in the UK in 1998. Total 4902 thousand tonnes [1]

Natural background levels of CO are around 10-200ppb [1], with levels in urban areas being highly variable, depending upon weather conditions and traffic density. The total average levels for the UK is around 1ppm [21], however central locations (e.g. London) are closer to 2ppm and may regularly vary up to 5ppm (8 hour mean) [1]. For comparison, levels in Mexico City are closer to 11ppm [15] (8 hour mean). Trends in recent years have shown a yearly fall in average concentrations of order 0.1ppm over most of the country and over twice that in central London [16].

Natural emissions of CO account for around 1300 million tons per year with emissions due to human activity being of order 700 million tons per year. The average residence time in the troposphere is around one month before oxidising to CO_2 [16]. Levels of CO may affect ozone levels by reacting with an hydroxyl radical [80] in a similar manner to VOCs, see section 1.2.4.

CO can have a serious effect on human health. It has an affinity for haemoglobin two hundred times that of oxygen [1], and thus impairs the oxygen carrying ability of blood resulting in oxygen starvation. The effects of exposure to CO are most serious for those who suffer from heart disease [19] (e.g. angina) in which case a single exposure to low levels may cause chest pain and reduce the ability to exercise. Moderate exposures to CO can affect the central nervous system, resulting in vision problems, reduced ability to work/learn and reduced manual dexterity [19]. The escalating symptoms of carbon monoxide poisoning are headaches, vomiting and in severe cases collapse and death, although the effects of brief exposure are reversible.

The current legislation aimed at cutting carbon monoxide emissions is almost entirely aimed at cars. Since January 1993, all new petrol-driven cars sold in the European Union (EU) have been fitted with a catalytic converter [23]. In the US during winter months in certain urban areas petrol is oxygenated by adding oxygen containing compounds such as alcohols or their derivatives [19]. At optimum operating temperature (around 300 degrees centigrade [24]) a catalytic converter will convert carbon monoxide, nitrogen oxides and volatile organic compounds (see later section on VOCs) to less harmful compounds (water, nitrogen and carbon dioxide) with an efficiency of up to 90% [23]. On short journeys this temperature is not usually reached, thus future legislation will probably include the fitting, given successful development, of some form of pre-heater for catalytic converter technology. Clearly converting CO to CO_2 raises as many questions as it answers, as CO_2 is itself a green house gas.

Currently levels of urban carbon monoxide in developed nations are dropping (see above). The primary source of urban carbon monoxide (road vehicles, see fig 1.4) is increasing in concern as the number of vehicle miles travelled rises (e.g. around 25%

every 10 years, 1981-1990 US data [19]), however current efforts to reduce the emissions from any given car engine appear to be very effective (e.g. around 25% decrease in 10 years, 1981-1990 US data [19]). The UK government's target [1] of 10ppm (maximum running 8 hour mean) seems likely to be met for the entire country by the set date of 31 December 2003 if current trends continue.

1.2.4 Volatile Organic Compounds

Volatile Organic Compounds, or VOCs are volatile carbon-based compounds that are produced as unburned fuel by-products in a combustion process. The most common types that are monitored are benzene (C_6H_6) and 1,3 butadiene (C_4H_6) primarily because they are known carcinogens. Benzene is a minor constituent of petrol (EC legislation states that it must be less than 5% by volume, average content in UK petrol is 2% by volume [23]). 1,3 Butadiene also comes predominantly from combustion employed by road transport, however it is not a constituent of petrol but comes rather from the combustion of olefins[6]. The sources in Europe for VOCs are around 70% from the combustion of petrol, 10% from petrol refining, distribution and evaporation and the remainder from other sources [21]. The specific breakdown by source for the UK is given below in figure 1.5 [1].

[6] The group of unsaturated hydrocarbons of the general formula C_nH_{2n} such as ethylene and propylene

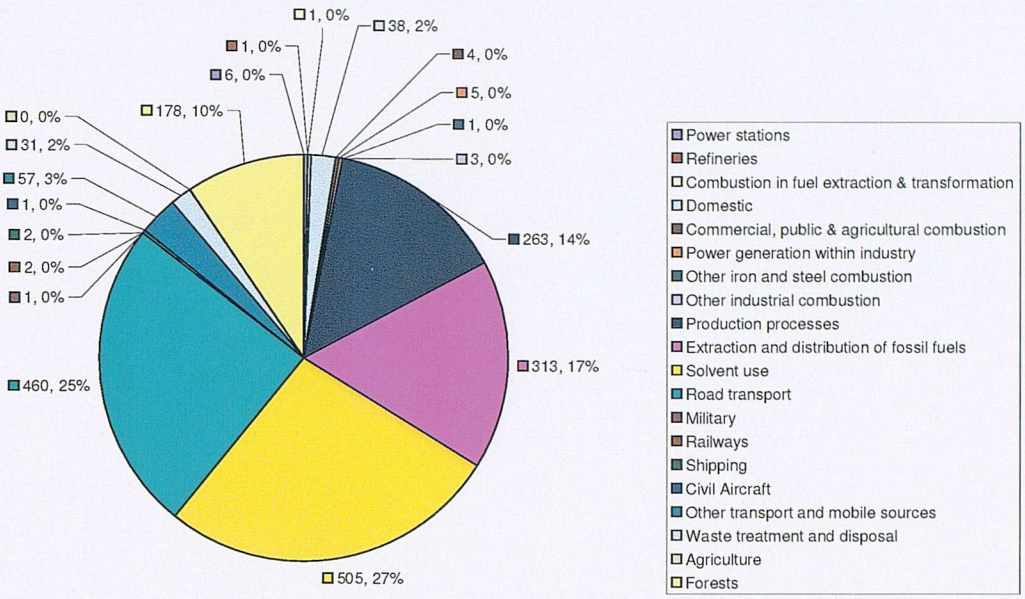


Figure 1.5: VOC emissions by source for the UK in 1998. Total 1871 thousand tonnes [1]

Background mixing ratios of benzene and 1,3 butadiene in the UK are around 0.1ppb, while urban levels (e.g. London) are around 1ppb and 0.5ppb respectively [21]. This compares starkly with the mixing ratios in Mexico City of maximum hourly averages of up to 0.2ppm [15].

Both benzene and 1,3 butadiene are carcinogens, however the main health hazard from these and other VOCs is their role in the production of ground level ozone. The production of tropospheric ozone is summarised in equation 1.1 [25]:

$$\begin{aligned} NO_2 + h\nu &\rightarrow NO + O(^3P) \\ O(^3P) + O_2 + M &\rightarrow O_3 + M \\ O_3 + NO &\rightarrow NO_2 + O_2. \end{aligned} \qquad (1.1)$$

Nitrogen dioxide is photo-disassociated to form nitrogen oxide and a free oxygen radical, and this oxygen radical then reacts with molecular oxygen to form ozone. The basic cycle is then completed by the reaction of the ozone formed in the second reaction with the nitrogen oxide from the first reaction to form oxygen and nitrogen dioxide.

Thus there is an interchange between NO and NO_2, and NO_X is the catalyst for the production of tropospheric ozone.

Hydroxyl radicals (free radical from the reactive oxygen species with formula OH, sourced by the photo-disassociation of certain atmospheric trace species [25], [26]) will oxidise VOCs and yield by products which further react so as to alter the balance of equation 1.1. The case for a generic saturated hydrocarbon (RH) results in the production of RO_2 which reacts with NO as in equation 1.2,

$$RO_2 \cdot + NO \rightarrow RO \cdot + NO_2. \qquad (1.2)$$

The reaction shown in equation 1.2 has two effects: firstly it results in the net production of ozone when taken with reactions in equation 1.1, that is $RO_2 + O_2 \rightarrow RO + O_3$, and secondly it eliminates ozone destroying NO. Both of these effects bias the equilibrium of troposphereic ozone levels to higher concentrations.

Another by product of the oxidation of VOCs by attack from the hydroxyl radical is to produce a further hydroxyl radical, thus perpetuating the reaction and allowing more VOCs to be oxidised in a similar manner. For a more detailed discussion of the other less dominant reactions and a quantitative treatment of reaction rates the reader is referred to [25]. The atmospheric half-life of benzene is around 12 days [25]; ground ozone is discussed in more detail in the following section.

VOCs, specifically benzene, do absorb considerably in the GUSTO wavelength range (see figure 2.3.6). Although the absorption cross sections are less than those of SO_2 they are larger than those of the other pollutant species in the same range. Thus along with SO_2 benzene is the logical first choice gas with which to characterise the system. At room temperature benzene is a liquid and thus harder to work with than SO_2. High concentrations (allowing a small sample to evaporate) are easy to obtain while lower concentrations require more technique (control of extremely small samples of liquid).

1.2.5 Tropospheric Ozone

Ozone (O_3) is a colourless gas formed at ground level by reactions involving VOCs and nitrogen oxides (see above equations 1.1 through 1.3). There are no terrestrial sources of ozone, however any that is formed, as in equation 1.1 above, will also be destroyed, assuming that the VOCs/other compounds that shift the balance of the reaction toward high ozone content are no longer present. Thus the levels of tropospheric ozone will fall only when the heat/sunlight required is not present or the VOCs have broken down. Ground level ozone can be transported great distances[7] by the prevailing winds [19].

Maximum hourly averages in the UK can exceed 200ppb compared to background levels of between 10-30 ppb [19], in Mexico City normal hourly average concentrations are of order 110ppb [15]. Both in the UK [21] and the US [19] there appears to be no noticeable trend in average yearly ground ozone concentrations over the past 20 years.

Short-term exposures (1 to 3 hours) to moderate ozone concentrations have been linked to increased hospital admissions for respiratory complaints. Repeated exposures are linked with increased susceptibility to respiratory infection, lung inflammation and aggravation of pre-existing respiratory diseases such as asthma. Other health effects of exposure to ozone are decreases in lung function and increased respiratory symptoms such as chest pains and coughing [19]. All of the symptoms of exposure to ozone appear aggravated by periods of moderate exertion and children active outdoors are the group at greatest risk of developing symptoms during levels of high ozone concentration [27]. In addition, long-term exposures to ozone present the possibility of permanent changes in lung function, which could lead to premature ageing of the lungs and/or chronic respiratory illnesses [19].

The primary targets to control the formation of ground level ozone are aimed both at stationary sources (of VOCs, NOx etc) from power stations/industry and at mobile sources (cars). Considerable improvements have been made, for example in the US, cars coming off production lines emit 70% less NOx and 80-90% less VOCs now

[7] Of order hundreds of Km, source US Enviromental Protection Agency.

compared to 1960 [19]. There are however substantially more car miles being driven (partly due to increased numbers of cars and partly due to increased use of each car). Current trends both in the UK and US do not show substantial change in ground ozone levels; probably the most promising idea (both UK [21] and US [19]) is to target the highest polluting vehicles as this small percentage is thought [19] to contribute substantially to the total observed concentrations.

Ozone is another example of a trace gas absorbing within the GUSTO wavelength range (see figure 2.3.6), however the small absorption features that would be used by a DUVAS system are lower in magnitude than both those of SO_2 and benzene. It would logically not be the first trace gas to characterise with the prototype unit.

1.2.6 Particulate matter

The classification of particulate matter in the atmosphere is done according to size, with particles that have a diameter of x microns or less, termed PM_x. Research during the 1990s provided epidemiological evidence that fine particles can damage human health even at concentrations previously thought to be unimportant [21]. In recent years $PM_{2.5}$ have also come under increasing study, as they are the type able to penetrate most deeply into the lungs (and lodge once penetrated). Particles larger than PM_{10} are not readily lodged in the lungs, and are removed relatively efficiently from the air by sedimentation. Particles are often classed as either primary (emitted directly into the atmosphere) or secondary (formed or modified in the atmosphere from condensation and growth).

The major sources of primary particles are combustion processes, typically diesel combustion; in general the transport of hot exhaust vapour into a cooler exhaust pipe or stack can lead to homogeneous nucleation of carbon-rich particles before emission [29]. Secondary particles are typically formed by chemical reaction, for example the oxidation of sulphur dioxide to sulphuric acid. The atmospheric lifetime of particulate matter is a strong function of particle size, but may be as long as 10 days [29] for

particles of about 1μm in diameter. The sources of PM_{10} in the UK are illustrated below in figure 1.6 [1]

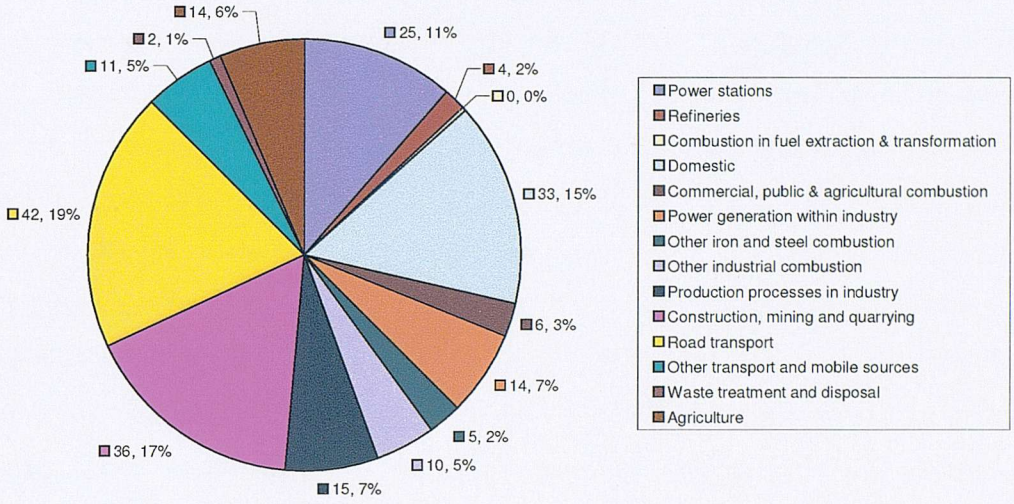


Figure 1.6: PM_{10} emission source distribution in the UK in 1998. Total is 216 thousand tonnes [1]

Since first monitored in 1992 particulate concentrations in the UK have been steadily decreasing from 35 to 25 μgm^{-3} [1]. A significant proportion of PM_{10} in the air over England and Wales comes from mainland Europe (highly dependent on prevailing weather conditions). This results in elevated concentrations in the east of England of around 17μgm^{-3} as compared with concentrations on the west coast averaging around 7μgm^{-3} [1]. Concentrations in Mexico City are a daily average of 150μgm^{-3} [15].

Exposure to inhalable particulate matter (PM_{10}) has been shown to have the following effects [30]: Increased respiratory symptoms and disease; and decreased lung function; alterations in lung tissue and structure; decrease in respiratory tract defence mechanisms. Groups that appear to be at greatest risk are children, the elderly and individuals with cardiopulmonary disease such as asthma. Emissions of particulate matter are also linked with the deterioration of construction materials and other surfaces.

The risk posed by PM_{10} to health is recognised as the greatest of the six primary atmospheric pollutants [32]. Legislation pertaining to the control of particulates in the UK revolves around the objective of $40\mu g/m^3$ (annual mean) set by the Air Quality Strategy for England, Scotland, Wales and Northern Ireland, which is likely to be met by the 31 December 2004 target date [21]. However the 24-hour objective ($50\mu g/m^3$; not to be exceeded more than 35 times a year) is likely to be exceeded in most large urban centres, especially alongside busy roads and in the vicinity of some industrial plants, e.g. large steel works [1]. The primary method to reduce particulate levels is to improve the combustion processes that are responsible for the primary emissions. For example, in the US, since 1994 engines for new diesel trucks have been built to meet the US EPA's emission standards, which requires that particulate matter emissions be reduced by 90%. Remodelled buses also reduce particulate matter emissions (more than trucks) while the EPA is continuing to work with manufacturers to further reduce emissions from diesel engines, including non-road engines (e.g. construction equipment) [19].

Particulate matter does not have distinct absorption features, making it unsuitable for a DUVAS system, however they do absorb with a wavelength dependence. The form of this dependence is predicted as λ^{-1} from theoretical consideration of small black carbon spheres with constant refractive index [34], however the assumption that all particulate matter may be treated as small black carbon spheres with constant refractive index is not universally accurate, with the reader being referred to [34] for a more detailed discussion. Particulates are usually measured by observing the scattering from laser based sources, however with suitable algorithms some type of concentration or change of concentration information may be obtainable from a GUSTO-like system.

1.2.7 Lead

Reduction of lead emissions is probably the biggest success story so far in the quest for control of urban air pollution since the problem was first recognised. Historically the

major source of lead emissions was from motor vehicles. Figure 1.7 illustrates the situation in the US [19] for 1970 and 1997 respectively.

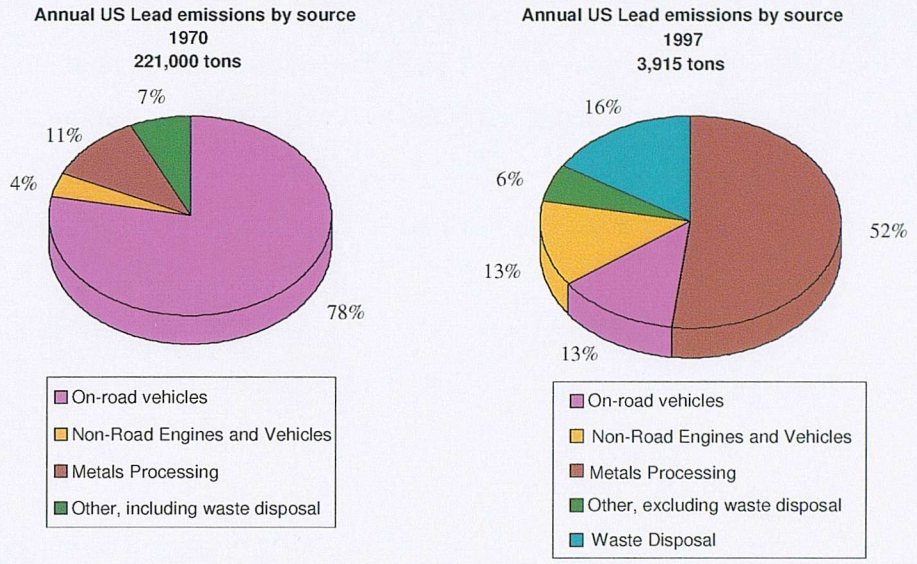


Figure 1.7: Annual US Lead emissions 1970 and 1997 showing 98% reduction in total emission

Annual average lead concentrations in the UK air have decreased from $0.59\mu gm^{-3}$ in the late 1980s to around $0.15\mu gm^{-3}$[1]. The only sites currently exceeding the 2004 objective ($0.5\mu gm^{-3}$) are in the vicinity of a small number of secondary non-ferrous metal smelters [21]. By 2005 virtually all urban and roadside sites are expected to meet the 2008 objective set by the Air Quality Strategy of $0.25\mu gm^{-3}$.

The health effects of lead are marked, especially in children. Exposure damages organs (including the kidneys, and liver) and affects the brain and nerves leading to risks of seizures, mental retardation, behavioural disorders, memory problems and mood changes. Even low levels of lead damage the brain and nerves in foetuses and young children resulting in learning difficulties and lowered IQ [19]. Other effects apparent particularly in men are high blood pressure and increased incidence of anaemia. Exposure to lead also affects animals, plants and fish in a similar way to humans, i.e. alterations to neurological and blood related systems.

The dramatic effects of legislation aimed at reducing levels of lead in the air have been achieved by the reduction and eventual elimination of lead used for road vehicles (e.g. initial legislation to phase out leaded fuel introduced in the US in 1975 and final legislation banning leaded fuel for highway traffic in 1995[8] [19]). The only remaining sources are high performance engines (e.g. drag racing), general aviation piston aircraft and industrial sources, with the emissions from the latter decreasing by a small amount (around 6% since 1988). Current efforts at reducing health risks associated with exposure to airborne lead are focusing on areas local to such industrial facilities.

A further reason for eliminating lead in petrol is the fact that it coats the catalytic surface of a catalytic converter, thereby totally suppressing the operation of the device.

1.2.8 Species Summary

A summary of pollutant properties is shown in table 1.2.1 [34], [80]

Species	Primary Source (UK)	Emission Level [35]	Mean UK Urban Concentration [1]	Recent UK Trend [21]	Average Residence Time*
SO_2	Industry	Global Natural: 50-70 Mtons Human: 150-200 Mtons UK 1kt	10ppb (2ppb background)	Down 85% 1962-1992	Aerosol – 1.5 days Deposition wet 7/3, dry 2/7 days (gas/aq)
NO and NO_2	Transport	Global Natural: 10-20 Mtons Human: 30-50 Mtons UK NO_2: 1Mt	NO_2: 20ppb (1-4ppb background)	None	Aerosol – 1.5 days Deposition wet 7/3, dry 3/7 days (gas/aq)
CO	Transport	Global Natural: 1300 Mtons	1ppm (10-200ppb background)	Down 80% 1976-1998**	1 Month

[8] US EPA data shows a 95% drop in lead emissions from the transportation sector (1980-1999) together with a 94% drop in lead air concentrations for the same period.

		Human: 700 Mtons			
VOC	Transport	UK 15kt	Benzene: 1ppb 1,3 butadiene: 0.5ppb (both 0.1ppb background)	UK trend not available	Days
Ozone	Not directly emitted		10-30ppb	None	Days
PM_{10}	Transport	UK data not available	$25\mu gm^{-3}$	Down 30% 1992-2002	Days
Lead	Industry		$0.1\mu gm^{-3}$	Down 75% 1980-2000	Days

*See main body text for references
** Figure is unreliable due to early measurements probably being overestimates

Table 1.2.1: Summary of properties of urban air pollutants

1.3 Legislative Controls

Legislation in the UK pertaining to air pollution has its origins as far back as 1306 when original controls on the use of coal in London were introduced. From this point onward various Royal Proclamations and Government acts [5] targeted specific primary pollution sources, for example the Railway Clauses Consolidated Act of 1845 required railway engines to 'consume their own smoke.' The Public Health Act of 1875 contained a section on smoke abatement from which legislation to the present day has been based. The first Clean Air Act in 1956 was targeted at the introduction of smoke control areas, chimney heights and the prohibition of dark smoke[9] emissions from installations (with some exceptions). This act was then extended[10] in 1968 to add further prohibitions on the emission of dark smoke. The founding of the European Community (European Coal and Steel Community, ECSC, 1951 followed by the European Atomic Energy Community, EURATOM, 1957, followed by the European Economic Community, EEC, Treaty of Rome[11] 1957) resulted in the first EC Directive (70/220/EEC) in 1970 introducing restrictions on the emissions of carbon monoxide and hydrocarbons (now VOCs) from motor vehicle engines. A further 14 EC Directives and

[9] The term 'dark smoke' is a legally defined shade of grey
[10] Also further extended in 1993
[11] As amended by the 1965 Merger Treaty and 1986 Single European Act

6 government acts followed between 1972 and 1995 when the UK passed the Environment Act providing a statutory framework for local air quality management. Current legalisation is related to the Environment Act of 1995; however further detail is beyond the scope of this text. In the interests of completeness and for the interested reader, a more in depth discussion of the role of government, current legislative framework and future legislative developments is given in the appendix.

1.3.1 Role of Government

The UK government objectives for the reduction of air pollution as from the last Strategic Air Quality Review (2000), are summarised in the following table 1.3.1 and 1.3.2 [36]

Pollutant	Objective		Date to be achieved by
	Concentration*	Measured as	
Benzene	16.25µg/m³ (5ppb)	running annual mean	31 December 2003
1,3-Butadiene	2.25µg/m³ (1ppb)	running annual mean	31 December 2003
Carbon monoxide	11.6mg/m³ (10ppm)	running 8 hour mean	31 December 2003
Lead	0.5µg/m³	annual mean	31 December 2004
	0.25µg/m³	annual mean	31 December 2008
Nitrogen dioxide‡	200µg/m³ (105ppb) not to be exceeded more than 18 times a year	hour mean	1 December 2005
	40µg/m³ (21ppb)	annual mean	31 December 2005
Particles (PM_{10})	50µg/m³not to be exceeded more than 35 times a year	24 hour mean	31 December 2004
	40µg/m³	annual mean	31 December 2004
Sulphur dioxide	350µg/m³ (132ppb) not to be exceeded more than 24 times a year	hour mean	31 December 2004
	125µg/m³ (47ppb) not to be exceeded more than 3 times a year	24 hour mean	21 December 2004
	266µg/m³ (100ppb) not to be exceeded more than 35 times a year	15 minute mean	31 December 2005

* Conversions of ppb and ppm to µg/m³and mg/m³at 20°C and 1013mb.
‡ The objectives for nitrogen dioxide are provisional [36].

Table 1.3.1: UK Government objectives included in regulations for purposes of local air quality management [36].

Pollutant	Objective		Date to be achieved by
	Concentration*	Measured as	
Objectives for the protection of human health			
Ozone‡	100μg/m³ (50ppb) not to be exceeded more than 10 times a year	daily maximum of running 8 hour mean	31 December 2005
Objectives for the protection of vegetation and ecosystems			
Nitrogen oxides†	30μg/m³ (16ppb)	annual mean	31 December 2000
Sulphur dioxide	20μg/m³ (8ppb)	annual mean	31 December 2000
	20μg/m³ (8ppb)	winter average (1 October to 31 March)	31 December 2000
* Conversions of ppb and ppm to μg/m3 and mg/m3 at 20°C and 1013 mb. ‡ The objective for ozone is provisional [36] † Assuming NOx is taken as NO_2.			

Table 1.3.2: UK Government national objectives not to be included in regulations for purposes of local air quality management [36].

1.3.2 International Influence

Once in the atmosphere, pollution may travel with the prevailing winds far from its point of origin (e.g. see paragraph 1.3.5 on tropospheric ozone) thus the problems associated with pollution are not in any way restricted to being purely a national concern. Many of the principles agreed at the UN Rio Conference [37] revolve around the concept of nations taking responsibility for the ways in which their pollution emissions affect other nation states. In the US, many of the pollution targets are derived in part for any given state by considering how they may be prevented from unduly contributing to another state's pollution levels, and a number of states have more stringent pollution objectives enforced by federal legislation specifically to enable downwind states to reach nationally set pollution targets [19].

Recent estimates suggest that more than 73% of the UK's sulphur dioxide emissions and as much as 85% of nitrogen oxides are transboundary, leading to deposition in other countries (14% of SO_2 and 24% of NOx) or the sea[12] [38]. There is increasing recognition that there are flows of pollutants in the other direction to the UK from

[12] Figures are for 1997, used in model referenced [38].

mainland Europe. Around 41% of the deposition (as opposed to the ambient levels) of oxidized nitrogen in the UK originates from sources outside the UK[13] and up to one half or more of peak ambient ozone levels across the south of the country also originate outside the UK [36]. Further there is evidence from the Airborne Particles Expert Group [39] (APEG) concluding that emissions in mainland Europe contribute up to 20% of annual mean urban background levels of primary PM_{10} in the UK and that this may be much larger during short-term peak episodes. Similar research from the APEG [39] concludes that the contribution to secondary PM_{10} concentrations (also annual mean average urban background) is, during years of typical meteorology, of order 15%, which is around 50% of secondary particles. In years where the easterly winds are more frequent they make up a higher proportion of PM_{10} concentrations, particularly in the south and east of England.

The UK government [36] estimates that around 80% of UK domestic environmental legislation results from European proposals, with membership of the United Nations Economic Commission for Europe (UNECE) bringing further responsibilities (see discussion in the appendix). An example of the Europe wide approach, giving rise to the need to change existing emission targets, already in part implemented by existing legislation is outlined below. The new targets are stricter than the older ones, due to the impact of the UK on mainland Europe; the corresponding targets for European nations whose emissions affect the UK have been similarly derived [40].

In December 1999 the UK signed a new protocol to reduce acidification, eutrophication[14] and ground level ozone. This protocol covers the same four pollutants as the proposed National Emissions Ceilings Directive (NECD) but is for the wider European region as opposed to the EU. Thus the new emissions ceilings for 2010 are as in the following table 1.3.3

[13] Source: As for 8.

[14] Process by which aquatic systems gradually age and become more productive, anthropogenic *eutrophication* is water pollution caused by excessive plant nutrients [19].

Pollutant	Emission ceiling for 2010 (kt)	Percent decrease from 1998 level
SO_2	625	61
NOx	1181	33
VOCs	1200	39
Ammonia	297	20*
* Percent decrease calculated from 1999 data.		

Table 1.3.3: Emission ceilings for NECD pollutants 2010.

1.4 Monitoring Techniques

The two main aspects of control with respect to urban air pollution are firstly targets for ambient levels for specific pollutants and secondly the monitoring network that gives feedback on the current/past ambient levels. Thus far urban air pollution has been detailed covering the history, concept, current framework, legislation and the role of the government together with details of the specific air pollutants and their current trends/ambient levels within the UK and wider context. Many methods exist to measure the level of a given pollutant over a given time either *in situ* or by taking samples.

The six main requirements for the monitoring of atmospheric trace gases are generally listed as [34]

1. Feasibility of detecting numerous compounds with one instrument
2. High sensitivity in order to permit the detection of very low concentrations
3. High selectivity in order to differentiate between different species present in a multicomponent mixture
4. Large dynamic range in order to monitor low and high concentrations with a single instrument
5. Good temporal resolution to enable on-line monitoring
6. Good portability for *in situ* measurements

In order to detect atmospheric trace gases one must at some point sample the atmosphere. This may be achieved by either directly extracting a sample or monitoring *in situ*/remotely. Direct sampling has two primary drawbacks. Firstly, there are possible problems involving reactions/contamination with regards the container used. Secondly, direct sampling complicates real time and or continuous measurements. Remote monitoring generally involves the use of satellite data and is thus extremely expensive.

A basic comparison of the concepts underpinning the standard monitoring techniques is given below. For the purposes of this text monitoring techniques will be divided into two groups, those that rely on spectroscopic measurement methods and those that do not.

Non-spectroscopic techniques do not rely on spectroscopic features but use some form of physical or chemical process for measurement. They are generally described according to categories based upon the primary process employed. Spectroscopic techniques are based upon some form of interaction between radiation and matter. They are generally further subdivided into those that use conventional light sources and those that utilise laser based sources. Laser based sources are extremely selective, being able to target specific excitation states of the target trace gas. The only way to monitor multiple target species with a laser based technique is to utilise a tuneable source and rapidly scan across the chosen wavelength range. Such systems are also prohibitively expensive.

1.4.1 Non-Spectroscopic Techniques

These types of techniques are generally applied to samples of gas removed from the point of origin. Sensitivity is generally good or excellent with variable selectivity depending on the chemical process. These techniques are often used in the laboratory, though some methods are used for *in situ* measurement within automatic monitoring networks [16]. There are a number of techniques that are not good enough to measure a

specific trace constituent without first separating it out of a group containing that constituent. These are usually combined with gas chromatography. Examples of these are flame ionisation detection where the ionisation occurs in combustion and yields the total hydrocarbon concentration, and mass spectrometry where fragmentation and ionization occurs by electron bombardment. One can also ionise a sample by UV radiation and β radiation (limited dynamic range). None of these techniques is suitable for real-time *in situ* measurements.

The most common non-spectroscopic detection scheme used in trace gas detection is chemiluminescence. This technique relies on the radiation emitted from exited molecules produced from a chemical reaction whose excitation depends on the type and concentration of compounds involved in the reaction. This technique is applied to monitoring levels of ozone NO/NO_2, wherein NO_2 can only be indirectly monitored by first converting it to NO by photodissociation. Commercial NO/NO_x monitors are commonly employed for routine *in situ* air control measurements [34].

The disadvantage of the above non-spectroscopic techniques is that no single method of this type can be applied to fast, *in situ* data retrieval of the main trace gas species detailed in the government guidelines table 1.3.1 and 1.3.2.

1.4.2 Spectroscopic Techniques

Properties of specific trace gas with respect to its interaction with specific wavelengths of light allow comparisons on the basis of wavelength response.

Laser induced fluorescence is a technique where the sample under investigation is hit with a specific laser frequency. The induced excitation of the target molecule results in fluorescence which may be detected, thus allowing the concentration of the original target to be evaluated. High resolution studies require cooling of the molecules to remove spectral congestion and to reduce the Doppler width of the transitions [74]. It is only suitable for detection of atoms or radicals and molecules that can be resonantly

excited. The excitation source for molecular laser induced fluorescence is typically a tuneable dye laser in the visible spectral region and with respect to air pollution this technique is usually achieved with an optic fibre output and input. For species of pesticides and oil derivatives that are studied in the UV, other chemicals need to be added in order to facilitate reactions to species that are optically active in the visible spectral range.

Long path absorption spectroscopy with lasers is a technique which refers to using lasers over a long path to investigate the associated absorbance of the sample under study. Often two laser sources are used, one on and one off resonance with a particular excitation of the molecule under study. This technique is often applied in long single-pass or multiple-pass cells or with single or double ended schemes in the open atmosphere. It utilises an integrated response and is generally within the UV-Vis range.

Light detection and ranging (LIDAR) is the technique based upon a pulsed laser light source emitting periodically very short pulses of light (duration of each pulse generally lower than 20ns). This light is backscattered in the atmosphere by particles and molecules towards the telescope which collects the light to be analysed, both for the time elapsed between light emission and reception as well as for the absorption of the light in the atmosphere [77]. Monowavelength LIDAR has a detector matched to the wavelength of the light emitted by the source and analyses the intensity of the backscattered light. This backscattered light is a function of the density, shape, size and refractive index of particles in the atmosphere. Differential LIDAR consists of two emitted beams at wavelengths chosen to correspond to a maximum and a minimum of absorption of the target gas. Raman LIDAR consists of a monowavelength light source but where the detector analyses the frequency shift of the light inelastically scattered by the target gas. Fluorescence LIDAR is the application of LIDAR to the fluorescence of the target molecule (inelastic scattering). All the LIDAR techniques utilise atmospheric back scattering of laser pulses. They allow three-dimensional profiling of pollutants but are complex and mainly applied to the UV and Visible (SO_2, NO_2, O_3).

Photoacoustic spectroscopy with diode lasers is the technique where the sample in a closed cell is irradiated with modulated light. The energy absorbed by the analyte molecules is totally or partially transferred to their adjacent gas molecules by inelastic collisions, inducing an increase of their kinetic energy and thus a rise of gas temperature. Since the exciting light is modulated with a frequency determined by the resonance conditions of the cell, and the gas is enclosed in a constant volume, a periodic change of pressure results. These periodic pressure fluctuations (sound) are detected by a sensitive microphone. Tuning of the emitted wavelength (bandwidth about 0.1pm) is possible by the temperature and/or the injection current of the laser within a range of a few nanometres. Thus, a certain component in the gas mixture can be excited selectively to a rotational-vibrational transition by absorption of the laser light. To modulate the light either a chopper is used or the emission wavelength of the diode laser is modulated. Thus the technique involves measurement of the absorbed energy in the sample cell over a short path length. This technique is mainly applied to the infrared [78] and allows selective monitoring of numerous pollutants.

Infrared spectroscopy, including FTIR (Fourier transform infrared) is a technique where the absorption of radiation by a typical organic molecule results in the excitation of vibrational, rotational and bending modes, while the molecule itself remains in its electronic ground state. Matrix isolation involves trapping the substance to be examined in an inert solid at low temperature; typically argon or nitrogen is used as the matrix material. The concept of Fourier transform as applied to IR spectroscopy is to pass the light through an interferometer, thus introducing a frequency shift (dependent on path difference). The output is thus taken in the frequency domain and must be Fourier transformed in order to recover the associated wavelength dependent intensity. This technique spreads the noise across all frequencies and thus reduces the associated errors. Absorption is in the fundamental IR range using large molecular cross sections, generally molecule specific. A typical light source is an electrically heated silicon carbide strip otherwise known as a glow-bar [75].

Raman scattering involves the inelastic interaction between photons and irradiated molecules. Like Rayleigh scattering, the Raman scattering depends upon the

polarisability of the molecules. For polarisable molecules, the incident photon energy can excite vibrational modes of the molecules, yielding scattered photons which are diminished in energy by the amount of vibrational transition energies. A spectral analysis of the scattered light under these circumstances [76] will reveal spectral satellite lines below the Rayleigh scattering peaks at the incident frequency. Such lines are called 'Stokes lines'. If there is significant excitation of vibrational excited states of the scattering molecules, then it is also possible to observe scattering at frequencies above the incident frequency as the vibrational energy is added to the incident photon energy. These lines, generally weaker, are called 'anti-stokes lines'. This technique utilises small cross sections and has selectivity by frequency shift. Only one wavelength is needed. Sensitivity is poor due to low level of scattered light.

Cavity ring-down (CRD) spectroscopy is a direct absorption technique, which can be performed with pulsed or continuous light sources and has a significantly higher sensitivity than obtainable in conventional absorption spectroscopy. The CRD technique is based upon the measurement of the rate of absorption rather than the magnitude of absorption of a light pulse confined in a closed optical cavity with a high Q factor. The advantage over normal absorption spectroscopy results firstly from the intrinsic insensitivity to light source intensity fluctuations and secondly, from the extremely long effective path lengths (many kilometres) that can be realized in stable optical cavities. In a typical CRD experiment a short light pulse is coupled into a stable optical cavity, formed by two highly reflecting plano-concave mirrors. The fraction of light entering the cavity on one side 'rings' back and forth many times between the two mirrors. The time behaviour of the light intensity inside the cavity can be monitored by the small fraction of light that is transmitted through the other mirror. If the only loss factor in the cavity is the reflectivity loss of the mirrors, the light intensity inside the cavity decays exponentially in time. If there is additional loss inside the cavity due to the presence of absorbing and light scattering species, the light intensity inside the cavity will still decay exponentially in time provided the absorption follows Beer's law [79]. The CRD effectiveness falls off very quickly in a kerbside environment as the mirrors get coated with PM10, also the laser needs to be tuned for each absorber.

Differential optical absorption spectroscopy utilises a broadband source (typically UV) over an open path in the atmosphere. The measurement is made from the differential (small changes) absorption signal. This technique was pioneered in the 1970's by Platt *et al.* [33]. This technique is further detailed in chapter 2 of this text.

Table 1.4.1 shows a summary of the standard methods of spectroscopic trace gas detection.

Detection Scheme	Main Features
Differential optical absorption spectroscopy (DOAS)	Broadband source, separated from receiver by open path in the atmosphere; dual-wavelength operation; integrated response; mainly applied to UV (NO_2, SO_2, O_3)
Laser-induced fluorescence (LIF)	Fluorescence of irradiated species; small cross sections; only suitable for detection of atoms or radicals
Infrared spectroscopy (incl, FTIR, matrix isolation, long path)	Absorption in fundamental IR range; large molecular cross sections; molecule specific
Raman Scattering	Inelastic interaction between photons and irradiated molecules; small cross sections; selectivity by frequency shift; only one wavelength needed; better for short wavelengths
Long-path absorption spectroscopy with lasers incl cavity ring-down spectroscopy (CRD)	Long single-pass or multipass cells; single or double-ended schemes in open atmosphere; integrated response; UV-vis-IR range
Light detection and ranging (lidar)/differential absorption lidar (DIAL)	Atmospheric backscattering of laser pulses; three dimensional profiling of pollutants; complex; mainly applied to UV and vis (SO_2, NO_2, O_3)
Photoacoustic spectroscopy (PAS) with lasers	Measurement of absorbed energy in cell; short path length; mainly applied to IR; selective monitoring of numerous pollutants

Table 1.4.1: Summary table of spectroscopic techniques for trace gas detection [34]

Clearly the most important aspect when considering pollutant species regarding spectroscopic techniques for monitoring air pollution are the spectral regions within which they absorb. This information is summarised for the UV in the following figure 1.4.1. An equivalent figure for the IR would be considerably more complex due to the fact that trace gas UV absorption is dominated by vibronic states while IR absorption is dominated by the considerably more numerous rotational states associated with each mode of vibration. For the purposes of this work it was decided to build the first GUSTO unit operating in the UV.

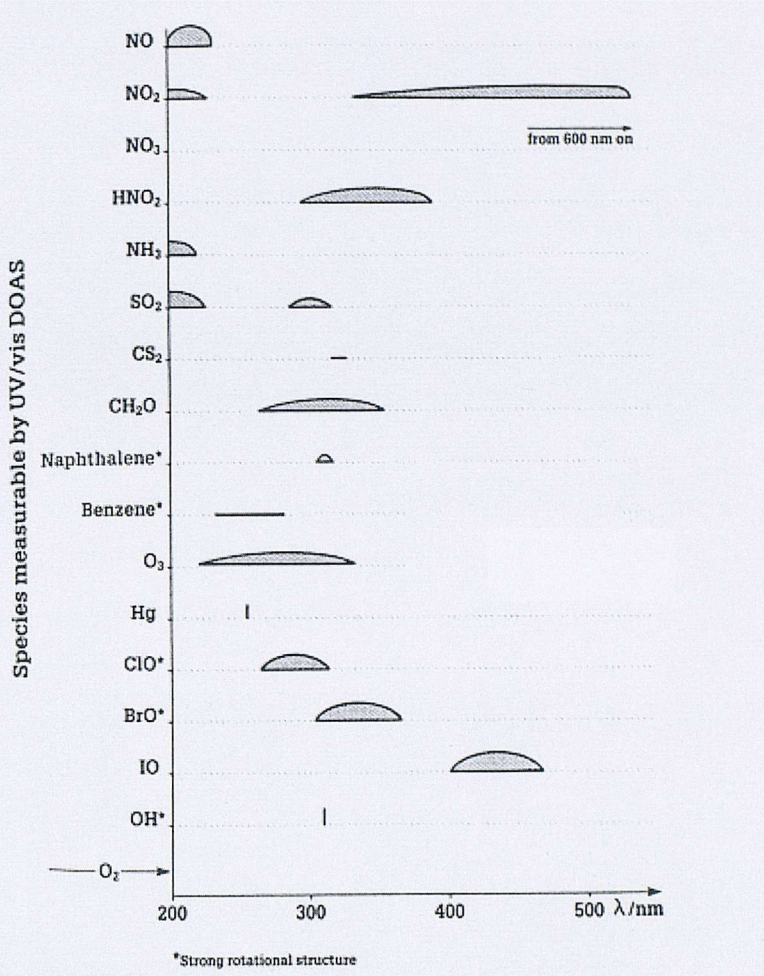


Figure 1.4.1: Summary of pollutant absorber ranges from 200 to 500 nm [34]

Monitoring networks require many instruments, which necessitates the unit cost being relatively low. This precludes the use of lasers as there would need to be multiple units that emit in the UV or a single scanning unit, both of which are (at the time of writing) prohibitively expensive. A broadband UV based source operating over an open path can make *in situ* measurements of NOx, NH_3 and SO_2 simultaneously. This form of system, commonly referred to as DOAS (see discussion in chapter 4) has been applied to urban pollution monitoring in the past, e.g. [41], [42].

1.4.3 Current data

The systems mentioned above that use open path *in situ* measurements on urban air pollution have certain things in common; *in situ* data are never taken on timescales of less than 5 minutes, and simultaneous measurements from the same data acquisition hardware (same source) are only possible with a broadband source (ie. DOAS). Path lengths for such systems are defined [43] as being 100m to km. In order to gain useful new information, one must implement some or all of the following concepts

- Fast data recovery (of order seconds) would allow time changing effects to be characterized. This is useful considering the non-linear response of human biology to air pollution [3].

- Short path length (of order 10m or less) would allow localized effects to be characterized, for example the roadside ambient levels that would affect small groups of people as opposed to the average for everyone on a given road.

- Multiple species, combination of UV and IR would allow all the trace gas targets imposed through legislation to be monitored by a single instrument.

- Real time functionality would allow systems of this type to be integrated into the existing monitoring networks, or to form their own monitoring networks, thereby helping to implement legislation.

1.5 GUSTO

Given the above discussion of current data it was decided that there was a place to develop and apply a new instrument for the purposes of urban air pollution monitoring. Similar instruments in development have been covered in other texts and thus detailed discussion will not be replicated here, however the need for an instrument of this type and a brief summary of the key concepts is outlined below.

1.5.1 Similar Instruments

There is only one similar instrument currently available for use within the urban air pollution monitoring community. This device has until recently been made by Siemens and is called the UV-falcon®. As of the first quarter of 2003 this unit has been discontinued by Siemens, along with the rest of their pollution monitoring programme. The UV-falcon is an open path differential absorption spectrometer based system that operates specifically in the environment of site monitoring (at power plants etc.). The primary differences between the concepts behind the UV-falcon and GUSTO are three fold. Firstly, the UV-falcon uses a Fourier Transform UV spectrometer (200-270nm) which means that the noise on any given measurement will be a function of the noise present on all elements in the 1024 element diode array. The GUSTO approach uses a non-Fourier Transform system and thus suffers from lower susceptibility to this source of noise. Secondly, the unit is designed to operate over path lengths of around 150m as compared to GUSTO where the path length is 15m. The unit can go down to 10m path length, however the sensitivity for any given species is only 50ppb even at 100m and is thus no longer comparable to the GUSTO concept. Thirdly, the unit takes data with an integration time of 5mins (sensitivity 50ppb), with an absolute minimum of 5s. This differs from GUSTO which is intended for taking data with an average integration time of 2s.

1.5.2 The Need for GUSTO

Data for high throughput (i.e. short integration time together with simultaneous multiple species) real-time system with a large dynamic range have not yet been obtained. Some form of DOAS system would theoretically be applicable to this quest. Low integration times necessitate the use of Photo Diode Array (PDA) based data acquisition hardware, thus fast, short open path and *in situ* measurements would be possible. The key elements to a DOAS system that have not been applied by similar instruments are first the short path length and secondly the high throughput (less than 100m and 500s

respectively). Thus to use such a system over 10m (roadside) and take data (single or multiple species) on a second by second basis, would be both useful and new. Given the number of pollutant species from figure 1.4.1 that absorb in the UV as opposed to IR it was decided to make a UV system. A DUVAS system termed Generic UV Sensors Technologies and Observations (GUSTO) was implemented. The author's main contributions have been on the hardware calibration/characterisation, data acquisition (both hardware and software), retrieval methods, testing and first use.

1.5.3 The Aims of GUSTO

The short term aim of the work detailed in this text was the production and first use of an inexpensive, robust short path, high throughput DOAS instrument operating in the UV capable of real time concentration measurements of trace gas/gases, with an accuracy of better than ± 10ppb for SO_2.

The long term objectives for the GUSTO project is the expansion of the concept to networking high throughput urban air pollution monitoring systems. Ultimately, GUSTO type systems could be used to monitor an entire city if large enough numbers of cheap and robust systems could become available to local authorities. This would enable real time and continuous monitoring of pollutant levels over short timescales thus providing a picture much less quantised than is available with the fragmented systems currently in use. With the addition of IR capability to complement the UV, every unit deployed would be capable of measuring around 13 trace gases including the ones already outlined in the species section of this text. Each unit would both measure the original data, do preliminary analysis to retrieve concentrations and transmit all or part of this to a central unit which would act as the node from which data could enter existing information systems (e.g. the world wide web, EIONET[15] and any other interested parties). See figure 1.5.1 below.

[15] European Environment Information and Observation Network [44]

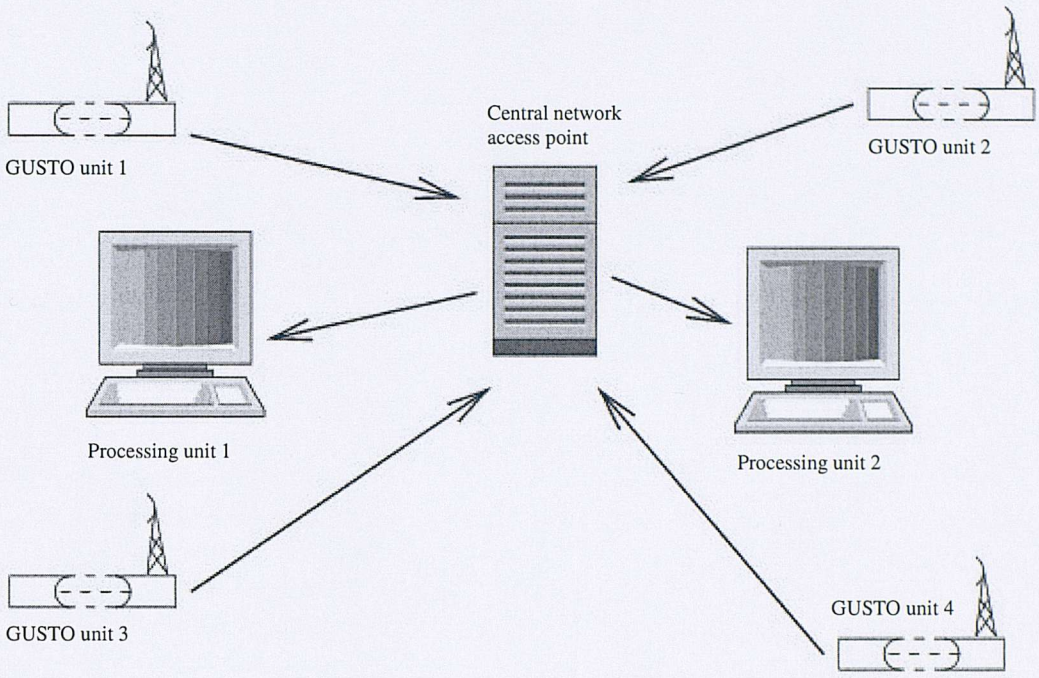


Figure 1.5.1: Networked GUSTO units forming part of a high throughput, real time, urban air pollution monitoring system [4]

This type of networking would allow the data to be processed in many ways not possible with a set of unlinked systems; for example it could allow forecasting of pollution levels [46].

Variations on the concept of the GUSTO instrument, especially if deployed across a road, would allow excessively polluting vehicles to be monitored. Again this requires the addition of complementary IR functionality, by measuring the NO and CO levels to give the NO/CO ratio which is a prime indicator of engine efficiency [43].

1.6 Summary

This concludes the introduction to urban air pollution; the background and concept of the instrument have been briefly outlined. The following chapters deal with the theory that underpins the instrument, its function, application, calibration and trace gas measurements obtained by its use.

Chapter 2

Theory

2 Introduction

The atmospheric concentrations of certain[16] trace gases can be determined using differential optical absorption spectroscopy (DOAS). This method makes use of the characteristic narrow band absorption of the gas under study in a given spectral range (here we use part of the UV, 200-260nm). DOAS was pioneered by Platt *et al.* [33]; the basic idea of this process is as outlined in Section 2.2.

2.1 Differential absorbance

The Beer-Lambert law describes the absorption over a path length x [m] of photons by a gas with number density n [m^{-3}] and absorption cross-section $\sigma(\lambda)$ [m^2] and is usually written as follows [47].

[16] Most commonly NO_2, SO_2 and O_3 but also applicable to NH_3, NO, NO_3, OH, CH_2O, benzene, toluene and xylenes [28] see also figure 1.4.1

$$I(\lambda) = I_0(\lambda)F(\lambda)\exp[-\sigma(\lambda)nx - \alpha(\lambda)nx] \tag{2.1.1}$$

where
$I(\lambda) = Measured\ Intensity\ (Wm^{-2})$,
$I_0(\lambda) = Lamp\ Intensity\ (Wm^{-2})$,
$F(\lambda) = Wavelength\ dependence\ of\ instrument\ (unitless)$,
$\sigma(\lambda) = Absorption\ cross\ section\ (m^2)$,
$n = Number\ density\ (m^{-3})$,
$x = Pathlength\ (m)$,
$\alpha = Scattering\ cross\ section\ (m^2)$.

However, with the GUSTO configuration, we do not know the lamp intensity. To determine this would mean removing all the air from the path, or providing a sealed reference channel.

If we consider the absorption to be made up of broad (σ_B) and narrow (σ_N) contributions to the total cross section, denoted by subscript N on σ and α features [47], then we may write,

$$\sigma = \sigma_B + \sigma_N \tag{2.1.2}$$
$$\alpha = \alpha_B + \alpha_N .$$

Thus if we fit I with a polynomial I_P such that the broad features are accurately reproduced then $I_B = I_P$ and we may write, assuming that I_0 and F vary slowly with λ,

$$I_B = I_P = I_0 F \exp[-\sigma_B nx - \alpha_B nx] \tag{2.1.3}$$

hence

$$\frac{I}{I_P} = \frac{\exp[-\sigma_N nx - \alpha_N nx + \sigma_B nx + \alpha_B nx]}{\exp[\sigma_B nx + \sigma_B nx]}, \quad and \tag{2.1.4}$$

$$\ln\left[\frac{I}{I_P}\right] = -(\sigma_N nx + \alpha_N nx). \tag{2.1.5}$$

Equation 2.1.5 holds only for a single specie, while for many species we need to sum for σ_N. In the limit of low particulate density we may assume that α_N is very small compared to σ_N so

$$\ln\left[\frac{I}{I_P}\right] = -x\sum \sigma_{Ni} n_i = dA \tag{2.1.6}$$

The quantity dA in equation 6 is known as the differential absorbance. Thus from equation 2.1.6 we may find n given the LHS and x.

2.2 Data retrieval method, DOAS

The method used to obtain the differential absorbance as defined in the above equation 2.1.6 from the actual data taken from a system that employs the DOAS concept consists of the following:

One measures the intensity of light that has passed through the given path length as defined by the above equation 2.1.1. Then a fit to these intensity data is generated as described by equation 2.1.3 and the differential absorbance is then calculated as given by equation 2.1.6. From this, given that the corresponding absorption cross-section may be looked up from an appropriate reference and the path length is known, the number density may be calculated.

Data taken with a real instrument will have features specific to the instrument and these must also be accounted for. The method of data acquisition will also affect the form of the data taken and any systematic errors or discrepancies present in these will too require attention. The types of signal that GUSTO will be used to study are summarised in the following figure 2.2.1 [48].

The measured signal is labelled I, dA is the differential absorption spectrum while P is the pixel sensitivity and S is the photoelectron shot noise. I_B is the background signal (fit to I). Thus one measures I directly and retrieves dA by applying equation 2.1.6. The typical signal to noise ratio for the DOAS system from which this plot is an example is 100:1 for the pixel sensitivity and 200:1 for the shot noise [48].

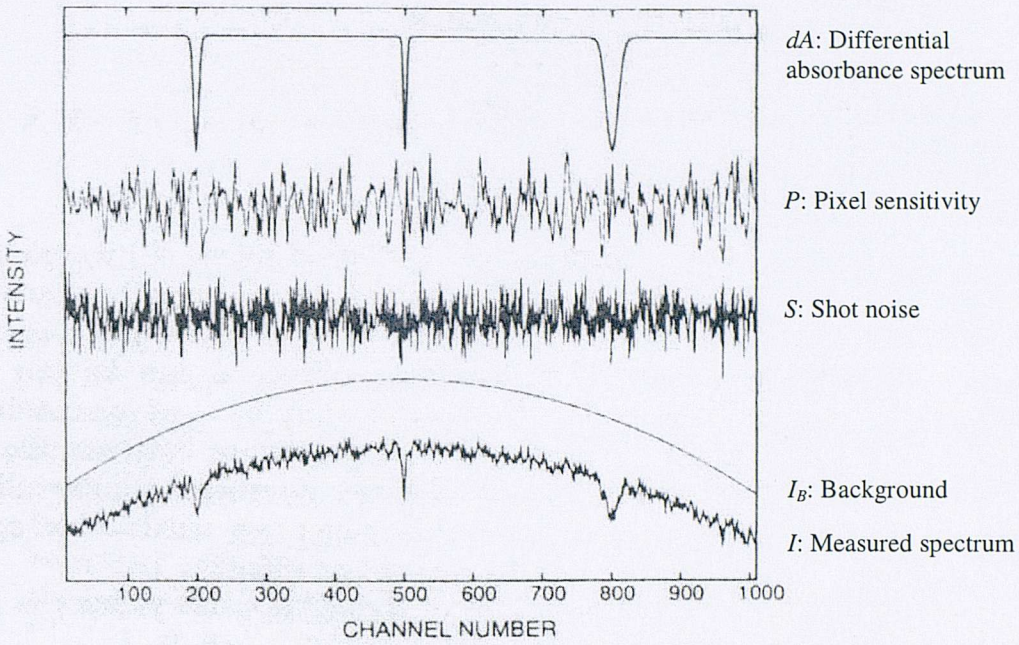


Figure 2.2.1: Breakdown of DOAS signal into its major components [48] Arbitrary scale is used.

Figure 2.2.2 shows a summary of the steps in the DOAS method, with each being described in more detail below.

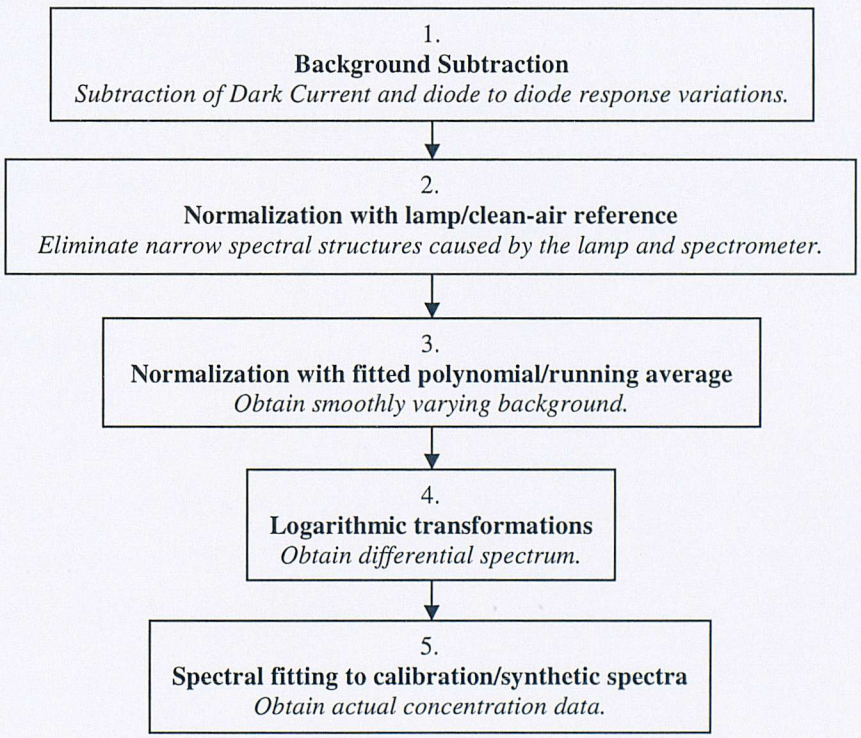


Figure 2.2.2:GUSTO data analysis method

2.2.1 Background subtraction

The subtraction of the background, the first step outlined in figure 2.2.2, is the removal of the pixel sensitivity and the photoelectron shot noise together with the smoothly varying background from the original measured signal. The photoelectron shot noise (S) is the noise present due to the finite number of photons interacting with any one diode and is dependent on the integration time used for the signals under study as well as the actual properties of the photodiode array. The pixel sensitivity (P) is a property of the diode array and originates from the manufacturing techniques employed in its fabrication. A numerical treatment of both of these noise sources is given in chapter 4. The photoelectron shot noise is by definition random as one cannot measure how many photons actually interact with any one diode. By using a suitably long integration time the variations between successive diodes become a progressively lower fraction of the total signal and thus the signal to noise ratio (SNR) increases. Further, if such random variations were to dominate any given signal present in the data set then multiple data sets can be averaged so as to reduce the noise by the square root of the number of data sets averaged. A numerical treatment of this averaging process is presented in chapter 4.

The pixel sensitivity can be determined by putting a known intensity of light on each diode in turn, or more usually by putting an unknown intensity on a given diode and then transferring it to each other in turn, thus giving a relative comparison of each diode with respect to one specific diode. It is also possible to shine the light used in the system through the optics and on to the photodiode array at a number of intensities over a long period of time. The resulting systematic differences in the light intensity from diode to diode will be partly due to pixel response variations and partly due to the lamp/optics combination; both must be calibrated and removed at the same time. Both of these approaches were tried with the GUSTO system and the results and conclusions are presented in the following chapter.

2.2.2 Normalisation with lamp/clean air reference

The second step from figure 2.2.2 is the normalisation of the signal obtained with some form of clean/known reference. The actual line shape of the instrument must clearly be measured and the pixel response variations discussed above need also to be considered, as do any properties of the optics. Extensive testing with known inputs and clean air (no pollutants present) must clearly be done and the effect of trace absorbers that are always present in air (e.g. oxygen and water) must also be considered. The results of this analysis are presented in detail in chapter 4.

At present the existing GUSTO system relies on good optics and calibration to minimize experimental errors. Other well-documented techniques exist within the DOAS community. For example one can use multiple scans over different wavelength ranges (Multichannel Scanning Technique [48]) to compensate for the pixel sensitivity pattern (*P* from figure 2.2.1). This technique utilises the ability to change the grating position within the monochromator and hence take a number (usually three) data sets all with approximately the same input light. Thus one has three data sets each containing two independent signals (the intensity signal one wants to measure and the diode response variations). One can also partially compensate for the noise generated by the unavoidable changes in illumination taken with the angular dependence of the response of the PDA, by using a quartz fibre bundle in place of the lens between the focusing lens before the spectrometer slit and the slit itself (see later discussion on the GUSTO hardware)[49].

2.3 Fitting the background

The most complex part of the DUVAS method, is the *normalization with fitted polynomial/running average* (third stage). A polynomial needs to be fitted to the entire scan. Best results were obtained with a 4^{th} order polynomial, while other work uses a polynomial of up to 6^{th} order [33]. It was first decided to build the algorithm around

smoothing implemented with running averages. However the resulting smoothed scan was not smooth enough to yield usable data [33]. This was determined by the comparison of 3 and 5 point running averages, outlined in section 2.3.2, where the respective differential scans differed by amounts comparable to the features present in each scan. Thus a more accurate and therefore more complex fitting method was required.

2.3.1 Savitzky-Golay Filter

The choice of smoothing filter for application to differential absorption spectroscopy is some form of polynomial least squares fit [33], and the exact form decided upon was the Savitzky-Golay Smoothing Filter [50]. In a Savitzky-Golay Smoothing Filter, each original data point is fitted by setting its value to that of an m^{th} order polynomial fitted in a window of size n points around the point to be fitted. For the purposes of this program m was set to 4 [33] while best results were found by trial and error to be obtained with n set between around 31 and 101. This may be seen by looking at data such as that shown in chapter 5 for SO_2 (e.g figures 5.1.5 page 133 and 5.1.14 page 142) where the broad features (figure 5.1.5) are of order 100 diodes in size while the fine features that are not to be fitted (figure 5.1.14) are of order 10 diodes in size. Thus in a data set comprising N equally spaced data points, N m^{th} order fitted polynomials are required. The process of least squares fitting involves only a linear matrix inversion and the coefficients of a fitted polynomial are themselves linear in the values of the data. Thus we may find all the coefficients in advance by fitting for synthetic data comprising of all zeros and a 1, then fit to the real data by taking linear combinations. This makes the method surprisingly fast as the Savitzky-Golay coefficients for any given m and n need only be calculated once and thus they may be read from a file and simply be applied as each scan is being analyzed.

In a general digital filter [51], including a smoothing filter, each data value f_i is replaced by a linear combination g_i of itself and some number of nearby neighbours (n_L are left nearby neighbours and n_R are right nearby neighbours), where i runs over all data points. This is described by the following equation 2.3.1

$$g_i = \sum_{n=-n_L}^{n_R} c_n f_{i+n}, \tag{2.3.1}$$

where c_n is the coefficient used to obtain a contribution to the sum from the data point f_{i+n}. One knows f_i and is free to choose the values for n_L and n_R, hence the filter is applied by finding c_n and applying them to the data set.

In the case of the Savitzky-Golay filter we want to fit a polynomial of degree m in i, that is we want to fit

$$a_0 + a_1 i + \ldots + a_M i^m \quad \textit{to the values} \quad f_{-n_L}, \ldots, f_{n_R}$$

Thus g_0 will be the value of that polynomial at $i = 0$, which is just a_0.

The matrix formulation of this is straightforward with the design matrix shown below in equation 2.3.2 and the normal equations being as per equation 2.3.3.

$$A_{ij} = i^j \qquad i = -n_L, \ldots, n_R \qquad j = 0, \ldots, m \tag{2.3.2}$$

$$\left(A^T \cdot A\right) \cdot a = A^T . f \tag{2.3.3}$$

where
a = *unknowns* c_n *from above equation* 2.3.1
f = *known data* f_i *from above equation* 2.3.1

Given that we know A and f then equation 2.3.3 above may be solved for a by any of the usual methods. In this case it was decided to use the so called Lower-Upper (LU) decomposition (described below), as it is both accurate and stable.

LU decomposition [52] is as follows: suppose we wish to solve the following matrix equation 2.3.4, where T and b are known. We are in fact rewriting equation 2.3.3 such that $(A^T \cdot A)=T$ and $A^T \cdot f=b$.

$$T \cdot a = b \tag{2.3.4}$$

Then if we find L and U such that equation 2.3.5 holds:

$$L \cdot U = T, \tag{2.3.5}$$

where L is lower triangular (has non-zero elements only on the diagonal and below) while U is upper triangular (has non-zero elements only on the diagonal and above). Then we may write,

$$T \cdot a = (L \cdot U) \cdot a = L \cdot (U \cdot a) = b.$$

Hence solving equation 2.3.4 is the same as solving the following two equations:

$$L \cdot y = b \tag{2.3.6}$$
$$U \cdot a = y. \tag{2.3.7}$$

This is straightforward as both L and U are by definition triangular. Thus equation 2.3.6 may be solved by simple back substitution [52] while equation 2.3.7 may be solved by simple forward substitution [52].

To solve equation 2.3.4, one first solves equation 2.3.6 by back substitution, where L and b are known and y is found. Then one solves equation 2.3.7 by forward substitution, where U and now y are known thus finding x, which is itself the solution to the original equation 2.3.4.

The most difficult part of this procedure is in going from equation 2.3.4 to equation 2.3.5. This is accomplished in our case by application of the standard Crout's method with partial (implicit) pivoting [52], which, in brief, solves for L and U in equation 2.3.5 by arranging the equations that make up the matrices in a certain order. Pivoting refers to the fact that one of the two iterative steps in Crout's method involves a division by one of the matrix elements and one needs to be sure that this division is mathematically stable. Implicit refers to the fact that the matrix equations are scaled so as to make their maximum element equal to unity.

Hence the coefficients of the Savitzky-Golay filter may be found, stored and applied to each successive scan, thus yielding the required smooth background. The chosen

implementation of this is to recognize that the filter coefficients are by definition symmetric about the centre point, thus they may be stored in a compressed form (one half only), read and then decompressed by the retrieval software, ready to be applied to the data.

2.3.2 Averaging techniques

As mentioned above we originally attempted to fit a smooth background by implementing a running average. The following figures 2.3.1 and 2.3.2 show how this running average fit compares to the polynomial fit that was finally decided upon.

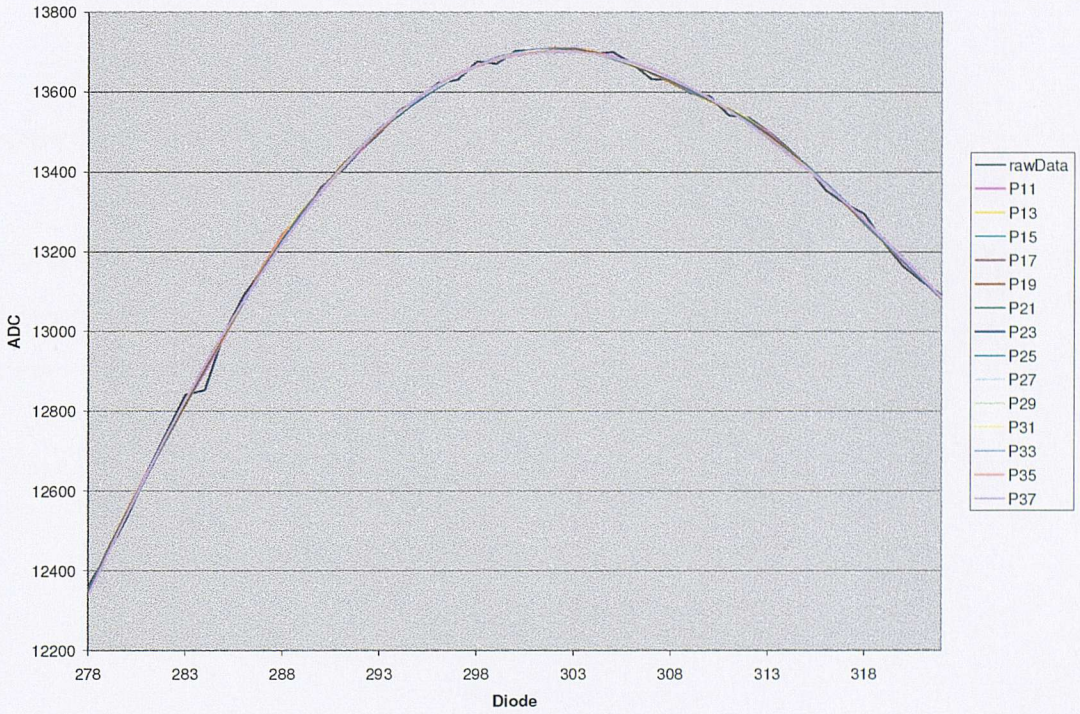


Figure 2.3.1: Comparison of polynomial fits to a GUSTO like peak, shown are fits for window widths from 11 to 37 points.

Figure 2.3.1 shows the polynomial fits (for windows of width 11 through to 23 points) as may be seen all the fits roughly follow the shape of the raw data. The raw data from which this graph is derived are shown in figure 4.3.1.

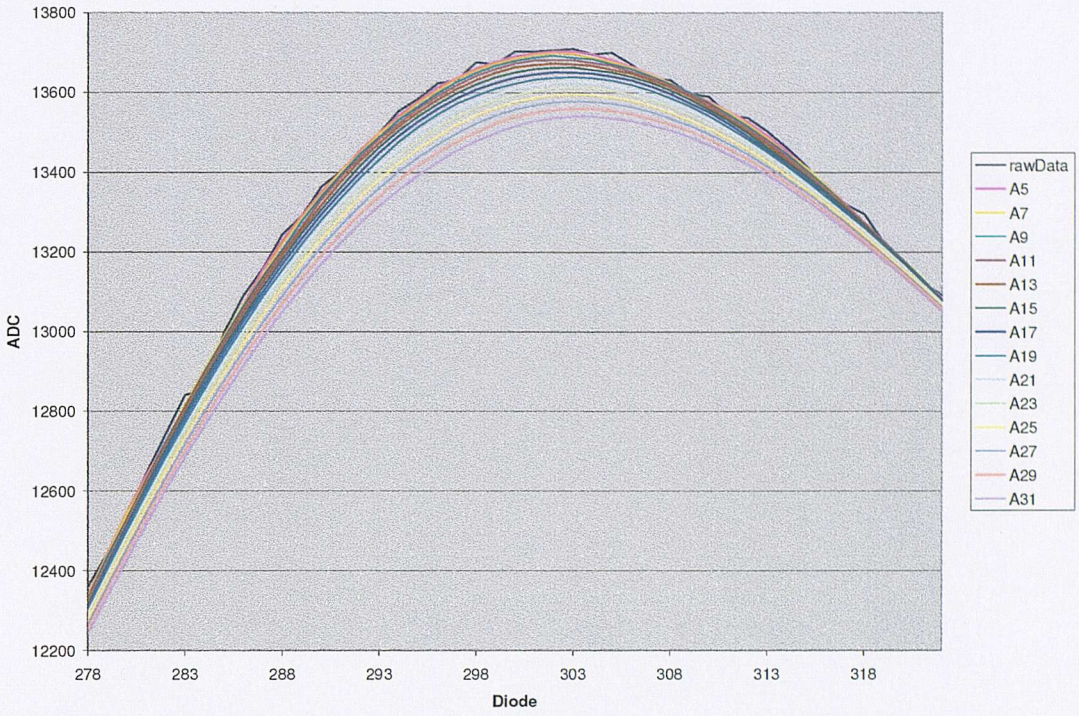


Figure 2.3.2: Comparison of running average fits to a GUSTO like peak, shown are fits for window widths from 5 to 31 points.

Figure 2.3.2 is the equivalent of figure 2.3.1 but for the running average fit (window regions from 5 to 31 points). The size of peaks in GUSTO like signals are of order 8 points (see figure 5.1.14 page 142) and thus running averages fits will need to be at least 8 points wide. Hence by comparing figures 2.3.1 with 2.3.2 it may be seen that a polynomial fitting method is the logical choice for a GUSTO like system.

2.3.3 Detailed look at peaks

As well as looking at data from 1999 it was decided to examine closely the effect of fitting to a single peak. To this end the fitting algorithm was run on single synthetic triangular peaks of differing sizes and depths (average width of GUSTO like peak is 8 points, see figure 5.1.14 page 142). Below is an example; figure 2.3.3 is a 7-point peak fitted with a number of running average fits from 5 to 31 points: figure 2.3.4 is the same 7-point fitted with a polynomial fit for 11 to 37 point windows. Quantitative analysis of this underestimate of the peak depth is outlined in the following section.

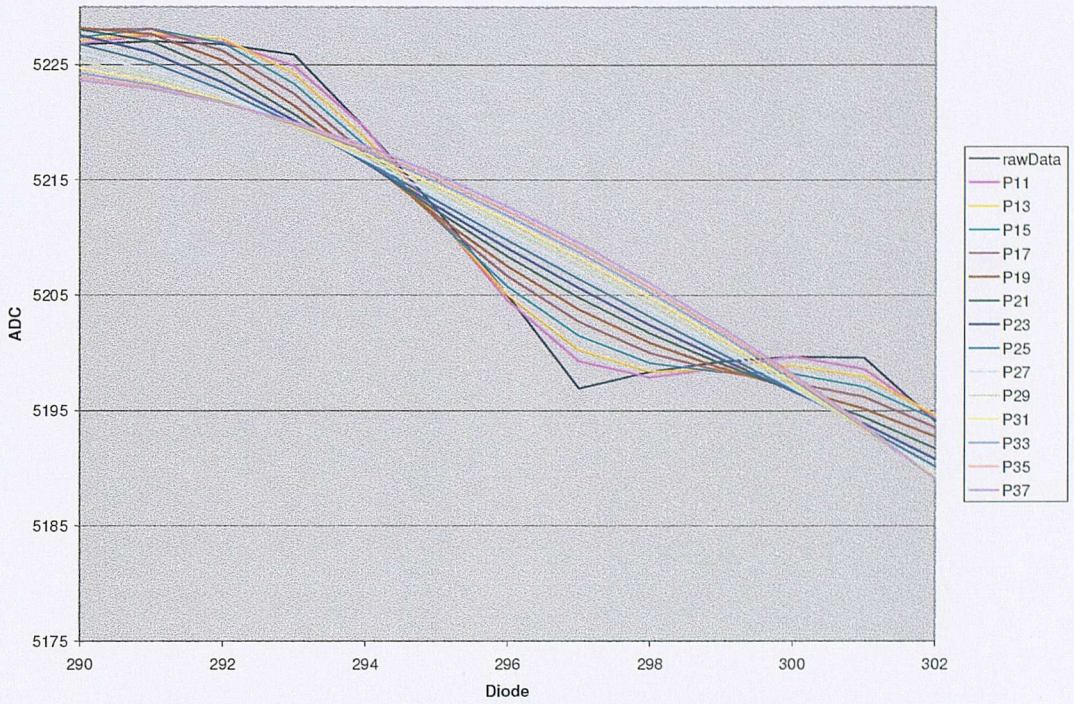


Figure 2.3.3: Seven-point peak fitted with a range of polynomial fits.

The raw data curve shown in both figures 2.3.3 and 2.3 4 is clearly on a background that is underestimated by both the polynomial and running average fits. This underestimate is greater for the running average fits (due to the fact that the peak is not in a flat background but is on a gross structure peak similar to those seen in figure 5.1.14 page 142).

Consider figure 2.3.5, which shows a signal (y-axis scaled to give maximum signal height of 1) with a peak introduced (depth d). If there are N points in the window used for the fit and the peak has width n points, and assuming that each point is given equal weight by the fit, n points will contribute approximately d/2 less to the background than the remaining N-n. Thus the error in the fit with respect to the background (as a fraction of the peak depth) is given by n/(2N). Hence with a 3-point peak of any depth fitted by a 77-point window the background will be consistently off by around 2 percent (3/154).

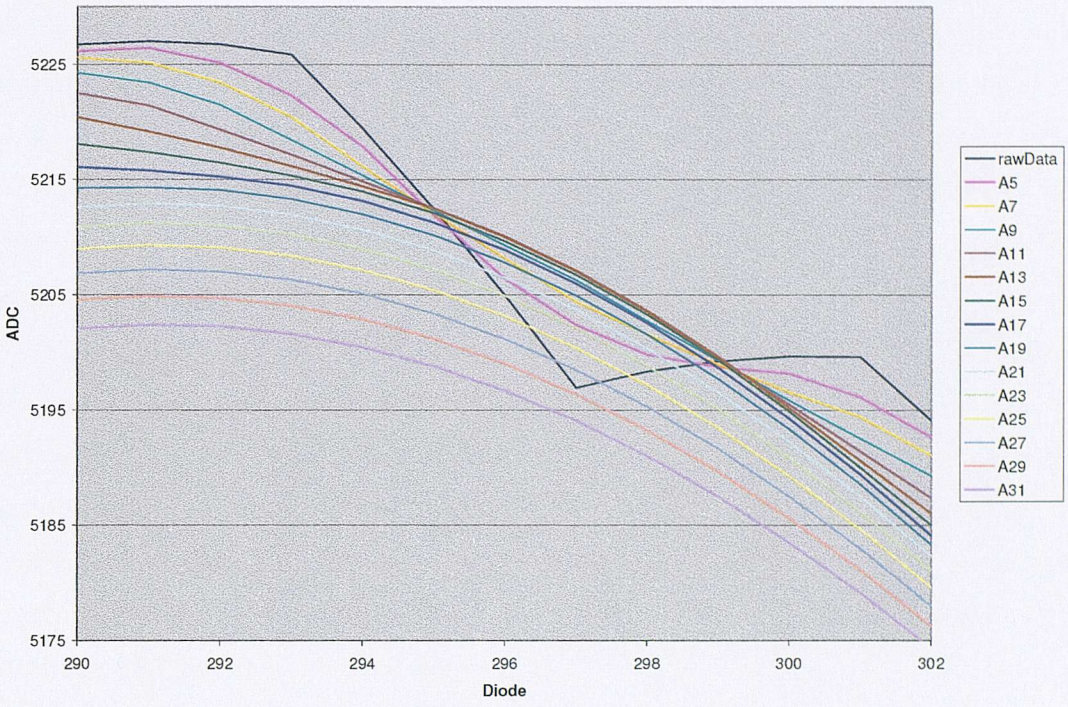


Figure 2.3.4: Seven-point peak fitted with a range of running average fits

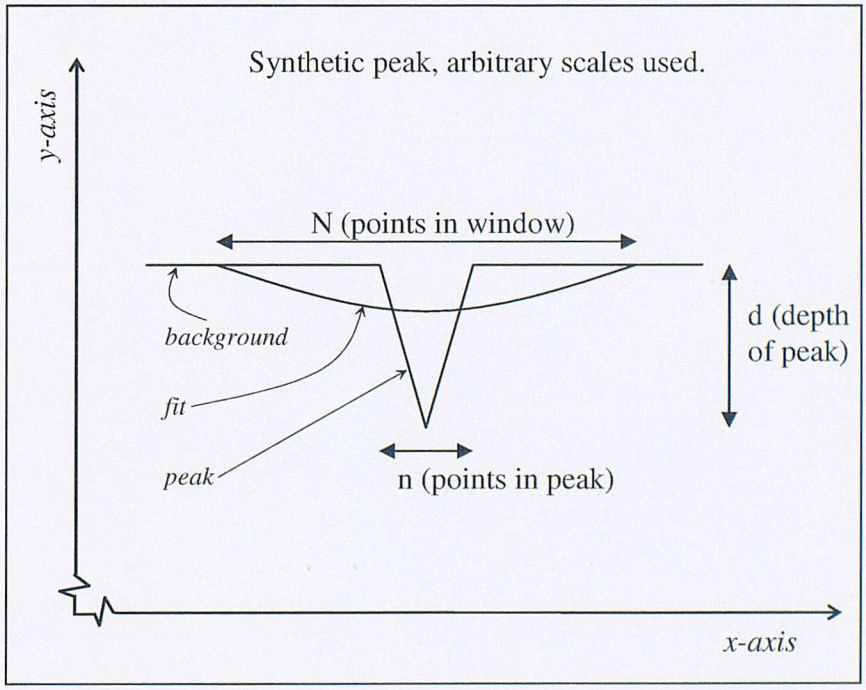


Figure 2.3.5: General synthetic peak in arbitrary (flat) background

The above discussion of the error in fitting to a background containing a peak only applies if the absorbing pollutant under study has discrete peaks on an otherwise smooth background. If it absorbs in a spectrally unresolved manner with one peak partly overlapping with another then any fit would not be biased in the same manner and the error in the fit would therefore be lower. Figure 2.3.6 shows the absorption cross sections of the species within the GUSTO range. The backgrounds for gases such as ozone or sulphur dioxide would be fitted well due to the spectrally unresolved nature of the cross sections, while gases such as nitrogen oxide and benzene would be fitted less well as the absorption peaks are more distinct. The detail shown is approximately correct for the resolution of the GUSTO instrument (0.2nm) thus the signals will not be appreciably smoothed beyond what can already be seen in the figure. Figures 2.3.7 and 2.3.8 are enlarged portions of fig 2.3.6 for nitrogen oxide and benzene respectively, illustrating the detail not shown in figure 2.3.6.

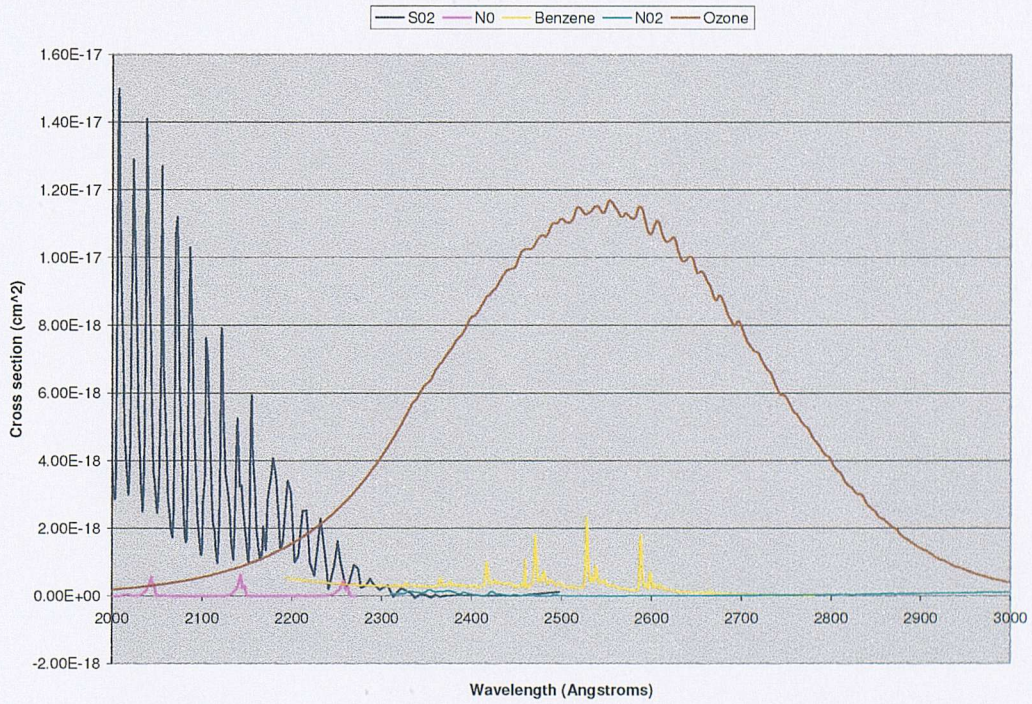


Figure 2.3.6: Absorption cross sections for SO_2, NO, Benzene, NO_2 and O_3 in the GUSTO wavelength range [53], [73].

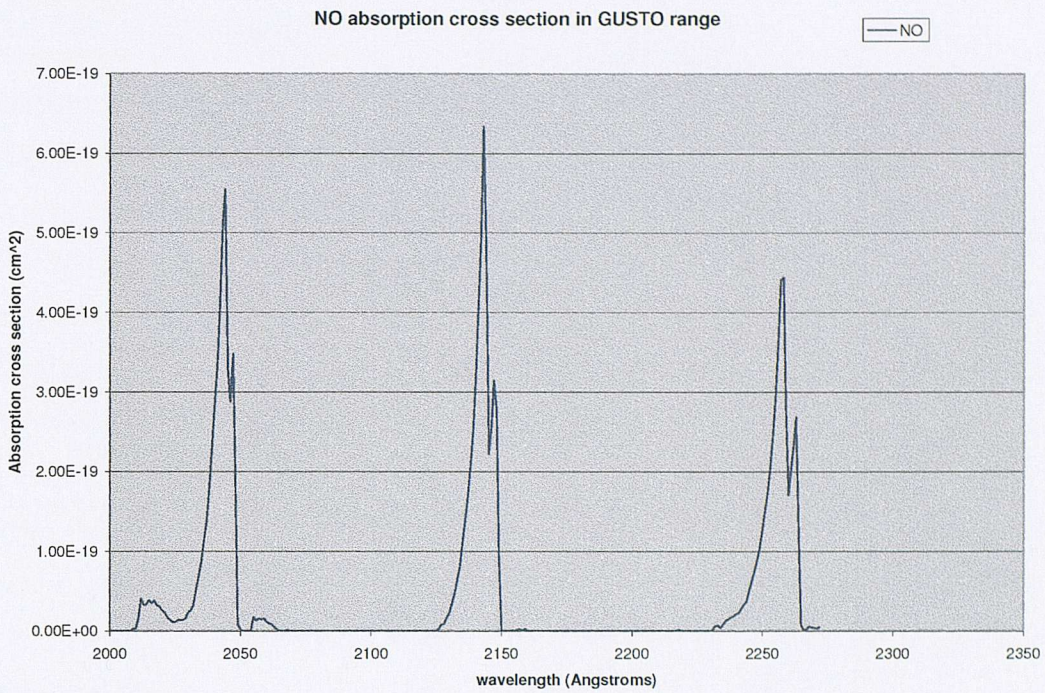


Figure 2.3.7: Enlargement of absorption for cross sections for nitrogen oxide within the GUSTO wavelength range

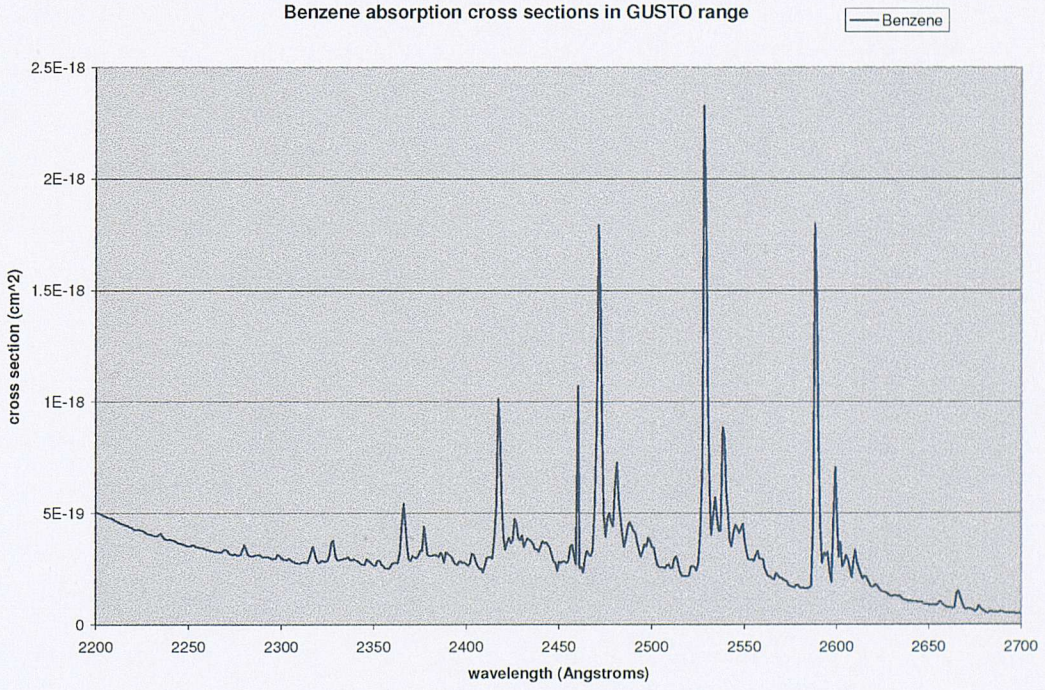


Figure 2.3.8: Enlargement of absorption for cross sections of benzene within the GUSTO wavelength range

2.3.4 End effect correction

When the moving window within which the polynomial is fitted with the Savitzky-Golay smoothing filter hits the end of the spectrum, the method fails. This is because you are then only able to include an asymmetric set of points in the window and thus the curve is no longer being well fitted. Obviously one cannot simply lose the end 37 points from one's data sets (as this would mean valuable data could be lost). The use of a running average fit, as discussed in the previous subsection, is not preferable to the polynomial. Hence it was decided to shrink the moving window as it neared the edge of the scan, thus allowing a polynomial to be fitted right up till the point where a running average would better fit the data. It is not logical to fit a fourth order polynomial to anything other than at least four points and in fact we decided on ten. Thus the end effect corrected algorithm uses a window that is as large as possible without going off the edge of the scan but smaller than 77 points, and a running average for all other points (best results were found with a 7 point average).

Figure 2.3.9 illustrates the application of this to a sample data set (from the outside optical setup taken in June 02), showing both the original data, together with the polynomial fits with and without the end effect correction applied. The intensity of light hitting the low wavelength diodes is reduced as a result of the attenuation of light from the quartz in the system (windows etc.). The middle of the diode array has an oscillatory response to incoming light due to the manufacture of the diodes that make up the array [58]. The light intensity output from the lamp decreases at higher wavelengths (above around 250nm). The photon count on the lower diodes is low (see discussion in chapter 6) while the photon count on the higher diodes is almost 20,000 higher (representing 30% of the full 16 bit range). The difference between the two fits illustrates what happens as the smoothing filter reaches the ends of the spectrum. At lower diodes the filter (without end effect correction) adds zeros to the data set (pads with zeros) in order to utilize the full 35 point window, however the signal is nearing zero and thus the difference between the end effect corrected curve and the non end effect curve is small compared to the higher diodes where the signal is closer to 20,000 ADC counts where the filter again pads with zeros.

While not at the ends of the spectrum, the filters with and without the end effect correction are working on exactly the same numbers and thus they give identical results. The residuals (fit-data) show the diode-to-diode variation (they are in fact differentials, see discussion in chapter 4), however the residual of these differentials illustrates the difference between applying and not applying the end effect correction. This is plotted in figure 2.3.10.

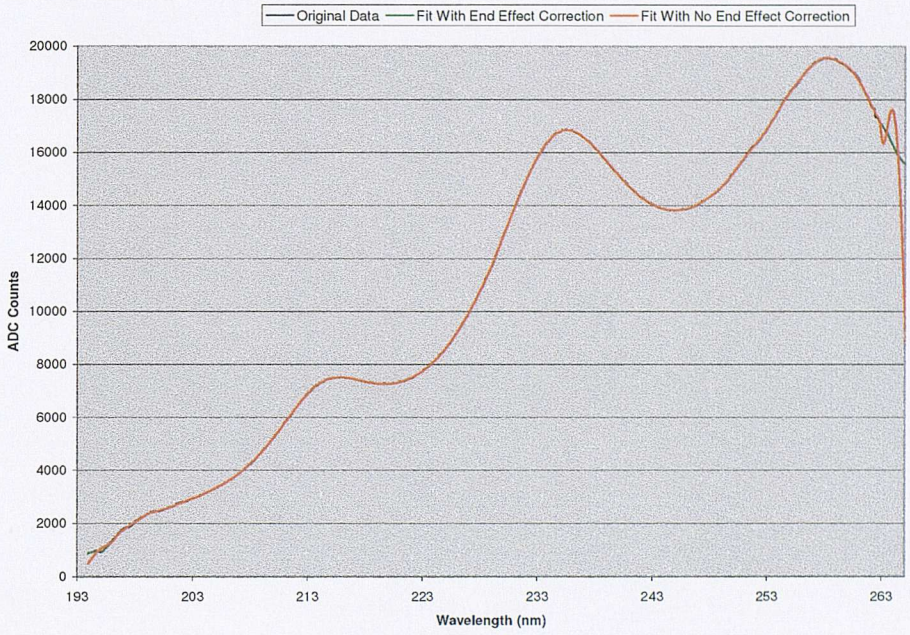


Figure 2.3.9: Implementation of end effect correction routine, shown is original data plus fits with/without end effect correction applied

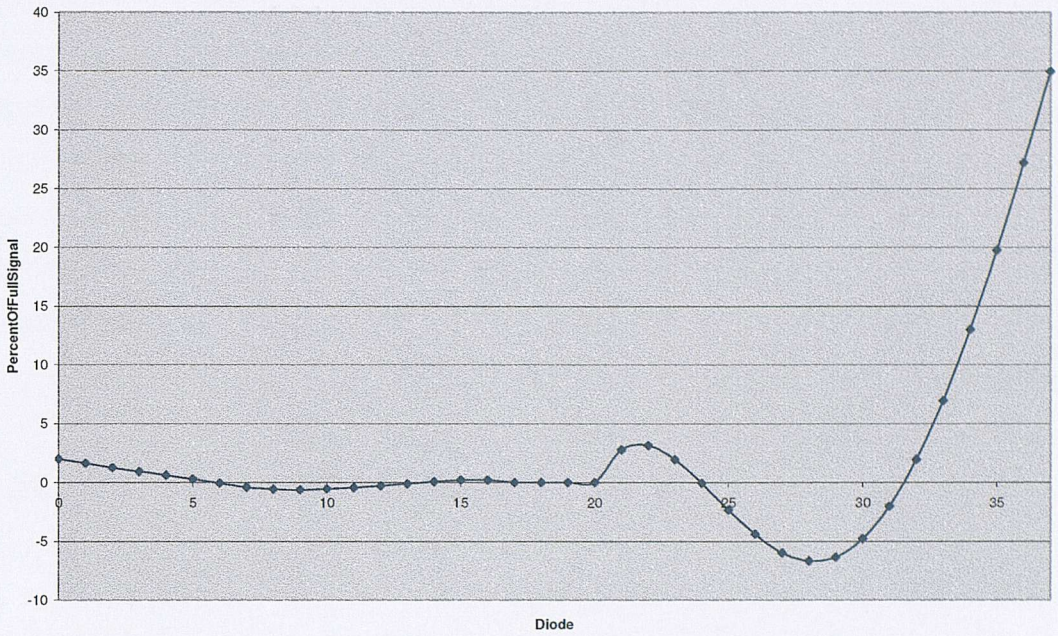


Figure 2.3.10: Residual of differentials between fits with and without end effect correction, y-axis scale is percent of full signal, x-axis scale is diode number, only first and last 16 diodes shown.

The y-axis scale in the above figure 2.3.10 is the magnitude of the residual as a percent of the full data signal from figure 2.3.9 while the x-axis is the diode where only the first and last 16 are shown (all the remaining are zero).

From figure 2.3.10 it may be seen that not applying the end effect correction makes a significant difference to the retrieval from the affected diodes, particularly as the introduced error is of the order of one percent while the signals under study are fractions of a percent. This effect is smooth, however any unwanted shape present in the retrieved differentials must at some point be removed. Application of the end effect correction is the most straightforward way of fitting the background.

2.4 Mask fitting

The final step in the method for retrieving pollution concentrations from measured data is the spectral fitting to calibration/synthetic data. At this point one simply has the differentials which have been measured. If only one pollutant is being investigated then

application of a standard single variable least squares fit will yield both the number density of the pollutant and the error on the fit can be thought of as a measure of the error on this retrieved number density. If multiple pollutants are present then either a multi-variable least squares fit, or application of multiple, single or low, variable least squares fits, in a pre-defined order, must be applied (for example if the concentration of the largest contributor is found first and those data fed back into the fit for the other contributors etc.).

2.4.1 Application to GUSTO data

In the case of the data taken during the course of 1999-2002, SO_2 and benzene were identified as being theoretically measurable. These species absorb at different wavelengths (see fig 2.3.6) hence they may be fitted separately. To illustrate the processes of this type of fit the algorithm was run on the old data from 1999 (ozone released under controlled conditions in the laboratory).

The measured absorption cross sections for the species under study were obtained from a previous study conducted at Imperial College [53] and then converted to the GUSTO resolution (512 element PDA covering 70nm). This must then be convolved with the instrument function to yield what one actually expects to see from the instrument. From there the process of mask fitting can be applied.

2.4.2 Illustration of mask fitting algorithm

As mentioned in the previous section, sample data were taken in 1999 in order to demonstrate the functionality of the first GUSTO prototype. Ozone spectra were fitted to the said data using a new algorithm [54]. Pollutant concentrations were obtained that agreed well with the results previously obtained for the same data set [55]. The actual data set from 1999 consists of some 1500 separate scans (each corresponding to a one second time interval). The size of the errors present in each scan proved to be too great

to allow actual concentrations to be obtained from any less than sets of 10 scan averages. The inferred ozone mixing ratio was 80ppb ± 10 over 15m for both the original analysis and after application of the new algorithm.

2.5 Summary of DOAS/DUVAS method

Application of the DOAS method to short open paths (less than 100m), with a photodiode array instead of a scanning monochromator and using the UV instead of the visible optical region is a substantial departure from the classic DOAS system [33] thus herein it will be referred to as DUVAS (replacing the O for optical with UV for ultra violet). The main points that must be addressed when analyzing data taken from a system using the principles of the DUVAS methods may be summarised as follows (this is a version of the steps shown in figure 2.2.2 but for the specific GUSTO instrument in use as opposed to a generic DOAS unit):

- Raw data must be normalized (output from our system is from 0 to 5V, see chapter 3, however the readin electronics has the range –5 to +5V). Failure to do this yields differentials of the order of 3 times larger than are actually present.

- The pixel-to-pixel error needs to be compensated for using previously taken calibration data.

- The photodiode shot noise is compensated for (if necessary) by averaging successive numbers of scans.

- The background must be found by appropriate fitting, paying careful attention to any systematic errors inherent in the processes.

- The measured differential is then calculated.

- The instrument function is convolved with the known absorption cross sections for the species under study and this is fitted to the measured differentials. The number density of the pollutant can be obtained and the error on the fit is one method of estimating the error in that retrieved number density [72].

Chapter 3

Experimental Hardware

3 Overview

Figure 3.1 shows the optics and basic set-up used for the GUSTO prototypes. Light from the deuterium lamp is focused by lens 1 and turned into a parallel beam by being passed through telescope 1. The distance between telescope 1 and telescope 2 is the open path within which are the pollutants under observation. Telescope 2 focuses the parallel beam radiation and this is then directed through lens 2 onto the spectrometer. The distance between the mirrors is 1m in the lab and around 15m outside. The specification and operation of each component is detailed below.

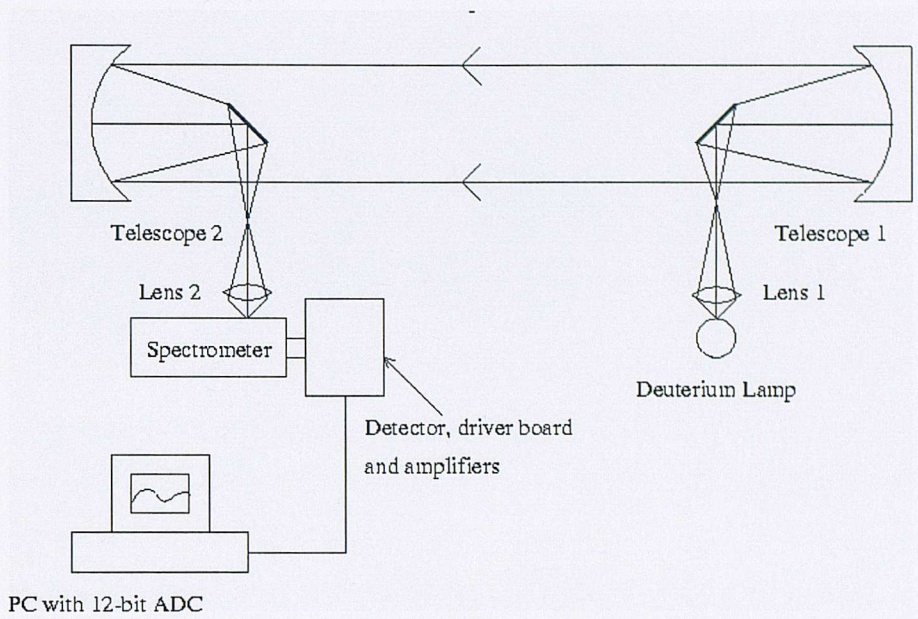


Figure 3.1: [4] Outline of optics and hardware used in the GUSTO prototypes.

3.1 Optical hardware

The optical configuration of the GUSTO system may be divided into five primary sections, these are the lamp with its associated optical components, telescope, windows, spectrometer and photodiode array. A detailed discussion of each subsystem follows.

3.1.1 Lamp and collimating lens

The lamp is from the L2D2 2000 series 30 Watt range of deuterium lamps from Hamamatsu. This is a powerful UV lamp with spectrally featureless and adequate light output from 200 to 270 nm. The exact model chosen is designated L6567 and differs from other lamps in the same range by way of having a point like source with no focusing optics in the lamp bulb to constrict the spot from the arc. The entire unit is housed in a custom-built aluminium housing that allows height adjustment. This housing gets extremely hot during prolonged use thus good airflow is maintained around the unit. Also mounted in the same housing is the collimating lens that collects light from the arc and focuses it onto the imaging lens sitting in a separate holder further forward. Correct alignment of this lens system is essential as the focus between lens 1

and the forty-five degree mirror before telescope 1 in figure 3.1 needs to be exactly at the focal point of the parabolic mirror in telescope 1 in order to obtain a parallel beam within the telescope subsystem. The focal points can be followed through the optical system to check that this alignment is achieved.

3.1.2 Telescope

The key to getting an open path system to work lies in achieving correct optical collimation in the beam that exists along the open path. If this is not achieved then much valuable light will be lost due to beam divergence. The GUSTO system contains a telescope that consists of light reflected off a forty-five degree mirror (diameter 23 mm to beam) onto a parabolic mirror (diameter 150 mm) which is then collimated. This light then travels along the open path and is focused down by the same type of optical configuration (parabolic mirror onto forty-five degree mirror) onto the photodiode array. This does achieve a collimated beam (details of the checks for optical collimation as applied to this type of telescope are outlined in chapter 4), however the forty-five degree mirrors are directly in front of their respective parabolic mirrors. Thus a shadow is cut from the beam, which is the same for each mirror. The collimation of the beam over 15m is not sufficient (even when correct optical collimation is achieved) to retain any of the shadow from the first forty-five degree mirror by the time the beam reaches the second parabolic mirror, thus the shadow by the forty-five degree mirrors is in fact cut from the beam twice. The only way to avoid this is to use angled parabolic mirrors, which are rather more expensive than the standard circular type and were thus not an option for the present work, although it is retained as a strong option for the future. Also the limited dimensions of the box set the distance between the parabolic mirrors and the forty-five degree mirrors and a reasonable amount of light is lost due to the first forty-five degree mirror being overfilled.

3.1.3 Windows

The windows through which the parallel beam passes to get from the lamp to the photodiode array are made of fused silica (transparent to the UV within our range). They have a diameter of 15cm and are around 5mm thick which is thin enough to allow good light throughput however both windows cracked due to mechanical strain a few months after installation outside the main Imperial College buildings. Thus the optical plane of the windows is not uniform and a reasonable amount of dirt is lodged in the cracks (one running the diameter of each window) also dust may well be getting into the boxes through the cracks. Ethanol is used to clean the window surfaces both inside and out regularly and only around 4.3 percent of all light is lost at each surface. The alternative of having either new or no windows would result in either substantially increased cost or more dust particles entering the boxes respectively.

3.1.4 Monochromator

The monochromator used is the Ebert-Fastie CVI CM110SP [57] which is a special form of the Czerny-Turner type shown in figure 3.1.1. This unit has two gratings although only one is actually used in our configuration (1200 lines per mm and blazed at 330 nm with physical size 30mm by 30mm and maximum efficiency 65%). The unit has focal length 110mm with f/3.3 and a maximum resolution of 0.2nm. The slit used in this case is 100μm. This represents the single most expensive piece of hardware in the GUSTO set-up. The mountings are all made in house and apart from ensuring that incident light to the slit is along the optical axis the only difficult part is interfacing with the PDA. The output must be focused onto the surface of the PDA and this is achieved using a custom (in-house) painted aluminium spacer.

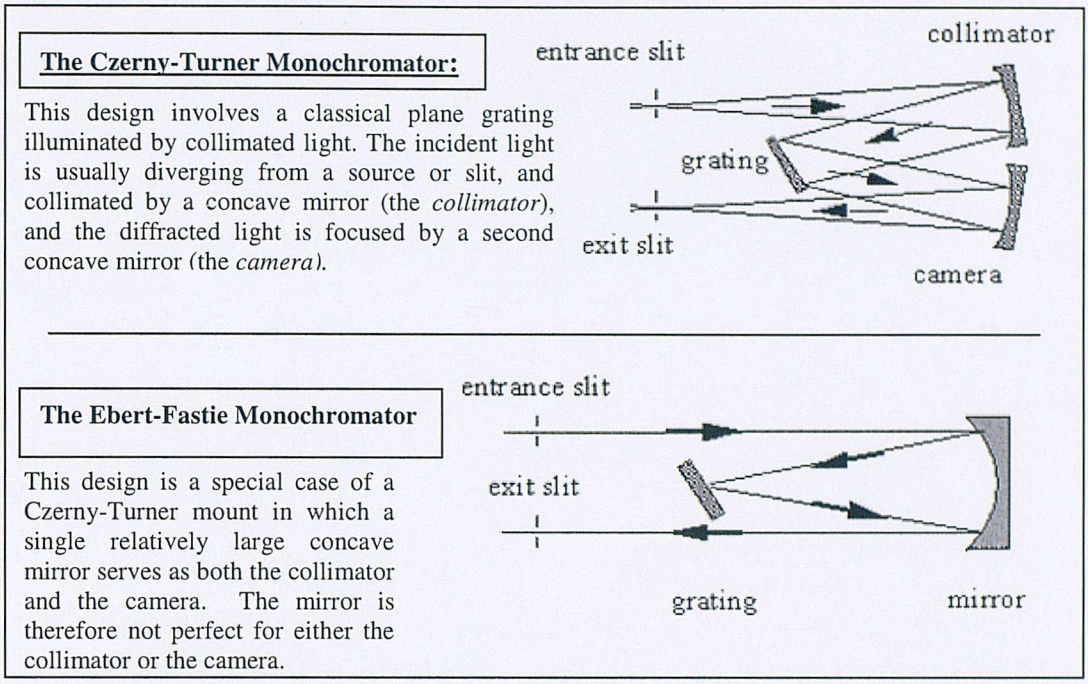


Figure 3.1.1: Schematic showing optical set-up inside monochromator [67]

3.1.5 Photo Diode Array

The Photo Diode Array (PDA) used is the Hammamastu S3904-512N, which is a 512-element linear array of 25μm pixel size capable of operating in the range 200-1100nm. The peak quantum efficiency is around 60% but falls to around 20% over the GUSTO range (200-270nm). The number of elements determines how many points on the differential one can actually resolve and therefore determines the quality in the quantization of the wavelength range. Also the shot noise over a two second integration time is roughly equivalent to a 10 ppb NO signal over 15m path length, which is clearly non-ideal. However the absorption cross sections of species such as SO_2 and benzene are larger in our range than those of NO and therefore the accuracy is not necessarily limited to the shot noise from the diodes. Also one can average successive scans together to reduce the random errors with better statistics. The array and its driver board are housed in the same mounting as the monochromator. The operation of the PDA is well understood [58] and thus only the characteristics relevant to its operation not already covered will be discussed further in chapter 4.

3.2 Data Acquisition Hardware

There are two main sections in the data acquisition hardware for the GUSTO system: the Analogue to Digital Converter (ADC) and the host PC within which it is housed and from which it is controlled.

3.2.1 Analogue to Digital Converter (ADC)

The output of the PDA and driver board runs directly to the ADC. Initially, a 12-bit ADC was used (PCDADIO ISA card from Arcom Control Systems Ltd.) however it was found that this does not give enough quantization levels to achieve reasonable resolution on the small differential signals that a typical DOAS or DUVAS system measures (discussed in detail in chapter 4). Consequently a 16-bit unit Computer Boards[17] PCMCIA card Model 'PCM 16D' was used, which sits inside a laptop and runs the real-time portion of the acquisition and analysis code. The ADC is in fact an 8 channel differential input unit where the first two channels are used as the input from the PDA; it runs up to 100KHz (output from the PDA is at 30KHz) and can only operate in ±5 or ±2.5Volt ranges. The PDA output is 0 to +5V, thus only half the available range is used (making it in fact a 15 bit ADC). The unit also has 4 digital control channels (for two way communication with the host PC) and two internal counters. One of each of these is utilised to inform the PC when a new scan is ready/finished (connected to the end of scan signal line from the PDA driver board) and one counter counts the number of diodes read in during each scan. This allows the user to be sure that no data have been missed as each separate data point is individually counted. The software records an error code if the count does not match, and the user is informed in real time that there is a problem. Control of the unit by the host PC is done in background mode which leaves the host system free to process analysis threads of the code as and when they are required.

[17] Company now called Measurement Computing [59]

3.2.2 Host PC

The ADC must be housed in some form of system; the simplest form of this is a basic computer. This system may then also store and perhaps analyse the data as they are taken. Prior to the end of 1999 the ADC used was a 12-bit ISA model and this was mounted in a computer of equivalent age (for 1999) in this case a 200MHz Intel Pentium unit. As data are taken at 2s intervals and the analysis requires reasonable computing power[18], no analysis was done in real-time. Once the 16-bit ADC was added to the system to replace the 12-bit unit a new PC was required (the 16-bit unit being PCMCIA instead of ISA), and a laptop was chosen. This unit is a 450MHz AMD K62 unit and is quite capable of storing and doing some preliminary analysis in real time. The hard drive of the laptop only runs at 4200 rpm (compared to 5400 rpm for desktop systems), however the increased processor and memory speed more than make up for the increase in hard drive seek time. The other main advantage of a laptop as opposed to a desktop computer unit is the ease with which it may be placed outside to take useful atmospheric data as opposed to laboratory data.

3.3 Summary of hardware

The output that is actually seen by the software environment within which all the analysis is performed is from the ADC. The ADC output is in the form of a stream of number sets, each consisting of 512 integers (in the range -32768 to +32767, with the current 16-bit unit) representing the signal on each of the respective 512 diodes in the PDA. Each number set is called a *scan* and it is from each scan that the various pollutant concentrations may be derived from analysis carried out by the computer. An example scan output is shown in figure 3.3.1.

[18] See algorithm outlined in chapter 5, most CPU intensive function is the real time graphical display done by automating Microsoft Excel [31] Tests showed the 200MHz unit to be insufficient.

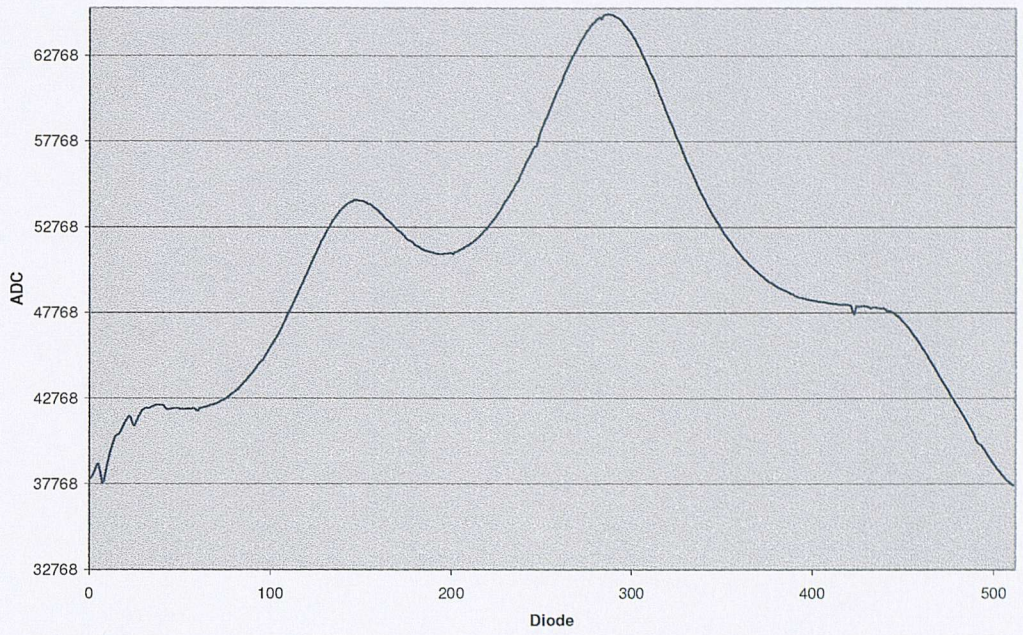


Figure 3.3.1: Example ADC single scan output, y-axis is ADC counts vs. Diode number.

Calibration of the components is discussed in the chapter 4. The hardware shown in figure 3.3.1 is summarised in table 3.3.1.

Hardware Component	Description	Section Referenced	Specifications
Lamp	Deuterium lamp acting as UV light source.	3.1.1	30 Watt Hamamatsu L2D2 L6565 [1] with corresponding Hamamatsu C4545 power supply [56].
Optics (see also table 4.1.1)	Provides parallel beam	3.1 3.1.2 4.1 4.4.8	f0.5 lens with 15 cm diameter diametric parabolic mirror for telescopes (aluminium on glass coated with magnesium fluoride).
Monochromator	Single Grating splits light into wavelength components	3.1.4 5.1.3	Ebert-Fastie CVI CM110SP with grating ruled at 1200 lines per mm and blazed at 330, maximum resolution 0.2nm. 100μm slit used[57].
PDA (Photo Diode Array)	UV sensitive PDA, measures intensity of output from spectrometer	3.1.5 4.4	512 element Hamamatsu S3904-512N, 200nm to 1100nm quantum efficiency 60 % (closer to 20% over range used) [58]

ADC (Analogue to Digital Converter)	Transfers signal from PDA to Computer	3.2.1 4.3	To 1999: 12-bit Arcom Control Systems Ltd. PCADADIO ISA card. Mounted inside a 200Mhz INTEL PC From 1999: 16-bit, 8 channel (diff) 100 kHz ComputerBoards PCM-16D/16 PCMCIA unit, mounted inside a 450Mhz AMD K6-2 Laptop [59]
Computer	Controls ADC and extracts concentration data from output from ADC	3.2.2	To 1999: 200 MHz Intel Pentium system with 2Gb HDD From 1999: 450 MHz AMD K6-2 Laptop with 6 Gb HDD

Table 3.3.1 Hardware specifications for the GUSTO system.

Chapter 4

Hardware Calibration

4 Outline

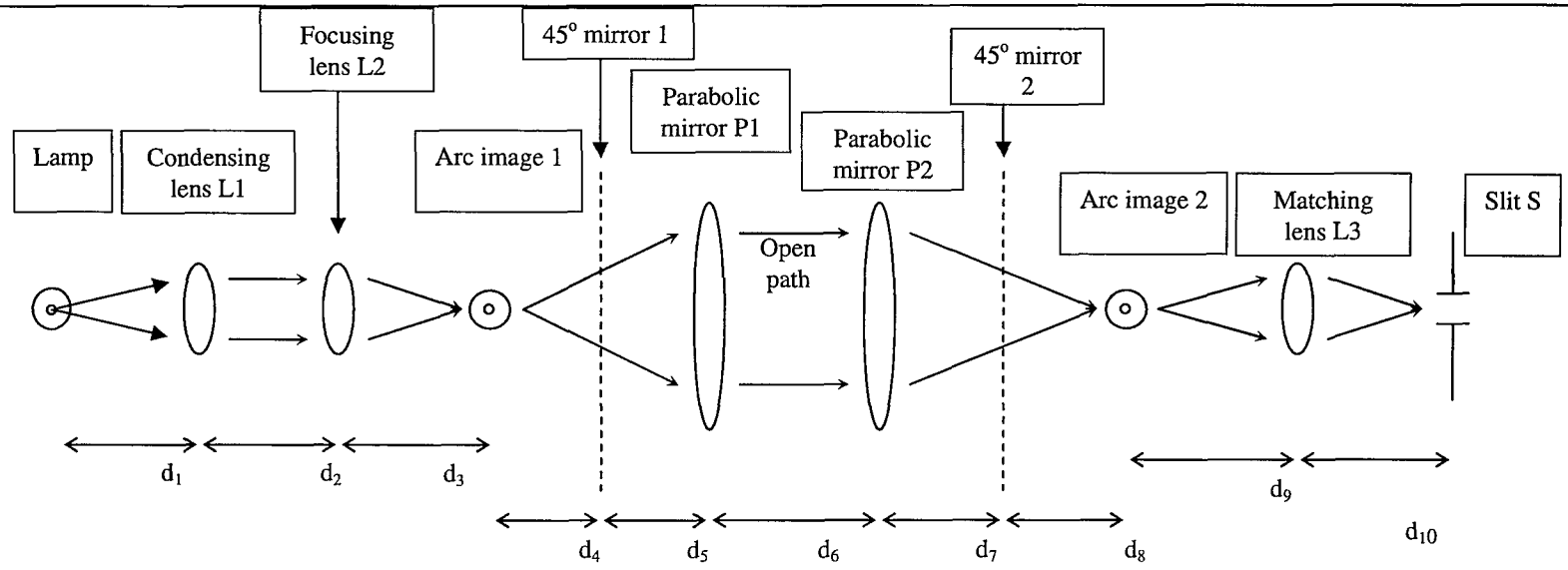
The DUVAS method outlined in chapter 2 relies almost exclusively on being able to resolve the relative size and shape of small absorption features present in the signals under study. The ability of any system to resolve such features is of key importance in its design and calibration phase. The three most important systems to characterise are: the optics as the light throughput determines the signal to noise ratio of the system before the PDA; the PDA as this measures the actual intensity of light as a function of wavelength; and the ADC as this interfaces the optical system to the computer based analysis system.

4.1 Optics

The optical set-up of GUSTO is outlined in chapter 2 (figure 2.6.1), with a more detailed schematic shown below in figure 4.1.1. The main optical components are further described in table 4.1.1. Light is emitted from the arc in the deuterium lamp and collected/focused by the lens combination L1 and L2. This focus is at the focal point of the first parabolic mirror P1 which makes the light along the open path plane parallel. Light is then collected by the parabolic mirror P2 and imaged again at the focal point of the matching lens L3 which focuses it with the correct f number (matching the f number of the spectrometer) at the spectrometer slit. The 45-degree mirrors are present to allow the optical system to be bent into two L shapes due to space restrictions, (see figure 2.6.1).

Optical component	Purpose	Characteristics
Lamp	Provides UV light input to system.	See figure 2.9.1
Condensing lens L1	Collects light from lamp.	f=25.4mm D=25.4mm f/1
Focusing lens L2	Focuses light at correct distance from P1 such that P1 output is optically collimated.	f=100.2mm D=38.1mm f/2.67
45 degree mirror 1	Changes direction of light so that system may fit in smaller container.	Elliptical with width = 38mm Height = 54mm
Parabolic mirror P1	Forms part of the telescope that is the two parabolic mirrors and allows collimated light to be used along the open beam path.	15cm diameter parabolic with f = 75cm f/5
Parabolic mirror P2	Completes the telescope that is the two parabolic mirrors and allows collimated light to be used along the open beam path.	15cm diameter parabolic with f = 75cm f/5
45 degree mirror 2	Changes direction of light so that system may fit in smaller container.	Elliptical with width = 38mm Height = 54mm
Matching lens L3	Matches the f numbers of the system to that of the spectrometer such that minimal light is wasted in interfacing between the two.	f=38mm D=19mm f/2
Slit S	The entrance slit of the spectrometer.	100μm, See figure 2.9.1

Table 4.1.1 Description of optical components



Distance	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
Ideal case	f_{L1}	Does not matter	f_{L2}	Fully fill 45° mirror	$d_4 + d_5 = f_{P1}$	Does not matter	Fully fill 45° mirror	$d_7 + d_8 = f_{P2}$	$d_9^{-1} + d_{10}^{-1} = f_{L3}^{-1}$	
Real case	Positioned such that Arc image 1 is f_{P1} from P1			Fully fill 45° mirror 1	$d_4 + d_5 = f_{P1}$	Does not matter	Fully fill 45° mirror 2	$d_7 + d_8 = f_{P2}$	$d_9^{-1} + d_{10}^{-1} = f_{L3}^{-1}$	
Where f_{XY} refers to the focal length of component XY										

Figure 4.1.1: Detail of GUSTO optics and description of distances.

4.1.1 Optical surfaces and light throughput

One of the main causes of photon loss in any optical system is the effect of the surfaces themselves. The GUSTO optics as outlined in figure 4.1.1 has 9 surfaces (including the two windows, one in each box, not shown) each of which causes up to around 4.5 percent loss. Thus the theoretical maximum throughput is of order 0.6. In fact there are many other factors that further reduce the light throughput, which are briefly discussed below.

The lamp has a quoted life of around 2000 hours; the light output decreases gradually over this range. The age of the lamp components will affect the actual wavelength dependence of the output as well as the overall intensity. The GUSTO analysis method as outlined in the previous chapter I is insensitive to on the gradual change of the light source, where gradual is taken to be much greater than the 2 second integration time of the system.

Any deposits on the optical surfaces will absorb or scatter in the UV. The GUSTO units are outside for a period of months at a time and the boxes within which the components are housed are not fully protected against the elements (except water) thus dust build up occurs on the surfaces. There is always going to be at least one surface that defines the boundary with the open path (in the case of GUSTO these are the box windows). The windows are transparent to the UV, however they are exposed directly to the elements and can get extremely dirty over a period of weeks/months. As with the lenses and mirrors (excluding the 45 degree mirrors which cannot be cleaned without severe surface degredation) these can be cleaned with ethanol. The box windows have also cracked along the entire diameter (one crack per window); because the box walls they are mounted in move relative to the windows. Clearly these cracks further open the boxes to the elements as well as affecting the light throughput directly. The total effect of these variables has been measured twice. In each case the total light throughput decreased from around 60 percent of saturation to around 10 percent over the 6 months of operation. The surfaces of the mirrors also degrade over time mainly due to their interaction with UV light and the resulting build-up of an oxidation layer over the

surface of the aluminium coating. In order to avoid the type of drop off in signal described above, the system should be cleaned on a monthly basis and the mirrors resurfaced around twice a year. Furthermore a UV transparent protective coating should be applied to the mirror surfaces.

4.1.2 Matching of components

If each optical component is not matched to the next (i.e. optimised) then light will be lost and the signal to noise will be decreased. Each lens should be fully filled by the previous component (the mirrors may be thought of as lenses) and the f number of the light entering the slit should match the f number of the spectrometer (as any other f number beam would either overfill or under-fill the grating).

The condensing lens L1 is positioned close to the lamp (around 3cm). Any closer and the heat begins to interfere with the lens mounting. Any further away and valuable light is lost. The focusing lens L2 is positioned such that the correct focus is achieved for the parabolic mirrors to provide a collimated beam. The 45 degree mirrors are positioned such that they are fully filled and thus no light is lost. The 45 degree mirrors cannot be positioned exactly as the boxes simply are not large enough. The amount of light lost is less than 10 percent in each case. The matching lens L3 is positioned to match the f numbers of the beam to that of the spectrometer. If the f number of the beam entering the spectrometer is higher than the f number of the spectrometer then the grating will be underfilled and spectral resolution will be lost. In the opposite case of the grating being overfilled valuable light would be lost.

4.1.3 Calibration of optical system

Once the system has been verified as retaining as much light as is possible the key to determining if the system is configured in an optimal manner lies in the collimation of the beam. This is not done by positioning components according to the theoretical

positions but rather by directly checking the optical collimation and making fine adjustments by eye. This is achieved by placing a flat reflective surface after the first parabolic mirror and directing the beam back into the system. Optical collimation is achieved when the back reflected focus from the parabolic mirror is in the same plane as the focus immediately before the said parabolic mirror.

The theoretical and real distances in figure 4.1.1 are shown in table 4.1.2.

Distance	Ideal case (mm)	Real case (mm)
d_1	$f_{L1}=25.4$	30
d_2	Does not matter	69
d_3	$f_{L2}=100.16$	106
d_4	$0.254\times(f_{P1})=190$	212
d_5	$(f_{P1})- d_4=560$	558
d_6	Does not matter	14870 (14.87m)
d_7	$0.746\times(f_{P2})=560$	578
d_8	$(f_{P2})- d_7=190$	175
d_9	$((f_{L3})^{-1}-d_{10}^{-1})^{-1}$	100
d_{10}	$((f_{L3})^{-1}-d_9^{-1})^{-1}$	61

Table 4.1.2 Comparison of distances shown in figure 4.1.1

4.1.4 Laboratory set-up

In order to obtain calibration data for the system warm-up and specific components a number of experiments were run in the laboratory. This differs from the set-up used at the roadside in a number of ways. The open path length must be smaller than 15m given our facilities. There is no need for condensing lens L1 or focusing lens L2 as the lower path length means that even without these lenses there is enough light throughput to saturate the PDA thus the mirror is illuminated directly by the lamp. There is no need to fold the light with forty five degree mirrors as the optical path in the laboratory is not constrained by the boxes used in the roadside system. The slit (100μm) is illuminated by being at the focal point of the final parabolic mirror (tilted so as to place the spectrometer away from the beam path). The parabolic mirrors used were similar to the ones in the outdoor system (15 cm circular diameter) but have a focal length of 60 cm. The matching lens L3 may also be ignored as it was found that light input to the

spectrometer was able to saturate the PDA without L3 being present. Thus the optical set-up for the laboratory measurements is as below in figure 4.1.2 with the distances used outlined in table 4.1.3.

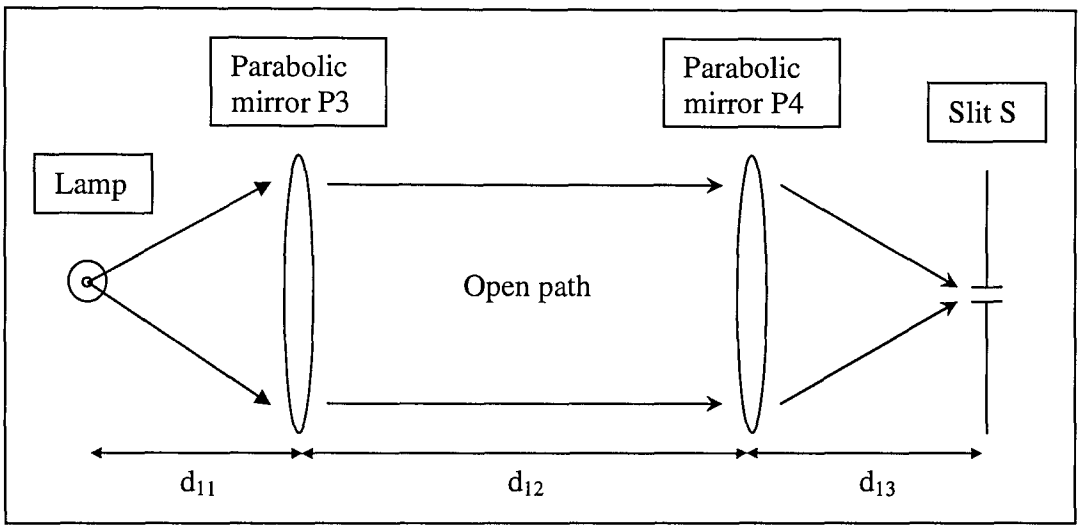


Figure 4.1.2: Laboratory optical set-up

Distance	Description	Value
D_{11}	Focal length of parabolic mirror P3	60cm
D_{12}	Open path for absorption cell	42.5cm
D_{13}	Focal length of parabolic mirror P4, slit is past focus so as not to saturate PDA. Focal spot size = 6mm	62cm

Table 4.1.3: Distances in laboratory optical set-up

A further use for the laboratory set-up was to enable an absorption cell to be used into which known amounts of pollutant could be introduced.

4.2 Data acquisition calibration

The basic operation of the data acquisition hardware was outlined in the previous chapter. The two main components that cause significant errors and therefore need rigorous calibration are the ADC and PDA. The ADC is the interface between the optical set-up and the computer that records and analyses the data, thus any systematic or random errors introduced will remain in the entire system from the point at which they enter. The three main aspects of the ADC are the spread of numbers given for any constant input (that is approximately the instrument function of the ADC), the possibility of missing or wrongly reported codes (the effective bit number of the ADC) and the configuration of the ADC (does one use the full range? etc). The PDA is strictly a part of the optical set-up however it is driven by an electronics board, simply a clock and start/stop signal lines, and itself sends outputs in the form of electronic signals. The three main aspects of the PDA are the random shot noise due to the finite number of photons interacting with any one diode (effectively dependent on the integration time used for the diodes), the systematic response differences from the diode array and the warm-up of the diode array (gradual increase in signal output over around the first one hundred minutes of data taking). The first two of these PDA characteristics are illustrated in figure 2.2.1 and all three are detailed in the following discussion.

4.3 Choice and characterisation of ADC

The first prototype of the GUSTO system was fabricated with a 12 bit ADC, however a single bit error would correspond to a differential signal (at two thirds maximum signal height) from the LHS of equation 2.1.6 of 3.67×10^{-4}.

The UK government guidelines (see tables 1.3.1, 1.3.2) state that concentrations of NO should be below around 16ppb, thus comparison of the above single bit ADC error with a 10ppb signal of NO over a 15m path in the 204nm range yields from equation 2.1.6,

$$15m\times\left(5\times10^{-19}cm^{2}\right)\times\left(1\times10^{-4}\right)\times10ppb\times\left(2.68\times10^{25}m^{-3}\right)\times\left(1\times10^{-9}\right)=2.0\times10^{-4}.$$

(4.3.2)

Thus if the ADC were the limiting source of error the signal to noise ratio (SNR) would be of order 0.5, which is clearly not acceptable. Thus a 16 bit ADC was bought and characterised.

The above is however a rather simplistic representation of how the quantization levels of the ADC affects the input to the analysis computer when taken with a DOAS-like method. This is illustrated in figure 4.3.1, which is a sample raw PDA output.

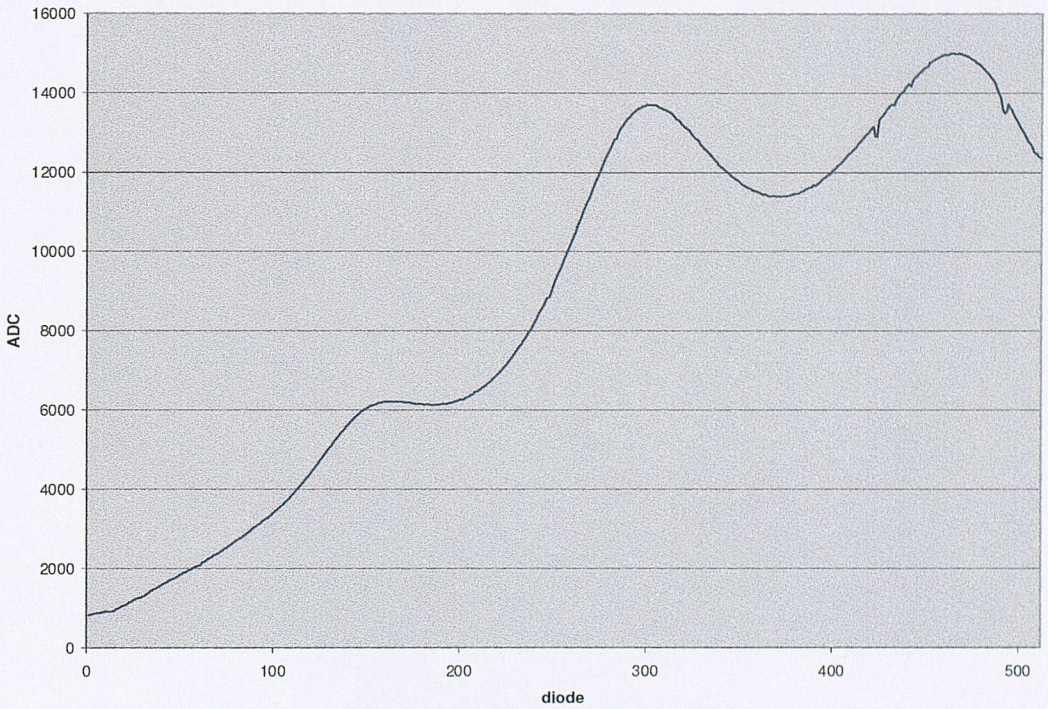


Figure 4.3.1: PDA output vs diode number, 0 = 200nm 511 = 260nm.

The 16 bit ADC used to obtain figure 4.3.1 was operating in differential mode in the ±5 Volt range; however, the output from the Hamamatsu PDA board is 0 to 5 Volts, thus zero input corresponds to 2^{15} and saturation input to 2^{16}. This gives 2^{15} available quantization levels and an effective bit number of 15. The interesting thing to note is that the minimum value of approximately 37670 and maximum value of 65220 mean that the signal over the 60nm range varies from around 15 to 99 percent of saturation.

Hence if the system were optimised to give the largest signal one would still have parts of the signal present at around 15% of saturation.

The key aspect of the DUVAS method is the relative size of the absorption features with respect to the background. Thus if the background signal decreases then so do the absorption features under study. The limiting accuracy of the ADC does not change, however, and hence the signal height has decreased while the noise level has remained the same. Therefore one must look carefully at the signal to noise ratio (SNR) for the ADC as it is not constant over the wavelength range of the data.

Figure 4.3.2 shows how the SNR of the raw ADC output varies for NO over increasing digitisation and fraction of full range. The x-axis scale represents the number of bits available to the unit while the y-axis shows how much (fraction) of saturation is being read by the input. The z-axis is the resulting signal to noise ratio given a 1-bit error on a DOAS-like signal.

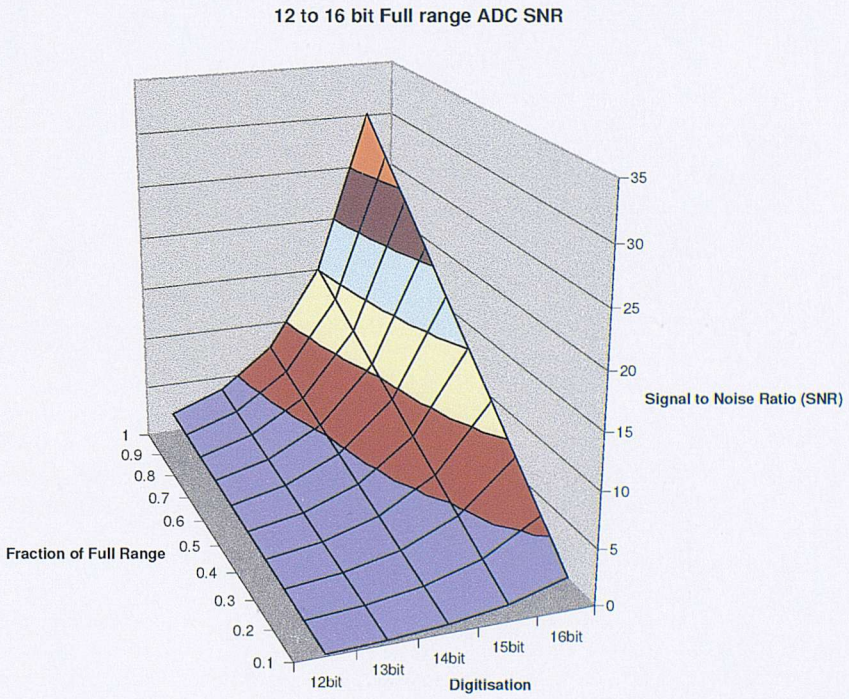


Figure 4.3.2: SNR of Fraction (0.1 to 1) of ADC full rage with Bit number between 12 and 16 for 10ppb NO over 15m path.

The 12 and 16 bit cross-sections are shown below for clarity in Figure 4.3.3

A 12 bit ADC would have a SNR all the way down to 0.2 while even a 16 bit ADC goes down to 3 at the lower end of the signal range.

As illustrated in figure 4.3.1 around 10 percent of saturation seems to be the lowest one expects from the GUSTO optics between 200 and 260nm. Figure 4.3.4 is another cross section from Figure 4.2.2, taken at a constant full signal range fraction of 0.1, showing the effect of increasing ADC resolution on SNR.

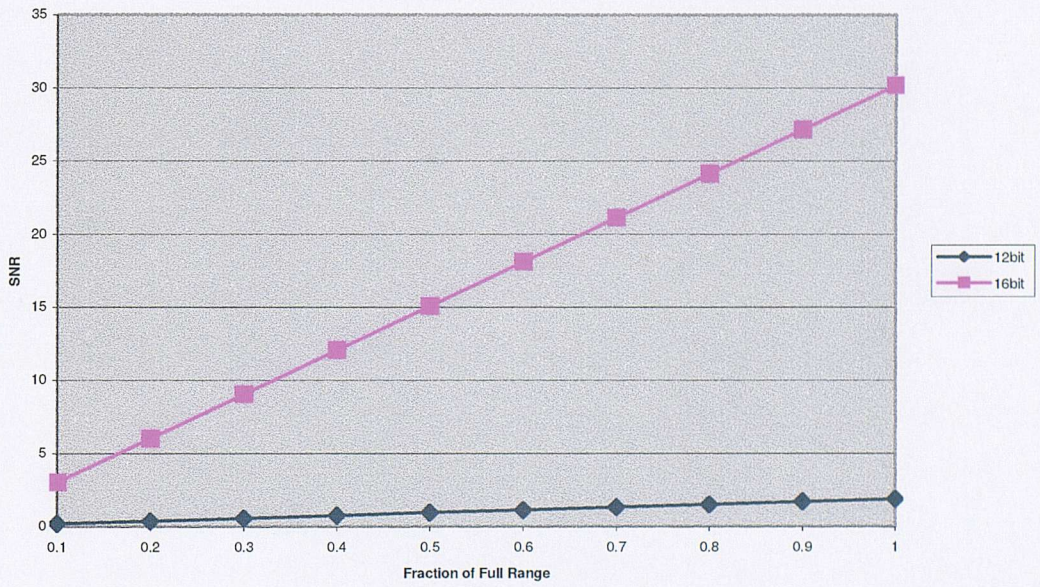


Figure 4.3.3: Superimposed cross sections of figure 4.3.2 at 12 and 16 bits, illustrating comparison between the original GUSTO ADC (12 bit) and the new (16 bit) component.

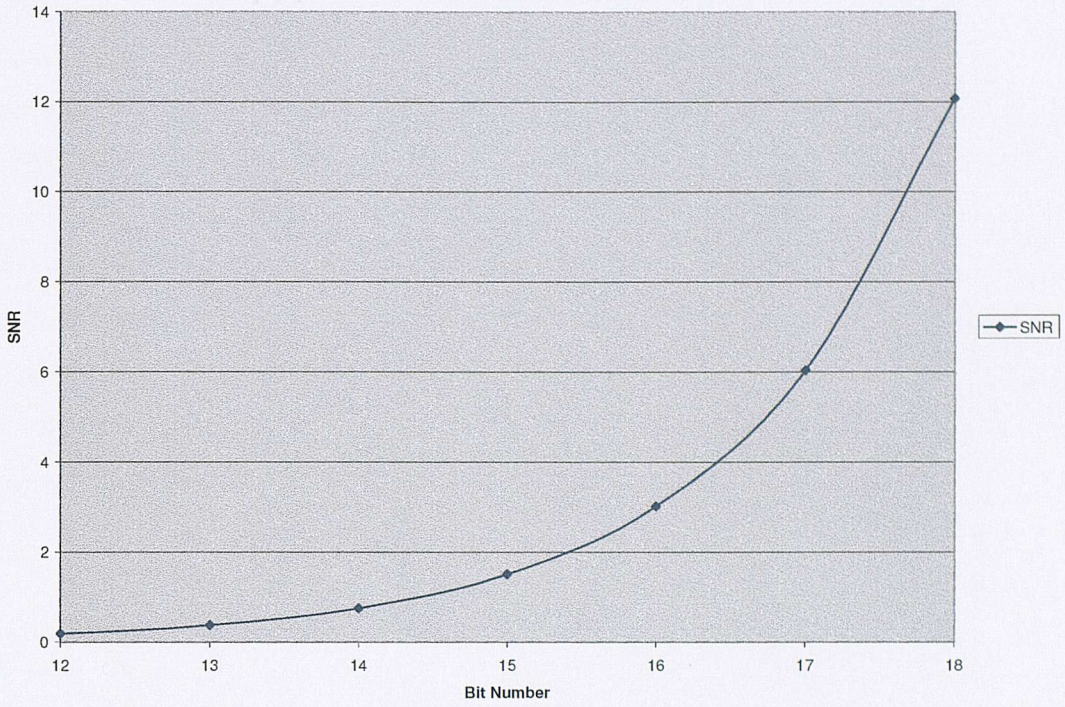


Figure 4.3.4: Extended cross section of figure 4.3.2 at constant full range fraction 0.1, illustrating change in SNR with increasing bit number for constant fraction of signal range.

The above discussion of SNR applies only to a theoretically perfect ADC which can accurately represent the sample input to a single bit error; in reality this is never the case and for a real ADC there will be two further considerations:

- The accuracy of any given ADC code (relationship between the theoretical input that should result in the given code) is not constant but varies from the expected theoretical value.
- As mentioned above our specific hardware only uses half the full range of our ADC, thus we use from zero to half of the above full range.

Data were taken with the chosen 16 bit ADC to characterise the above properties and these are summarised below.

4.3.1 Differential input

The chosen ADC has differential inputs (the output is not an absolute measure of voltage relative to ground, but rather the difference between the input connections). This has obvious advantages with respect to noise (any noise present on both input connections is automatically removed by the differential nature of the system). Thus the first characteristic to look at is the output if both inputs are physically connected together. This standard deviation of the resultant output is thus a very good description of the absolute minimum error for any ADC output code. Figure 4.3.5 shows the resultant histogram for our ADC with the differential inputs tied together.

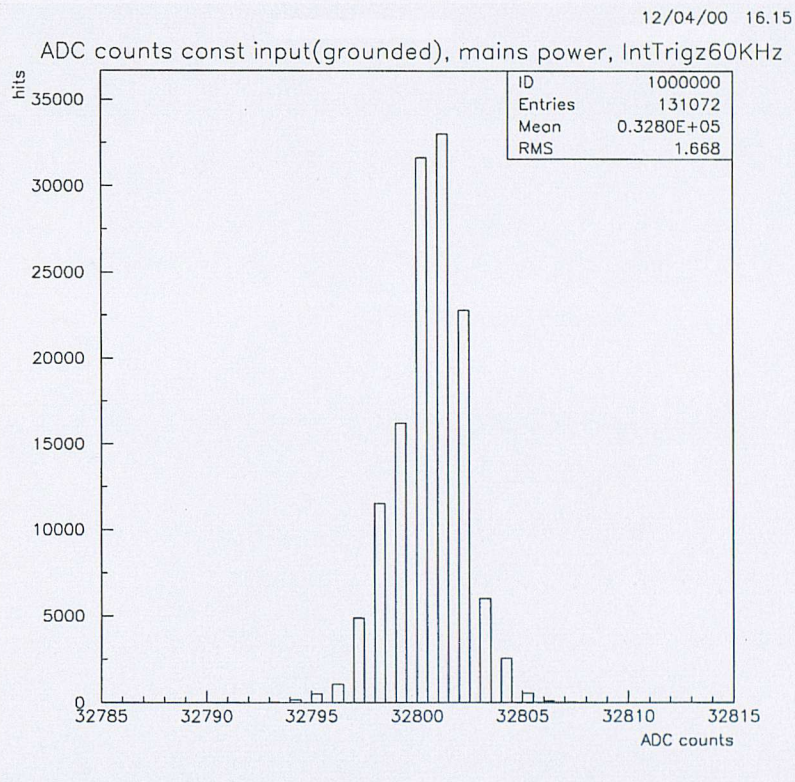


Figure 4.3.5: Differential input to ADC

Figure 4.3.5 is two seconds of data taken from a 60kHz internal trigger giving 131072 entries. The ADC was triggered at 60kHz, the frequency for which data points from the PDA are actually read. The integration time of the system is two seconds, thus changes

in the input on a longer timescale than this will not affect the final data. An RMS of around 1.7 is shown which is thus the lowest possible for this ADC.

4.3.2 Further Characterisation of ADC

A perfect n-bit ADC, given an input with a uniform probability density over the ADC's full range, would have an equal probability of getting any of its output codes equal to $1/2^n$. In a real ADC this is never exactly the case because the probability of getting any given code will be slightly different from every other. With an off-the-shelf unit this variation should be small but there may well still be missing codes and/or codes with probabilities off the quoted norm. In order to verify that the GUSTO ADC could be used reliably, it was decided to test this nonlinearity by measuring the transfer function for the unit (i.e. measure the probability of each code independently). The simplest way of achieving this would be to obtain a signal generator with substantially better precision than the ADC and measure at what input levels each code changed into the next, thus measuring the size of the bin corresponding to each independent code. Or one could use a random noise generator as an alternative. A sufficiently accurate system/noise generator was not available. The method decided upon was that of 'code density histogramming' [60], [61], whereby a signal of known shape is inputted to the system which samples at random intervals over a long period of time. Comparison of the resultant measurements with that from a theoretically perfect ADC can be used to determine the actual accuracy of the unit under study.

Investigations of this type need careful consideration with respect to the sample frequency (i.e. how random is it?) and also the signal shape (i.e. how accurately is it known? and how stable will it be over the course of the investigation?). A detailed discussion of this investigation for the GUSTO ADC has already been presented [4]. The author was involved in the investigation and the outcome is extremely relevant, thus a summary is included for completeness.

The main advantage of code density histogramming over any other method for characterizing an ADC is that it does not require a source of precision better than that of the ADC being tested and it is also not necessary to know precisely the amplitude, offset or frequency of the measured signal. In the case of the GUSTO unit a sine wave (peak to peak 9.9 volts, frequency 11.7KHz) was used as input with the sampling controlled by software [4]. The dynamic nonlinearity[19] and integral nonlinearity[20] were both calculated.

The dynamic nonlinearity gave the size of every ADC code bin between +0.6 and –1 LSB (quoted figure is difference from expected value of 1, thus range is 0 to 1.6 LSB), that is no ADC code was more probable than 1.6 LSB (+60% above the expected) and there were some codes with zero bin sizes (missing codes, 310 in total). Other than the 310 missing codes, no bin was less than –0.6 LSB (60% below the expected). Of the 65535 codes around 45% were ±0.6. Around 0.4% were missing codes and the remaining 54.5% were either ±0.3. Thus the result suggested by figure 4.3.5 that the rough variation for a fixed input should be around 1.7 LSB, appears justified and thus any given code measured stands approximately a 50% chance of being incorrect by 1 LSB.

Clearly if all the missing codes occurred in the same range then there would be a substantial problem with any data taken in that range. Fortunately this is not the case as the first missing code is 127 (binary 1111111) with every successive missing code being +211 (binary 11010011) greater than the previous.

During the 310 missing codes every bit changes value, making it unlikely that the missing codes are due to stuck bits; further testing proved that varying either the sample or signal frequency did not change any missing codes, thus they are also not due to beating occurring in the sampling process. The best explanation for the presence of this type of missing codes is some form of bit inter-modulation[21]. Also note that that the ADC codes are converted in software from signed short integers (-32768-32767) to

[19] Size of bin for each ADC code measured in least significant bits (LSBs) [61]
[20] Integral of dynamic non-linearity over all ADC codes, i.e. a measure of linearity for unit over all [60]
[21] The change of one, or a group of, bits affecting the probability of a change of other bits [60], [61].

unsigned short integers (0-65535), which changes only the MSB of the bit pattern. The effect of these missing codes on data will be an error of 1 bit in around 300. As data taken for samples less than 1ppm (SO_2, see figure 2.3.6 and equation 2.1.6) will be less than 500 ADC counts (residual between signal and background) this does not present any form of limiting error, c.f. PDA response in section 3.4 which is of order 20 ADC counts in any range. Mathematically one can express the effect of the missing codes as an alteration to the effective bit number of the unit given in equation 4.3.3

$$\log_2\{65536-65535/211\}=15.99 bits \tag{4.3.3}$$

The technical documentation for the ADC claims an input error of less than 2 (clearly the system would not be 16 bit if every input data point was only accurate to 2 or more input codes and thus cannot be sold as such). However, the above measurements imply that in a real configuration reading at 60 percent of maximum throughput (the unit is quoted as 100kHz compliant) there is noise present which limits the accuracy to +/- 1 LSB. This taken with the fact that we can only use half of the full range of input codes means that with maximum light throughput the highest part of our signal would only be seen by an effective 14 bit ADC. This means that the SNR shown in the first part of this chapter must all be divided by a further factor of two for our real ADC. The Signal to Noise ratio for the chosen 16 bit ADC over its full range is represented by figure 4.3.6.

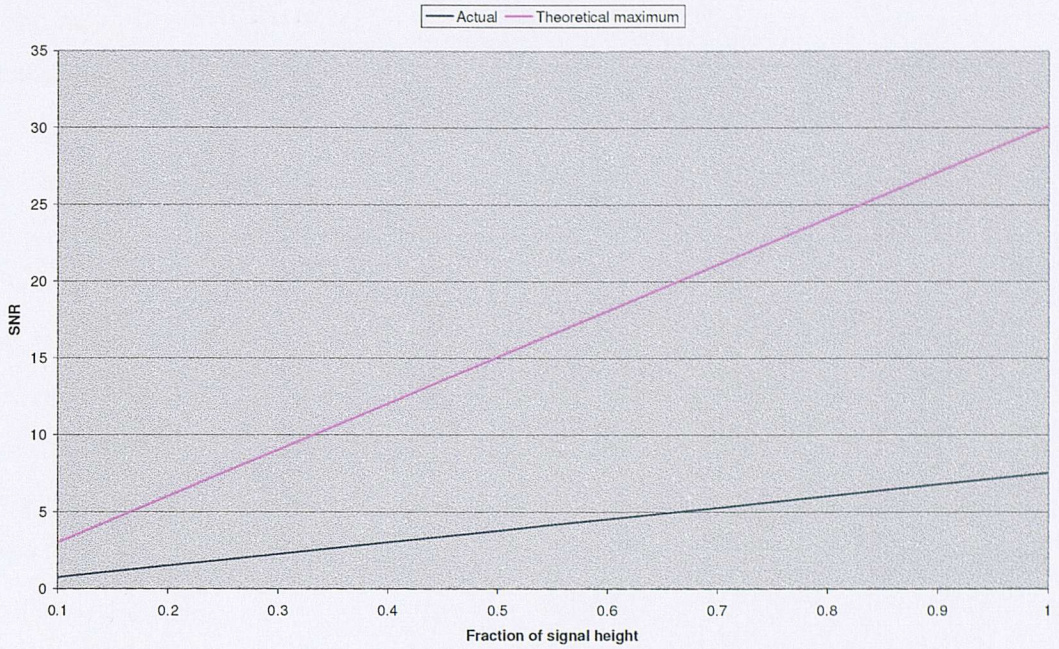


Figure 4.3.6: Predicted SNR for chosen ADC over full range assuming saturated light input and that the ADC represents the limiting source of noise.

To obtain a realistic signal to noise ratio of around 10 means that the data at the sides of the PDA (particularly the lower wavelengths) may be lost (see shape of Figure 2.2.1), although further analysis using statistical techniques such as the mask fitting outlined in the previous chapter have yet to be applied and will be further discussed along with real data in following chapters.

4.4 Characterisation of the Photo Diode Array

The other main component of the GUSTO system whose behaviour must be characterised is the Photo Diode Array (PDA). The DUVAS method (as outlined in the previous chapter) is very sensitive to small changes between successive wavelength intervals and thus the relative response of the PDA between wavelength intervals is of paramount importance. Any change pixel-to-pixel may otherwise be taken for signal.

The quoted specifications of the PDA [58] indicate a maximum variation pixel to pixel of 3%. This corresponds to a differential signal (at two thirds maximum signal height)

from the LHS of equation 2.1.6 of 3.67×10^{-4}. If this were the noise in the system it would give a SNR relative to a signal from 10ppb NO from equation 3.2.2 of 6.6×10^{-3}.

Given typical signals this would correspond to a noise level 150 times the signal. The pixel-to-pixel variation is clearly a feature that needs to be characterised. There are three aspects of the system that require attention. First for any data to be taken with the PDA the system must be left on for a number of minutes and it was discovered early in the life of the data acquisition hardware that this causes the temperature of the system to rise by a few degrees. This temperature rise causes the output of the system to change by a few percent over approximately the first 100 minutes of the warm-up. Secondly, the PDA gives out non-zero data for zero input (dark noise): this is due to the thermal properties of the system. At non zero input the diodes also exhibit shot noise, another form of white noise due to the finite number of photons interacting with the diodes. Thirdly, the diode-to-diode response is different to uniform light input. These three characteristics of the diode array are examined below.

4.4.1 Dark current, small timescale variations

The way in which the diodes that make up the PDA respond to zero input is to leak current. This leakage is quoted as being maximum 0.3pA and typically 0.1pA [58], the capacitance of each diode is 10pF [58] thus we may calculate the expected dark signal as between 0.02 and 0.06 Volts.

This voltage is reached assuming a 2 second integration time, the maximum output is 5 V and thus the dark current is expected to be 1.2% at maximum with 0.4% typical. Each diode should behave approximately the same as the others, thus this should be the overall trend of the array with the small-scale variations that we are interested in being of a lower magnitude. Figure 4.4.1 shows real dark data (diode array input covered with black card) while figure 4.4.2 is the first 50 diodes shown in figure 4.4.1

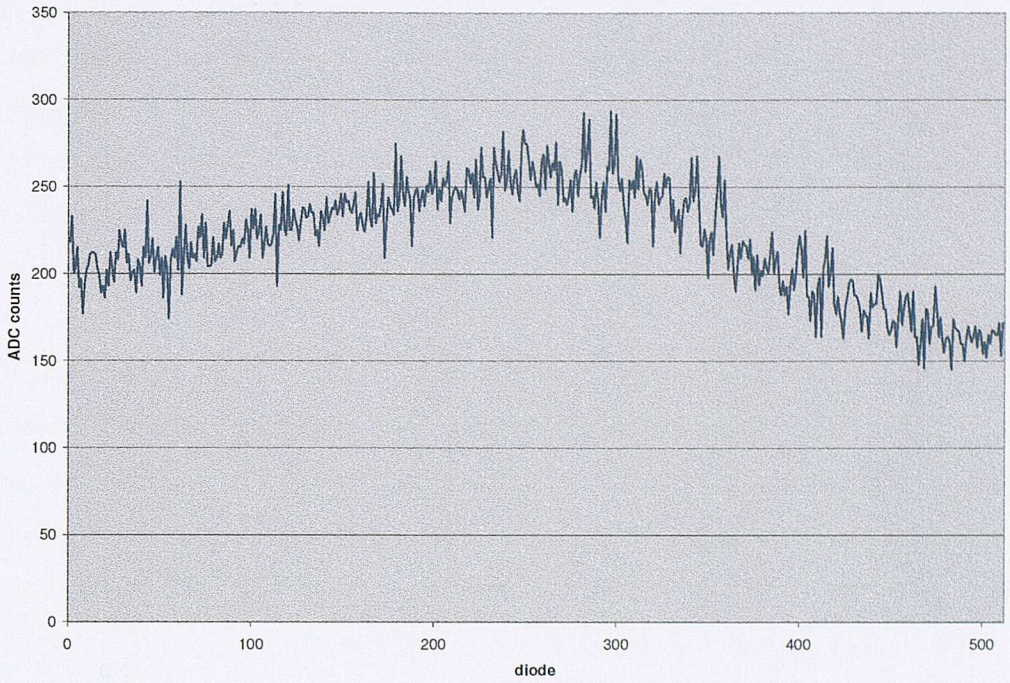


Figure 4.4.1: Sample dark current data set from July 01. ADC used was 16 bit in differential mode (zero is 32768 and saturation is 65535)

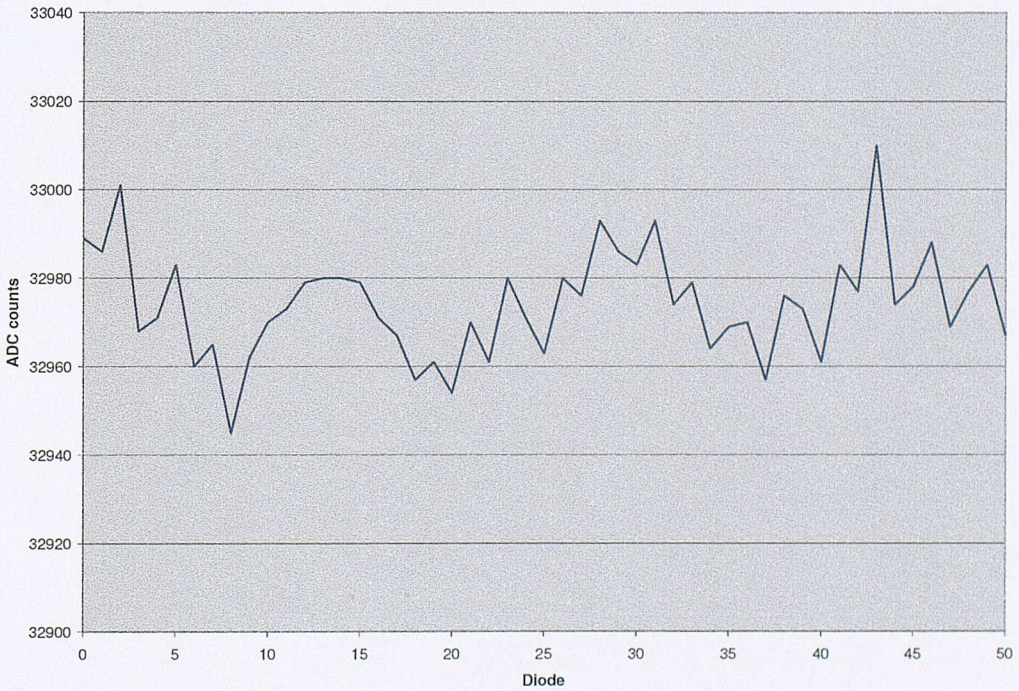


Figure 4.4.2: First 50 diodes from Figure 4.4.1.

From Figure 4.4.1 we may see directly that the approximate average for the entire 512 element PDA is 32990 which corresponds to 220 ADC counts and around 0.6% of the full range. This compares well with the quoted typical value of 0.4%. The spread of the array is around 140 ADC counts corresponding to 0.4% of the full range and the maximum value of 33060 is 292 ADC counts and representing 0.9% of the full range, well below the quoted maximum of 1.2%.

Figure 4.4.2 shows that the variation pixel-to-pixel is of order 20 ADC counts, which represents around 0.12% of the full range. This represents a differential signal of around 6×10^{-4} or a SNR of order 0.5. Thus the dark current must be compensated for if a usable signal is to be achieved. The standard way to achieve this is to recognise that it is due to the thermal response of the diodes and therefore a form of white noise, thus averaging of scans may be used as a method of eliminating its effect. Figure 4.4.3 illustrates how averaging scans affects the overall shape and is from the same set as the example scan in figures 4.4.1 and 4.4.2.

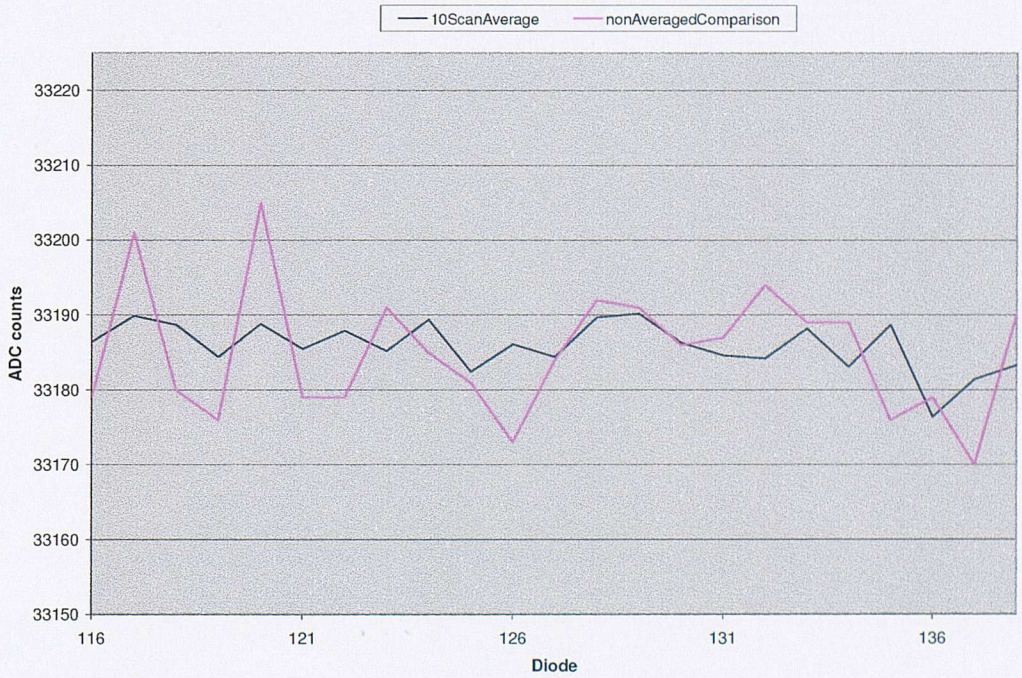


Figure 4.4.3: Effect of averaging real dark current data (10 scans) shown with non-averaged comparison

As one would expect, with around 10 scans averaged, the SNR has dropped by a factor of 3 to around 2.

Clearly the dark current variations are large enough to produce a SNR down to 2 even with 20 seconds of averaging. The characteristic that really matters is not the output without light but the output with light as this also includes the dark variations but allows systematic calibration. Once the diode-to-diode response had been characterised (see sections 4.4.3, 4.4.4 and 4.4.5) the error diode-to-diode is in fact much lower as even the dark current output of the array is consistent and not completely random. This is illustrated in figure 4.4.4 which shows the first same scan as in figures 4.4.1 and 4.4.2 plotted over a single scan taken exactly two minutes later. There is a shift of about 20 ADC counts between the data sets that has been removed for comparison.

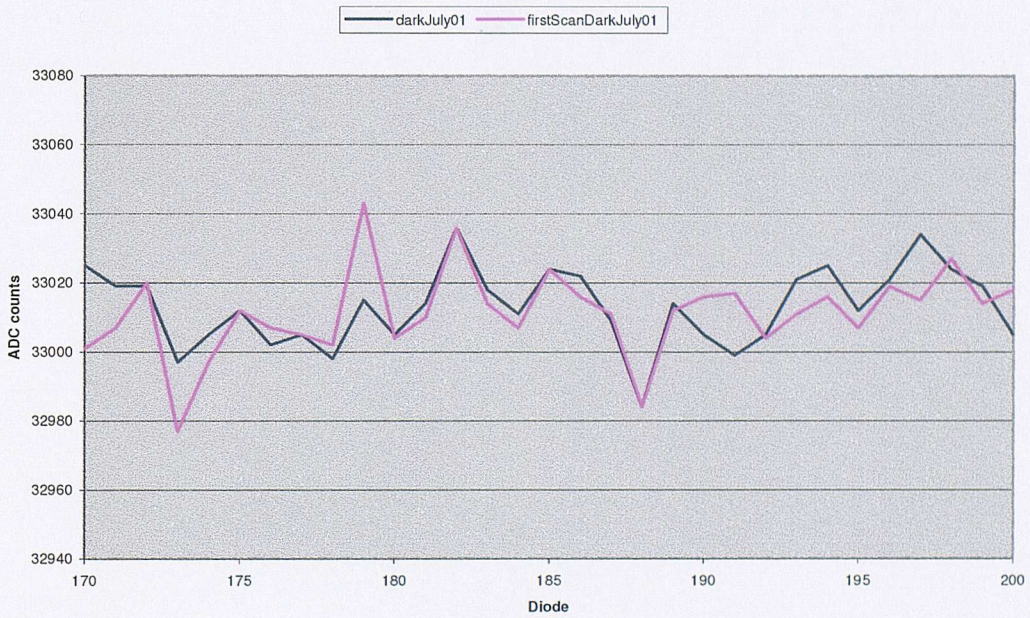


Figure 4.4.4: First and 120th scans from dark data

The similarities between the data sets in figure 4.4.4 can be easily seen, showing that the systematic differences diode-to-diode need to be exactly calibrated if a usable SNR is to be obtained. Three independent methods were used to measure this diode response calibration and they are outlined in sections 4.4.3, 4.4.4 and 4.4.5.

4.4.2 System warm-up

Once the optical system is activated, with light throughput from the lamp to the PDA, light interacts with the diodes, which change in temperature, as does the lamp itself. Both of these components are sensitive to temperature and the signal will change as a function of time until equilibrium is reached. The exact temperature dependence of the lamp and PDA is only useful information if the temperature changes in the system are also accurately known, however the system change may be directly measured and, with the 5V configuration as outlined in figure 4.1.1, the warm-up is as shown in figure 4.4.5.

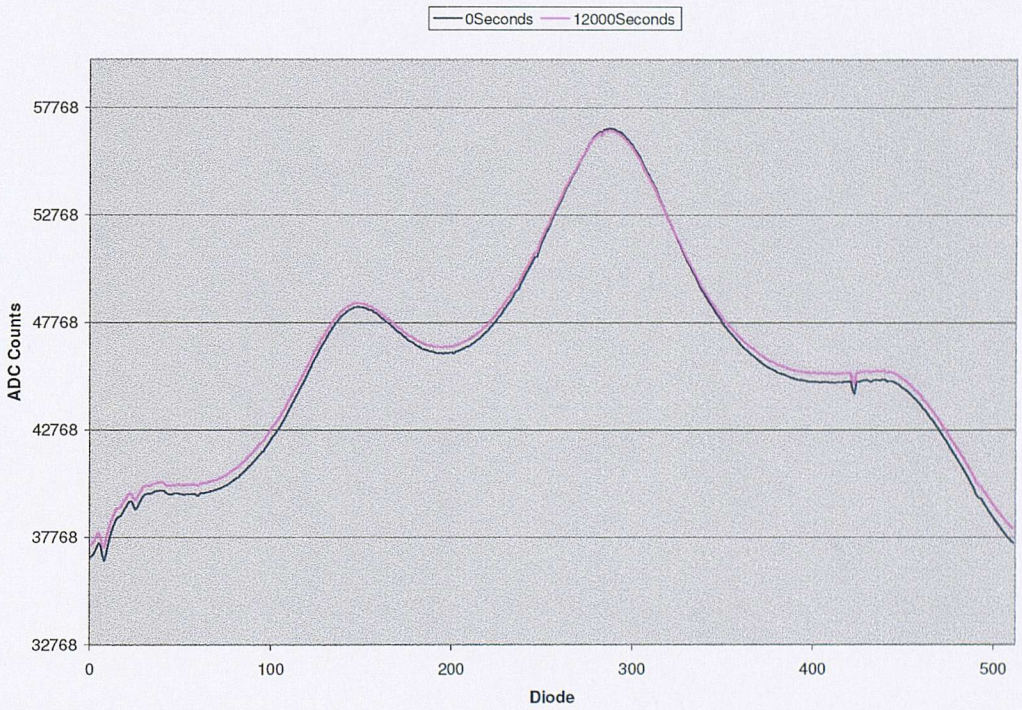


Figure 4.4.5: System warm-up first and last scans shown with total experiment duration at 4 hours.

Data were taken for 12,000 seconds with the system starting from cold (left over the weekend). Clearly there is a significant rise in the signal over the four hour period, however it is not uniform over the diode array. This may be partially due to the fact that the centre of the array is close to saturation and the PDA response is no longer uniform. The diodes do however change in a similar way over the course of the warm-up but by

different amounts, that is they vary slowly around their initial values. Figure 4.4.6 shows the change of the first and middle diode (number 1 and 255) as a function of time with the initial starting value scaled to 1. Thus each x-axis point is the average of 60 actual scans divided by the initial start value and each value is scaled to reflect zero light input giving zero ADC output.

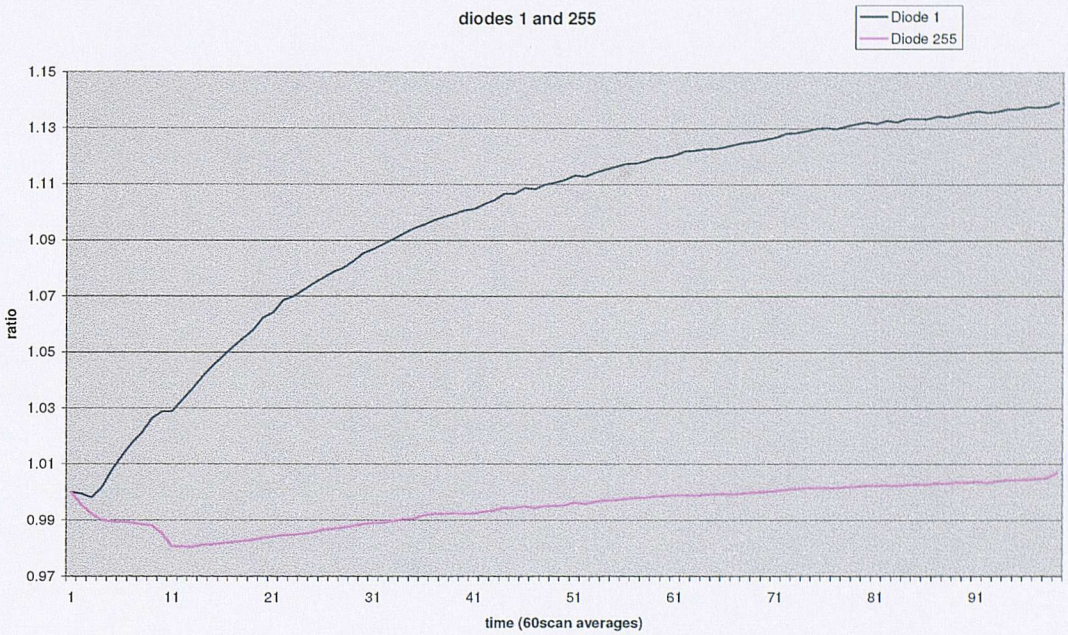


Figure 4.4.6: First and middle diode changing as a function of time over four-hour period

The first 22 minutes of the warm-up actually result in the diodes closer to the centre of the array going down in value, which could be due to the power supplies warming up or the system settling down to a steady state. After the 11^{th} time interval (22 minutes) all the diodes show a continuous rise that gradually slows. It was decided to use this long timescale rise to compare the diodes by calculating the percentage rise compared to the starting value between the end of the data set and the 11^{th} data point. This comparison is made in figure 4.4.7, which shows all 512 diodes.

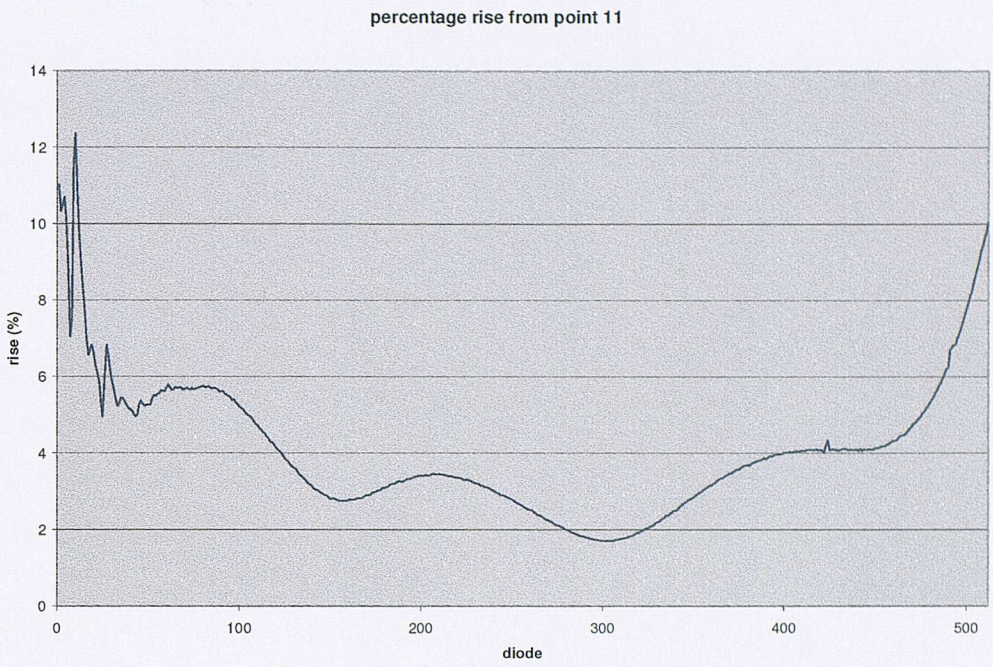


Figure 4.4.7: Percentage rise of ADC counts from 11th time interval for each of the diodes in the PDA

Figure 4.4.7 is interesting as it appears to have an inverse relationship with the actual signal on the diode array shown in figure 4.4.5. If lamp output were stable and the PDA was responsible for the structure in the percentage rise graph then the intensity of light interacting with the diodes would govern the shape of figure 4.4.7. However relatively small features in the lower portion of the array shown in figure 4.4.5 turn into relatively large features in the percentage rise graph of figure 4.4.7. The peaks at the low end of the spectrum (0-50 diodes, 200-207 nm) show considerably more variation compared to the remaining structure in the percentage rise graph than the original data graph. This suggests that the peaks themselves are changing relative to the rest of the signal. This may be understood by the fact that 200 nm is the lowest wavelength produced used from the lamp, and the lamp specifications [62] quote a drop off of around 35 percent from 250 to 200 nm. Furthermore the output of the arc is quoted as having features similar to those shown at around 200 nm for the low regions of its output spectrum. The overall shape is governed by the spectral response of the PDA.

The actual change of the signal over time may be seen in figure 4.4.7 to vary from around 10% for the end diodes to around 1.8% around the middle of the array. Assuming data were to be taken with approximately five scans being averaged for every data point then we may calculate the effect of the system warm-up on the quality of data taken as follows:

If the change is around 5 percent for the entire array over 4 hours, for a half height signal this corresponds to around 16000 ADC counts ± 800, which equates to a 0.5 ADC count error on five scans. From equation 2.1.6 this corresponds to a differential signal of 3×10^{-5}, which in turn corresponds to a signal to noise ratio from equation 4.3.2 of around 7, (10 ppb NO in 15 m of air).

The shorter timescale variations in the signal shown in figure 4.4.6 are for the middle diodes and although the changes around point 11 in figure 4.4.6 are of the order of 10 times faster than the gradual "four hour" changes later in the warm-up, they are all of order one percent. Therefore, the signal to noise for the non gradual features within the warm-up work out at around the same value as those for longer timescales and are therefore within the usable limits of the ADC subsystem.

4.4.3 Uniform intensity input for diode response calibration

Figure 4.4.4 demonstrates the need for accurate calibration of the diode-to-diode response as scans at very different times for no light input are not random. The response to actual light input will also be different and the combination of these two is the actual diode-to-diode response. The relationship of this response relative to the other components of the signal is outlined in figure 2.2.2. The most logical method for characterising this response is to give all diodes exactly the same input at the same time (uniform light input).

The optical set-up for such a uniform light intensity measurement [63] is shown in figure 4.4.8.

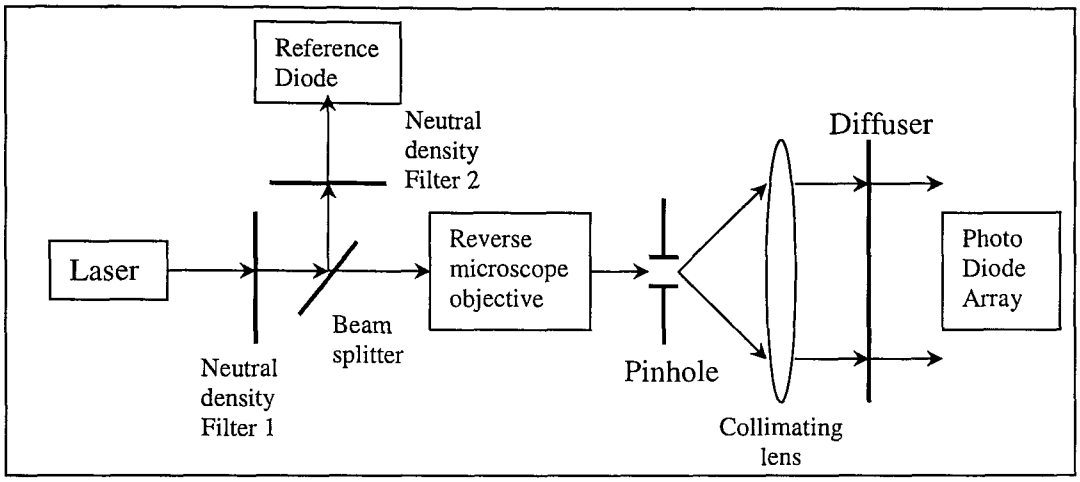


Figure 4.4.8: Optical set-up for uniform light intensity calibration of

The light from the laser is far too intense to be allowed to fall directly on the diode array or reference diode even after it has been expanded into a diffuse beam, thus neutral density filters are required [64]. The laser will have intensity that is a function of time, thus a reference diode is used to keep track of the change in intensity of the diode. This reference diode was connected to the same laptop that runs the ADC via an RS232 cable and set to scan once every two seconds so as to match the data taken from the PDA. In order to compare the reference diode output with the PDA, the reference diode's signal should be integrated over time. Once the intensity of the beam has been reduced by the neutral density filter and split so that intensity variations may be recorded, it is passed through a reverse microscope objective. A correctly placed and small aperture (pinhole) may then be put across the beam to remove all spatial information. The reverse microscope objective and pinhole operate as a system from which a point source is formed at the focal point of the collimating lens, which in turn produces a plane parallel beam without spatial variations that illuminates the PDA. Thus a Gaussian intensity beam hits the PDA whose overall changes in intensity are monitored by the reference diode. In order to correctly interface with the PDA the diffuser should be spun with a period of around $1/10^{th}$ the integration time. The purpose of each component is summarised in table 4.4.1.

Component	Description	Purpose
Laser	Helium Neon 4mW	Provides light input to system
Neutral density filter 1	0.1-0.9 transmission depending on which filter chosen	Cuts intensity to a level low enough to avoid damage to PDA
Beam splitter	Half silvered mirror	Allows reference diode to measure overall beam intensity
Neutral density filter 2	0.1-0.9 transmission depending on which filter chosen	Cuts intensity to a level low enough to avoid damage to reference diode
Reverse microscope objective	Standard microscope objective mounted with integral pinhole for IC GERB calibration	Serves with pinhole to provide point source for collimating lens
Pinhole	Variable narrow diameter aperture on micrometer controlled three axis mounting	Removes spatial information present in output from reverse microscope objective
Collimating lens	5cm diameter 10cm focal length convex fused silica lens	Provides parallel beam to PDA
Diffuser	5cm diameter diffusing disc	Removes interference pattern caused by multiple reflections within neutral density filter 1.

Table 4.4.1: Summary of each component in figure 4.4.8

If the input to the PDA were uniform then the relative differences in the resultant output of the PDA would be a direct measure of the diode-to-diode response differences. However there are in fact four signals that make up the output of the PDA. The diodes have a random shot noise output, thus data must be taken over a long time to eliminate these variations by use of good statistics. The intensity of the laser changes with time, which is illustrated for the 500 scan duration of the experiment in figure 4.4.9. The diodes each respond differently to a given intensity/wavelength of incident light. The optical system itself introduces changes of order one diode in width (no monochromator is present thus the array is measuring input intensity as a function of distance along its length of 1.5 cm). The only way to compensate for the spatial non-uniformities due to the optical set-up is to move the PDA relative to the input light. This is difficult unless the array is mounted on some form of movable stage, which it was not as the equipment for this type of operation was not available at the time.

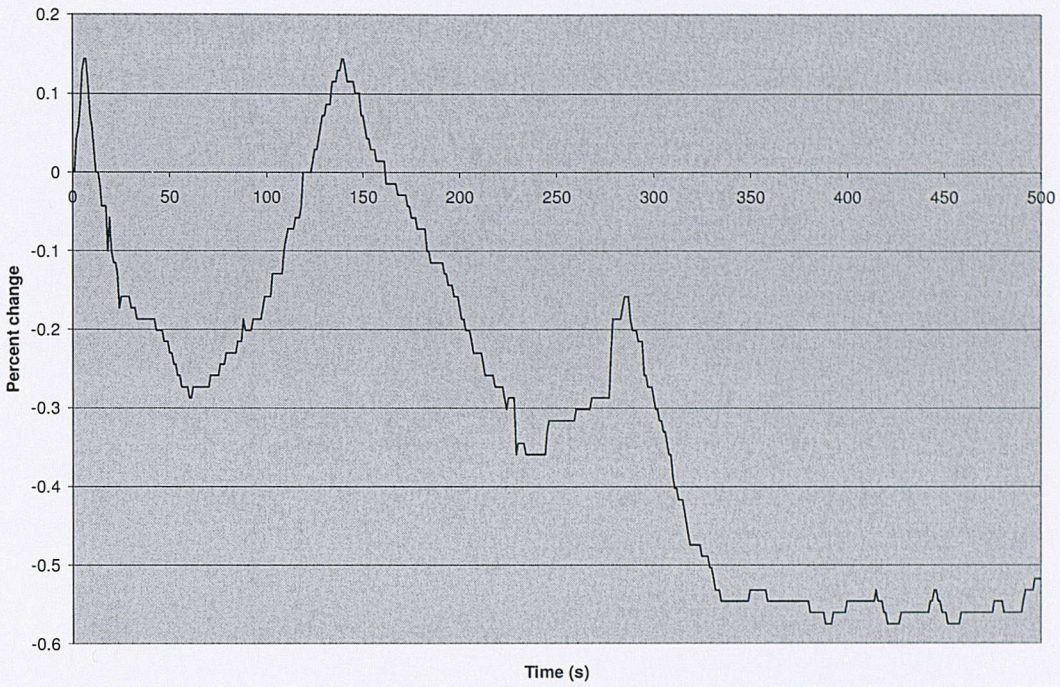


Figure 4.4.9: Laser intensity percentage change from initial value

Without the diffuser present there is an interference pattern present affecting around 20 percent of the signal intensity resulting from multiple reflections within the first neutral density filter. This pattern is non random and makes moving the array relative to the signal almost impossible. Adding a diffuser removes this interference pattern, however, it introduces large random fluctuations in the intensity (not dependent on time). This is illustrated in the sample output for figure 4.4.10.

The peaks present in figure 4.4.10 are of order 1000 ADC counts and represent the form of the spatial variations in the light beam present in the calibration set-up. A number of experiments were done to move the PDA relative to these spatial variations, however the standard methods for compensating for such a systematic error [48] require distance accuracy of order one diode in magnitude (0.03mm) which is simply not attainable without some form of high precision movable stage (see following section 4.4.4). An alternative method of removing the spatial signals arising from the high level of coherence of the light source (termed laser speckle) is to rotate the diffuser relative to the light beam. As the diffuser rotates the speckle pattern oscillates within its limits. If the period is fast enough (around 10 rotations per data scan) then the speckle signal will

be effectively obscured during the integration time of the PDA. A mechanical system to rotate the diffuser was also not available at the time of the experiment.

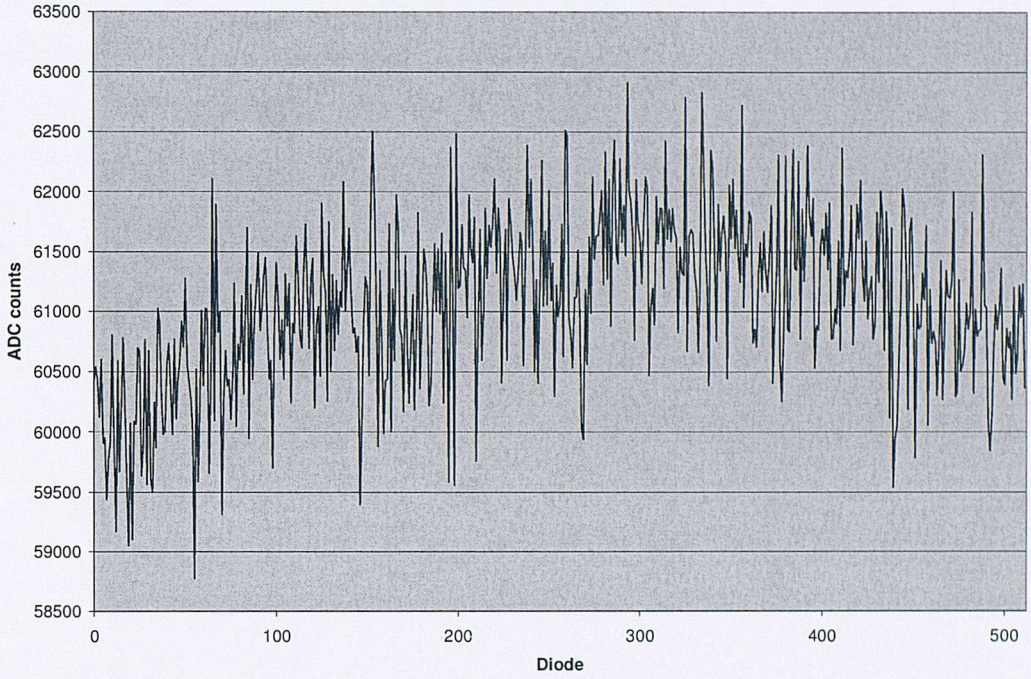


Figure 4.4.10: Sample data scan from uniform light calibration of PDA with diffuser present

Thus the first method for calibrating the response of the PDA to uniform intensity input yielded a set of spectra that contained both the variations in intensity due to the method and the PDA, however it proved to be simpler to use a different method (in fact two further methods of calibrating the diode-to-diode response were implemented and are discussed below) than to separate these two signals from the retrieved data.

4.4.4 Single diode calibration of diode-to-diode response

The uniform light (Hg line at 254nm) input calibration for the diode-to-diode response had two major flaws. Firstly, the light input to the PDA was not uniform and was not controllable across the array with enough precision. Secondly the light was from a visible laser and the diode response also varies as a function of wavelength. The PDA thus needed to be illuminated with a UV light source and if each diode was to be separately analysed this would need to be a narrow source. Therefore a mercury lamp

was used. The lamp was positioned behind a metal shutter (two 3mm thick aluminium plates separated by a 3mm air gap) and in front of a slit (25μm to match the pixel separation) such that the mercury line was narrow enough to illuminate only 3 diodes at a time. The entire PDA with power supplies and cables was mounted on a translation stage [64] capable of moving the array across the light source with micrometer precision. Data was taken at each diode for a time (50s, at one data element every 2s) for each diode and then back to diode 255. Dark current data was taken for 10s between every diode data block. Thus each diode may be compared directly to similar data taken for diode 255 taken immediately before and after. Further each diode data block may be normalised with dark current data taken immediately before and after the corresponding diode data block. The corresponding data is illustrated in figure 4.4.11.

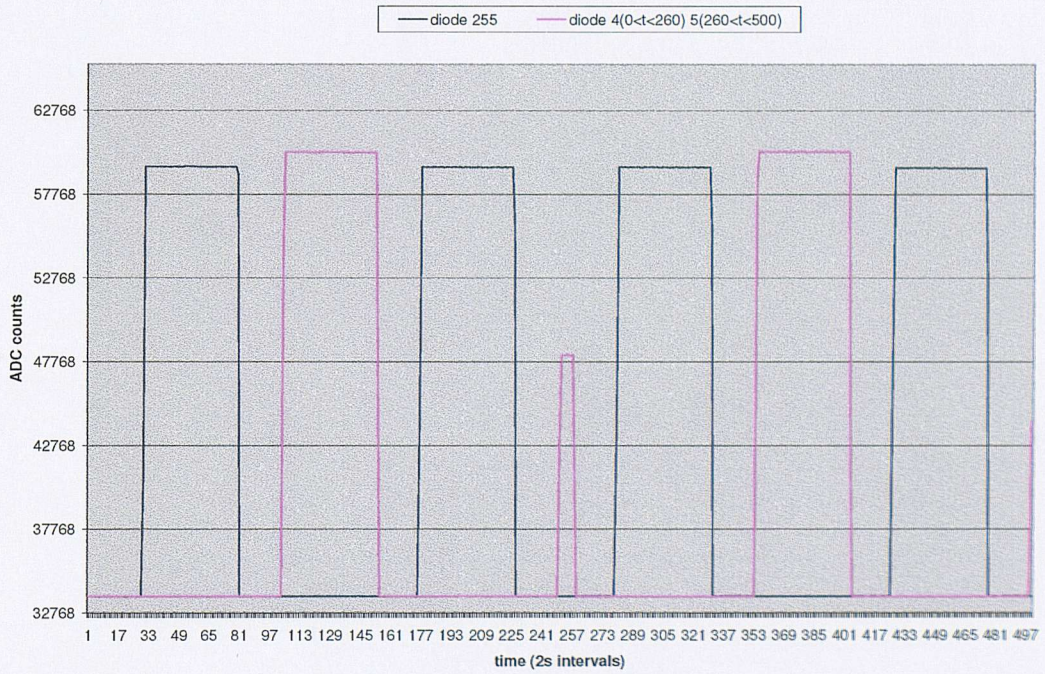


Figure 4.4.11: Data from diode by diode characterisation of PDA, x-axis is time y-axis is raw ADC output. The peak in the middle of the spectrum is a product of the way in which the data were obtained from the original input file, corresponding to light falling on diodes near diode 255 (hence the lower signal height).

Thus each diode data set may be normalised with the dark current data around it and then divided by the similarly normalised data for the diode 255 data sets immediately around it in time. This leads directly to a vector of the pixel-to-pixel responses relative to diode 255, the data for which are shown in Figure 4.4.12.

Diode 255 has a value of 1.001 in the above Figure 4.4.12 while the average difference of each diode relative to the next is around 0.01. Thus while the overall shape of the response varies by around 10% the pixel-to-pixel error is only of order 1%. The fact that diode 255 relative to itself is off 1 by nearly 0.1% is due to the overlap of light on diodes 254 and 256 when data were being taken for diode 255. The two peaks at the ends of the vector (diodes 61 and 490) account for small systematic variations seen in all data sets (for example figure 3.4.5). Use of this vector in its raw form would therefore allow the response of each diode to be known to around 0.1% and thus the SNR changes from the predicted value of 1/150 above to around 5, which is comparable to the SNR of the raw ADC output.

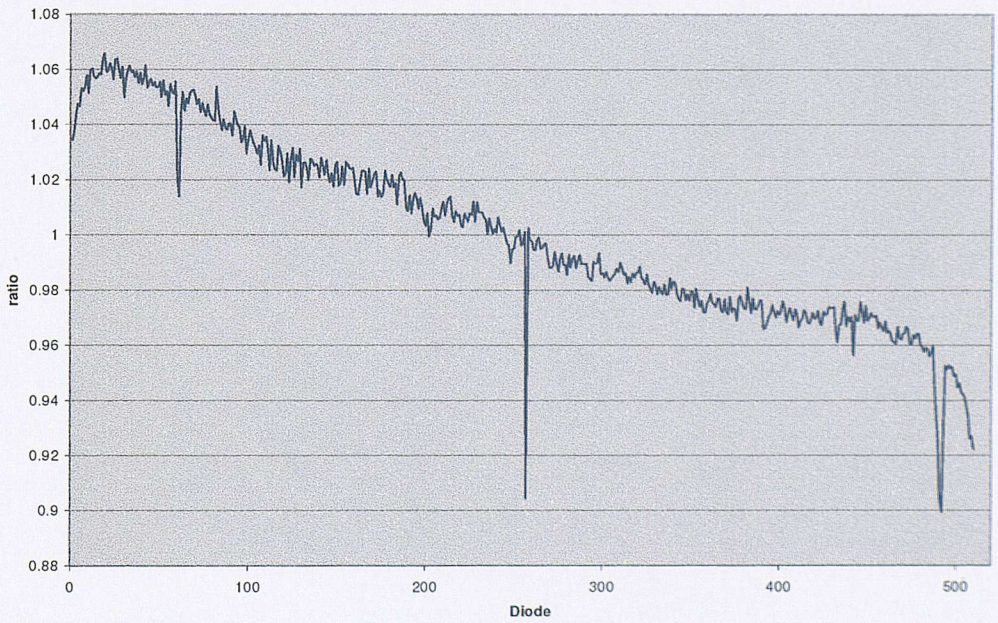


Figure 4.4.12: Diode-to-diode response relative to diode 255

It is worth pointing out here that the diode calibration data were taken with a 16 bit ADC (real accuracy approx 14 bit, see section 4.3) at saturation in differential mode (half full saturation) which from the discussion above implies a SNR in the data taken of around 7 from Figure 4.3.1. While 25 data points were used for each diode (and

twice that for each measure of diode 255) only 10 dark current data points were used (5 on each side of each diode data set) thus the limiting error is the dark current data of which only 10 points were used. Theoretically this therefore lowers the statistical error by a factor of root ten to a SNR of around 3 which is approximately equal to the observed SNR of 5.

When applied to real data the above vector yields no increase in accuracy. This is due to the fact that the diode-to-diode response difference is dependent on the angle of light hitting the diodes as well as the light intensity [49]. Thus in order to account for the diode-to-diode response variations, a measurement must be made using the same optical set-up that will be used to take data. This is outlined in the following section.

4.4.5 Background fit calibration of diode-to-diode response variation

A useable diode-to-diode response calibration may be obtained for any optical system which can be calibrated without any absorbers present in the beam path. The laboratory set-up of figure 4.1.2 is such a system as the absorbers are placed in an absorption cell present in the open path. The light throughput for the system may take any shape within the zero to five volt output of the PDA depending on the position of the lamp and spectrometer slit. Thus the diode-to-diode response must be calibrated for any intensity within this range. As each diode is manufactured in the same way one would expect the response to any given input intensity to scale with the magnitude of the input light for each diode and to test this, the response was calculated for a number of input intensities.

In order to calculate the response for such a DUVAS system, the system (figure 4.1.2) was run with air in the open path over an extended period for each of five intensity inputs covering the full useable range of the ADC. Each data run was of 500 scans so as to allow averaging of the final output to reduce the effect of the random electrical noise from the PDA. The system was left on for a period of six hours prior to the start of data taking so as to allow the warm-up period to finish and a steady state to be reached. Data from each input intensity is then scaled such that saturation input

corresponds to unity output and is then fitted with the moving window polynomial fit described in chapter two (see section 2.3.1). Each diode is further divided by its fitted value to obtain a difference relative to a theoretically smooth background. Thus once an average response has been calculated any data set may be divided by this calibration response to obtain a signal that will be on top of the smooth background. The DUVAS method may then be applied to this calibrated signal as normal. The averaged input signals together with their fitted backgrounds are shown in figure 4.4.13.

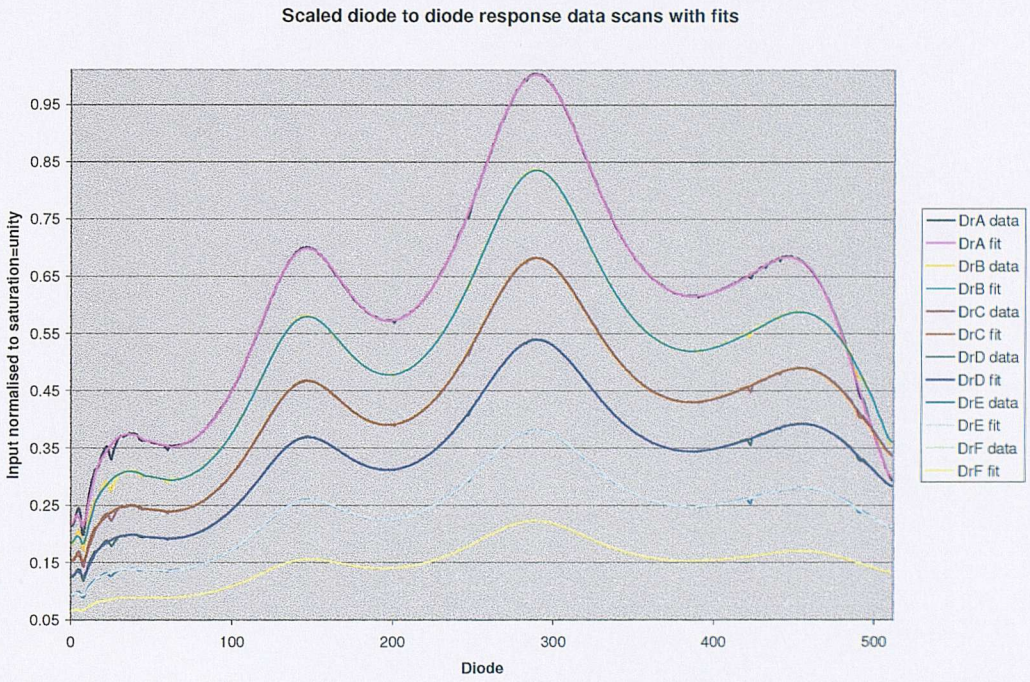


Figure 4.4.13: Calibration data for diode-to-diode response at varying intensity. Dr stands for Diode Response. A, B, C etc is an index of each individual signal where A has the highest magnitude and F the lowest. "Data" are the raw data and the "fit" curves are the fit to this data. Thus 12 curves are plotted on the same axes, 6 each of data and fit. The increase between each data set is not equal, the aim being to obtain sample data sets covering the majority of the ADC range.

The highest input shown in figure 4.4.13 overlaps with the other inputs at high diode number due to the fact that the spectrometer was twisted slightly whilst moving the slit forward to increase the signal on the PDA. The use of background fitting means that data from this set are still useful as it is not the absolute value of any given diode that is used for the calibration but rather the ratio of this value with the corresponding fitted value. The ratios do vary slightly over the input signal range, thus only the four input

signals within the expected 0.1-0.9 of full range were used to calculate the diode-to-diode response vector. These ratios are shown in figure 4.4.14.

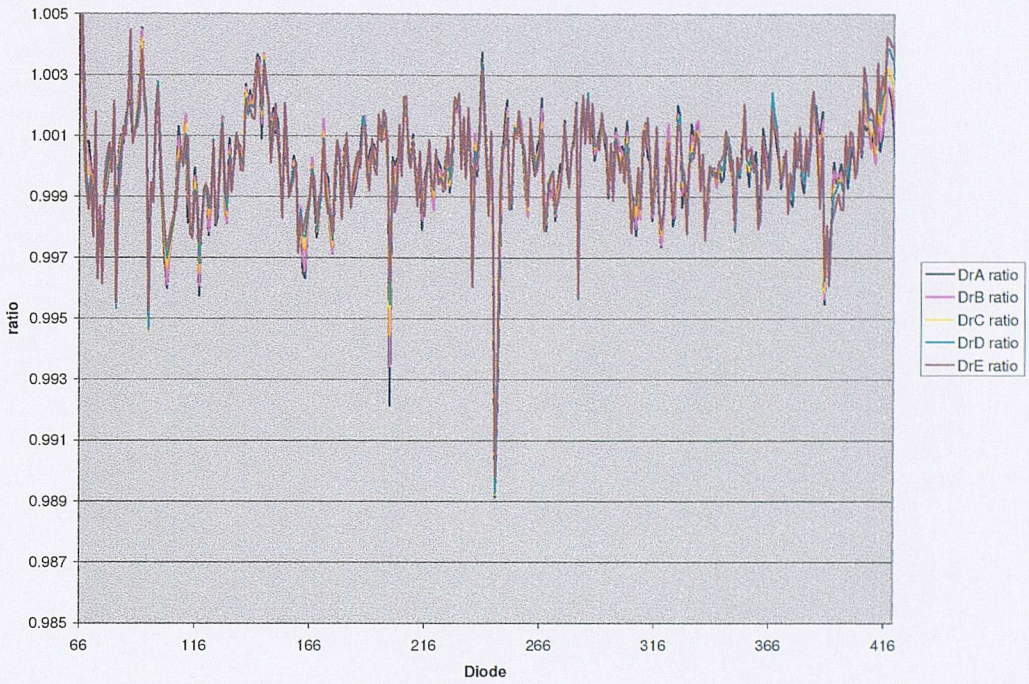


Figure 4.4.14: Ratios of input data with fitted background for signals in expected range (end diodes not shown due to low photon count/background drift)

The average of the signals shown in figure 4.4.14 are the diode-to-diode response data for the laboratory system shown in figure 4.1.2. In order to test this, calibration data were taken and the differentials were calculated from equation 2.1.6 with and without application of the vector. The result is illustrated in figure 4.4.15 where "data-differential" is the differential without any diode-to-diode response calibration whilst "corrected-differential" is the differential after the response calibration has been applied. Both differentials are from 500 scan averaged data sets.

Figure 4.4.15 shows the successful application of the diode-to-diode response calibration as the uncorrected differential has a standard deviation from 0 of 0.021 whilst the corrected differential has a standard deviation from 0 of 0.004, a factor of five times lower. Thus calibration of the diode-to-diode response variations reduces the differentials obtained by half an order of magnitude. This becomes significant when

analysing data from outside as the photon count is more limited than in the laboratory (longer path) and thus the diode-to-diode response becomes a limiting effect.

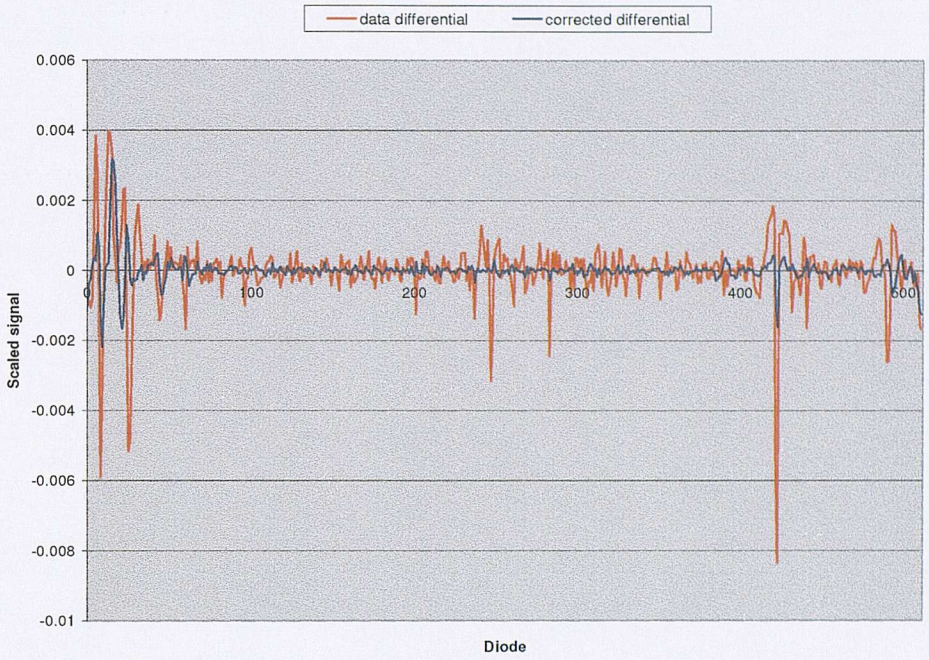


Figure 4.4.15: Differential with and without diode response calibration

From equation 4.3.2 the differentials shown in figure 4.4.15 correspond to a SNR of order 1, thus unless better calibration can be obtained one would have to look primarily at species with a larger absorption than NO. From figure 2.3.6 both benzene and sulphur dioxide have larger absorption features which would have, at 10ppb over 15m path, a SNR relative to the differentials shown in figure 4.4.15 of 5 and 10 respectively. Instrument line shape, wavelength and quantitative calibration was thus done with the laboratory set-up using both benzene and sulphur dioxide with this diode response calibration, the results of which are discussed in the following chapters.

4.5 Summary

The calibration of the subsystems outlined in this chapter and their effect on the system as a whole are summarised in the following table 4.5.2.

Component		Section Reference	Quoted		Measured	
Type	**Details**		**Spec**	**SNR***	**Spec**	**SNR***
ADC	12 bit	2.3	1 in 2^{12}	1/2	1 in 2^{12}	1/2
	16 bit	4.4.1 and 4.4.2	1 in 2^{16}	7	1.6 in 2^{16}	7
	½ Range	4.3	1 in 2^{15}	4	1.6 in 2^{15}	4
PDA	Diode-diode response	4.4	3% of full range	1/150	See figure 4.4.4 (0.1%)	1/5
	Corrected diode-diode response	4.4.5	See above	1/150	See figure 4.4.15 (0.1 of dark current)	1/2
	Dark current	4.4.1	0.4-1.2% of full range	1/20	0.4-0.9% of full range	1/20
System warm-up		4.4.2	na	na	0.5 ADC counts on 5 scans	7

* SNR relative to 10ppb NO over 15m path

Table 4.5.2: Summary of sub-system contributions to signal to noise

The diode-to-diode response limits the resolution of the system to species with absorbance in the UV like NO such that a SNR of one corresponds to a signal of around 50 ppb over 15 m. To achieve the same SNR the dark current noise would have to be limited by averaging over around 10 scans. This amount of NO is in fact larger than that generally found outside (see section 1.2.2 and table 1.3.1) during most of the year, thus it was decided to target SO_2 owing to its larger absorption in the GUSTO range. Concentrations outside during most of the year are of the order of 5ppb see section 1.2.1 and table 1.3.1), which would give a SNR of order 2 given the above characteristics of the system.

Chapter 5

Data, Methods and Results

5 Introduction

The application of the methods outlined in the theory chapter to retrieve concentration values from the data taken by the instrument need to be both verified and applied in a wider context. The method chosen by which to calibrate the full data retrieval method involved using an absorption cell, into which a known concentration of pollutant could be inserted. The instrument function, wavelength, multi-species capability and quantitative retrieval may then all be tested. The early part of this chapter deals with these questions, whilst the latter part is a presentation of actual data taken from outside the lab with corresponding results.

5.1 Absorption Cell

The optical set-up of figure 5.1.1 shows an open path in the laboratory within which sits an absorption cell. The cell is simply an aluminium tube with fused silica windows and

three natural rubber septa each covering an input tube set at intervals along the cell length. This is illustrated in the figure 5.1.1 and accompanying table 5.1.1

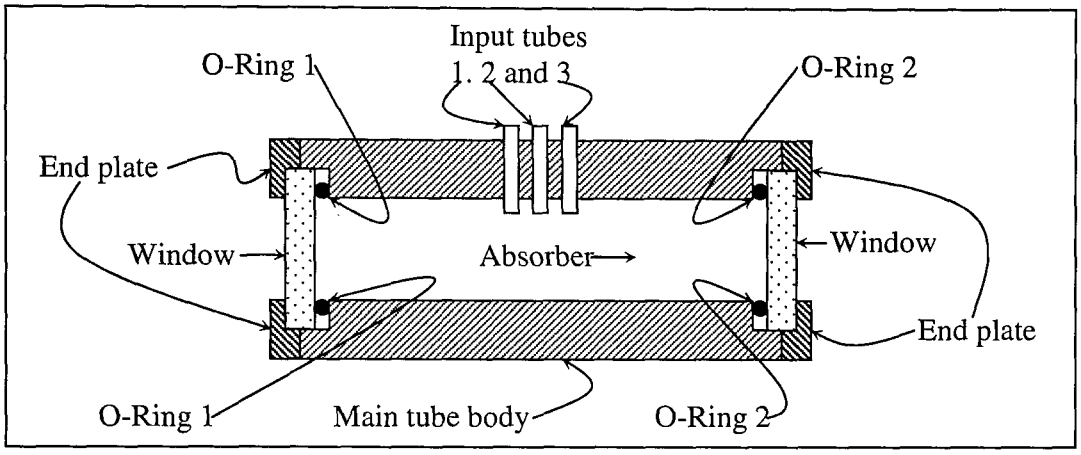


Figure 5.1.1: Absorption cell cross section

Component	Description	Specifications (units are mm)
Main tube body	Material: Aluminium Source: Custom machined	Length = 425 Diameter = 150 Thickness = 5
End plates	Material: Aluminium Source: Custom machined	Diameter = 223 Thickness = 12
Input tubes	Material: Aluminium Source: In house Set in centre of cell with 3cm spacing	Length = 50 Diameter = 10 Thickness = 2
O-Rings	Material: Viton Source: Bought to order	Diameter = 136.5 Cross-sectional diameter = 3.2
Windows	Material: Fused silica Source: Bought to order Light diameter = 123.5mm	Diameter = 152 Thickness = 10 Flat to 400nm
Septa covering input tubes	Material: Natural rubber Source: Merck	Length = 12 Thickness = 2

Table 5.1.1: Specification of absorption cell

The cell above in figure 5.1.1 together with the set-ups from figures 4.1.1 and 4.1.2 were used to do quantitative work with actual absorbing gases in the lab. Both sulphur dioxide and benzene were obtained [65].

5.1.1 Instrument Function

In order to obtain the instrument function for the GUSTO optics shown in figure 4.1.1 a mercury line spectrum was put through the entire system. The mechanism for doing this is to replace the 'arc image 1' of figure 5.1.1 with the pinhole of the penray mercury lamp. This was duly done and data were taken over a two hour period. As the mercury line is from a single transition a Gaussian may be fitted to the mercury data to obtain the width of the instrument function. Thus the instrument function is a unit area Gaussian of the same width, which may then be convolved with the synthetic data to obtain the same signals that one should actually 'see' with the real system.

The histogram for the two hours of mercury data (3600 scans) is shown below with the Gaussian fit in figure 5.1.2 together with the fitted parameters which are shown in the accompanying table 5.1.2. A small ghost (presumably due to an errant reflection) was removed from the main peak.

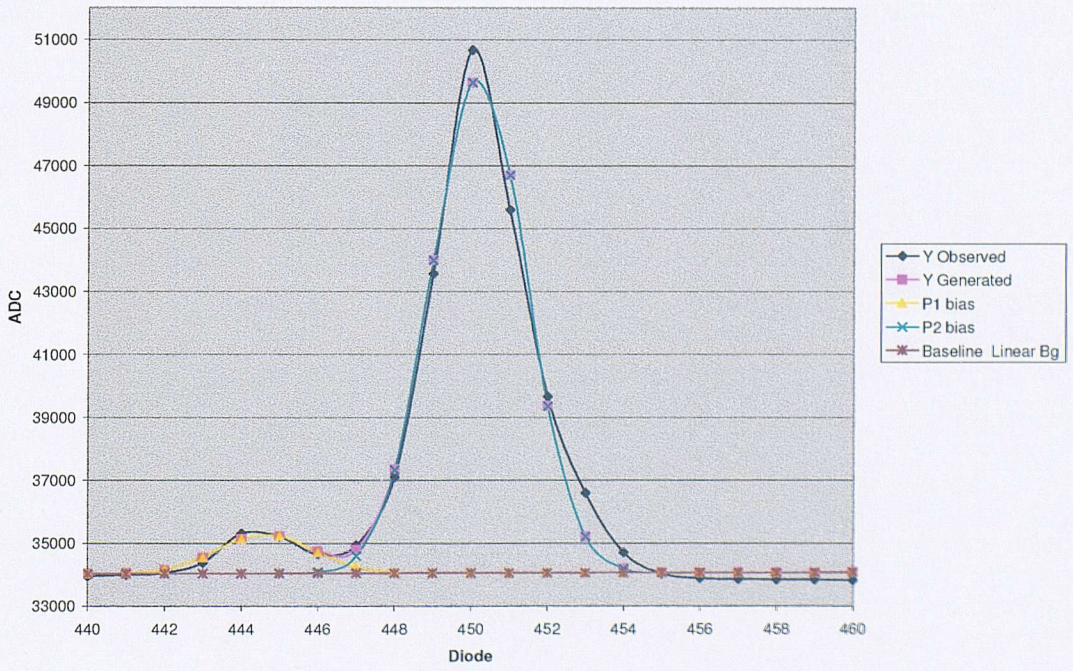


Figure 5.1.2: Histogram of mercury line data with Gaussian fit

Fitted Parameters			
r^2 Coef Det	DF Adj r^2	Fit Std Err	F-value
0.98772705	0.98111854	607.181584	187.786663

Peak Type	Values		
	Amplitude	Centre	FWHM
1 Gauss Amp	1244.83691	444.619765	2.90275525
2 Gauss Amp	15765.8696	450.182516	2.90275525

Details of Fit				
Set Convergence	State	Iterations	Minimization	Extent
1E-6	Converged	7	Least Squares	1/1

Table 5.1.2: Fitted parameters for figure 5.1.2, Produced with PeakFit from systat

The width of the Gaussian fit from figure 5.1.2 is 2.9 diode numbers, thus the corresponding instrument function for the same optical set-up is a Gaussian with zero mean and width 2.9. Using the usual form from equation 5.1.1 below, this function is plotted below in figure 5.1.3

$$f(x;\mu, \ ) = \frac{1}{\sqrt{2}} e^{-\left(\frac{(x- \ )^2}{2 \ ^2}\right)} \qquad (5.1.1)$$

where the mean $= \mu$ *and the width* $= \sigma$

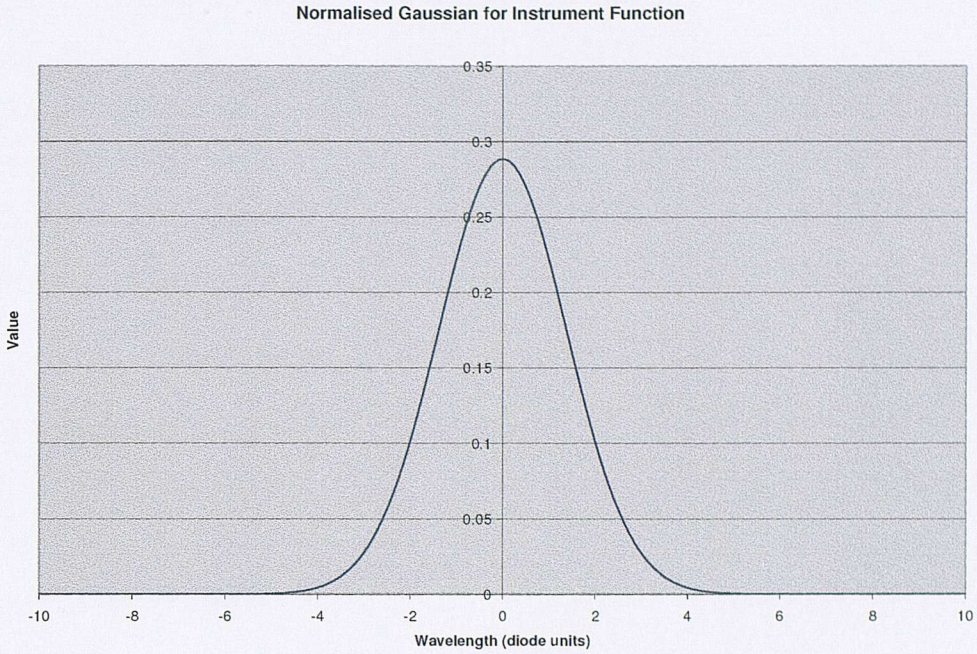


Figure 5.1.3: Instrument function for optical set-up of figure 5.1.1

The instrument function of figure 5.1.3 may thus be directly convolved with the differentials obtained from the synthetic absorption cross sections shown in figure 2.3.6 in order to compare the expected spectra with those actually measured. It is computationally faster to convolve once with the cross sections as opposed to de-convolving in real time with the measured data. This will be possible given a faster data acquisition and analysis computer system.

5.1.2 Method of Data Retrieval

In order to calculate the concentration of an absorbing gas using the DUVAS method (equation 2.1.6) one must fit the differential obtained from a given data set with the expected differential (dependent on concentration). Thus there is a heavy emphasis on accurately calculating the expected differential, and the way in which this is fitted to the measured signal.

The measured signal has an unknown initial intensity. Dependence on this is removed in going through the DUVAS method and one is left with a differential. One may recreate this mathematically for a synthetic spectrum by starting with the absorption cross sections for the species under study and calculating the expected absorption signal for any background (or no background) and then applying the instrument function and the same DUVAS method. This process is summarised in the following figure 5.1.4

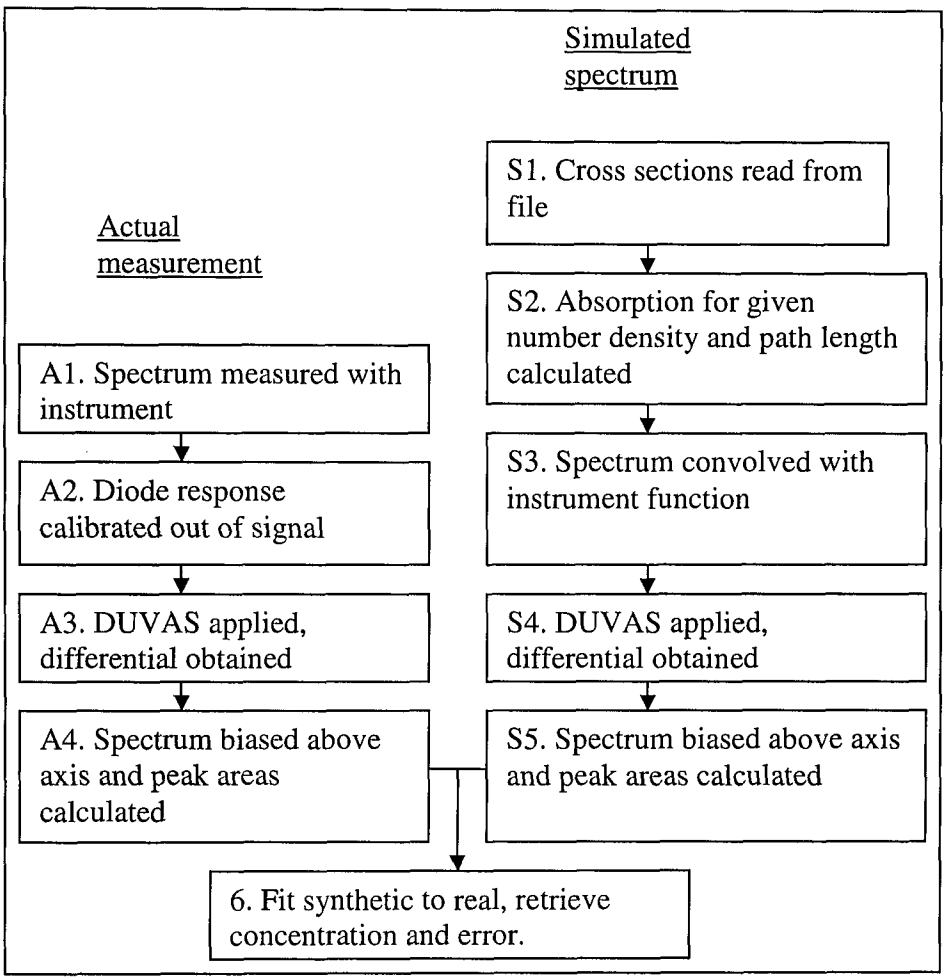


Figure 5.1.4: Method of data retrieval.

The penultimate step in the above figure 5.1.4 is to recognise that the area under an absorption spectrum is the conserved quantity and thus it is the area of the absorption peaks shown by the differential spectrum that one must compare between the measured and synthetic differentials. Clearly the area between a DUVAS differential and the x-axis is zero, and the amount that one biases the signal above the axis will greatly affect

the retrieved area. Thus one must ensure that the peak shapes are the same (achieved by convolution with the instrument function) and each signal must be treated exactly the same from the point of application of the DUVAS method.

Figure 5.1.5 illustrates the A1 in figure 5.1.4. It shows the original raw data from a measurement done with the absorption cell on 29th of June 2002 containing SO_2. A balloon of known volume was filled with a known quantity of pure gas (see later discussion) which was left to diffuse, then another known quantity of diffused gas was injected into the cell.

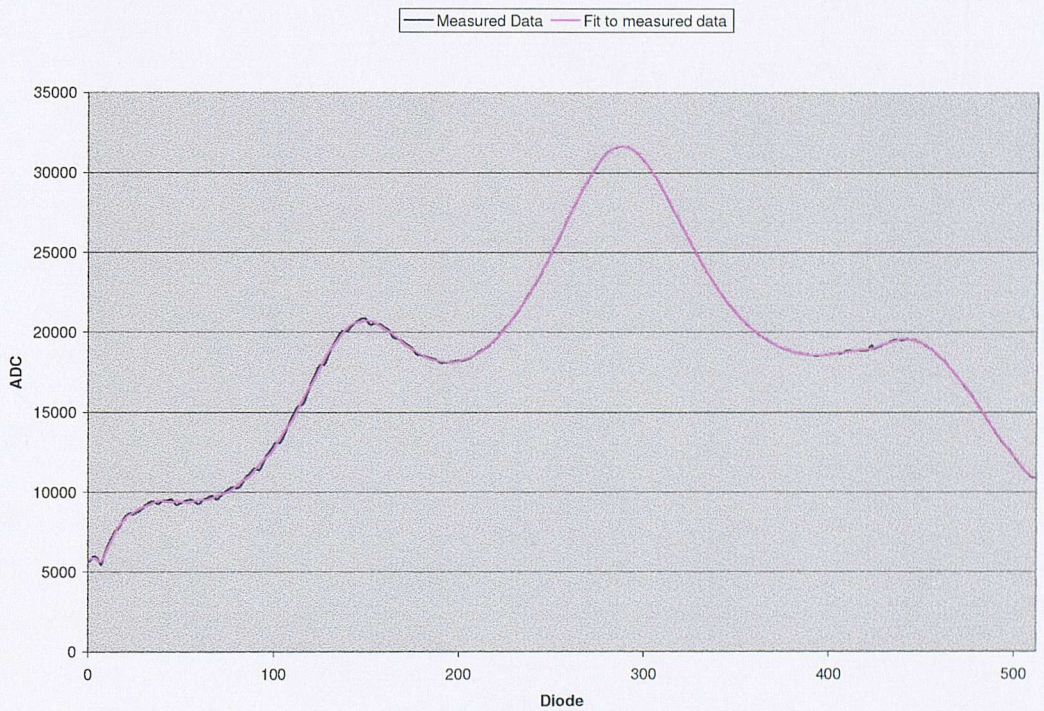


Figure 5.1.5: Corrected ADC output from SO_2 experiment 29/07/02 shown with fit after diode response calibration, 200-260nm x-axis scale.

The SO_2 absorption peaks between diodes 40 and 200 (206-230nm) are visible in figure 5.1.5 (compare with fig 4.4.5). The experiment was run for 500 scans (approx 16 minutes) and the data shown in figure is the average of all scans.

The effect of the diode response calibration (step A2 in fig 5.1.4) on data like those shown in figure 5.1.5 is most easy illustrated in a graph of the differentials such as figure 4.4.15, which has no SO_2 present. The effect of SO_2 on the shape of the

differentials is illustrated in figure 5.1.6, which is the final differential spectrum after diode response calibration (step A3 in figure 5.1.4) for the same data as shown in figure 5.1.5. The fit used to obtain figure 5.1.6 is shown on figure 5.1.5.

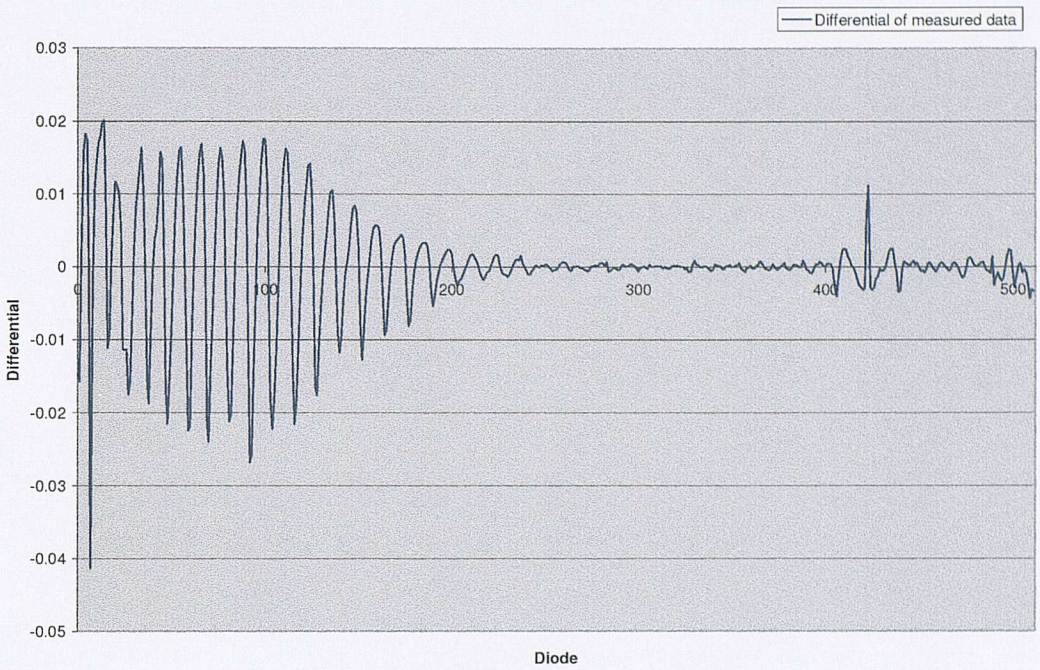


Figure 5.1.6: Differential for data shown in figure 4.1.5

The magnitude of the SO_2 peaks in figure 5.1.6 is considerably larger than the background (compare with figure 4.3.15) owing to the fact that a large amount of SO_2 was added (of order 200ppb) to allow for distinct peaks to be seen. However, the SNR is very large, exceeding 200.

Step S1 in fig 5.1.4 is illustrated by the below figure 5.1.7 showing sulphur dioxide cross sections [53] in cm^2 over the GUSTO range

The actual absorbance of a unity light input to a 15m open path system given the cross sections from figure 5.1.7 may be modelled with equation 2.1.6, retaining the exponential and setting I_p to unity yields the intensity curve shown in figure 5.1.8. Also illustrated in figure 5.1.8 is the convolution of this intensity spectrum with the instrument function shown in fig 5.1.3.

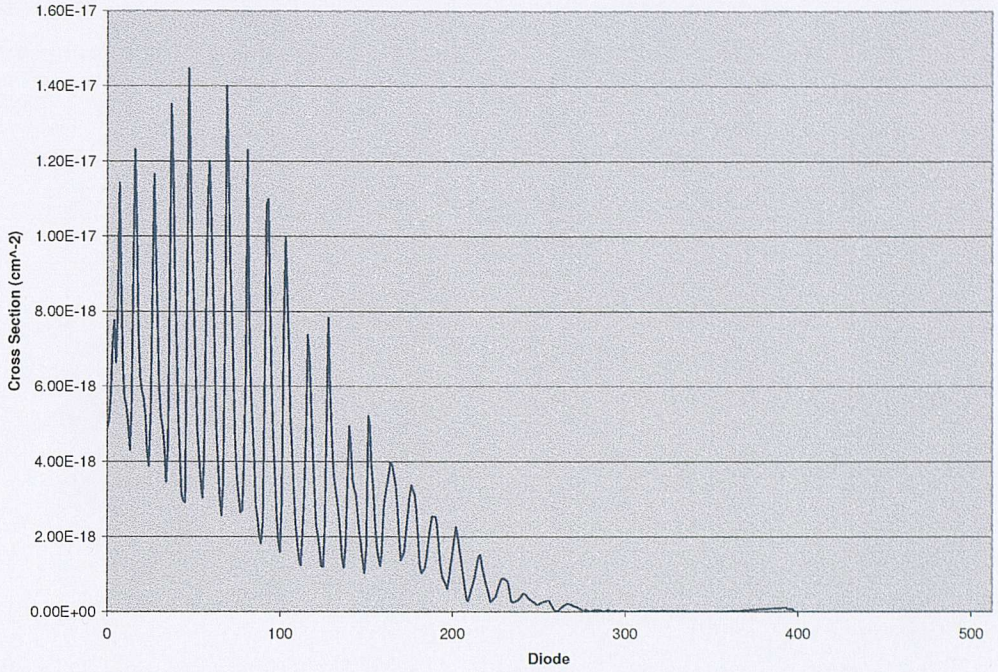


Figure 5.1.7: Absorption cross sections for sulphur dioxide over GUSTO range

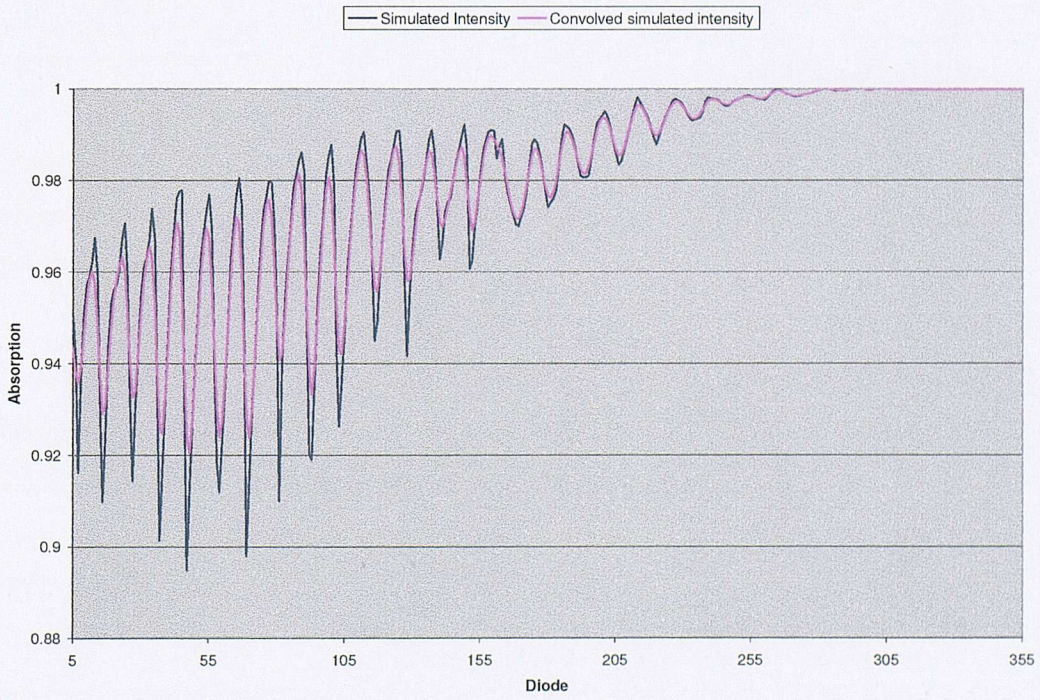


Figure 5.1.8: Absorption intensity over 0.5m open path for 6ppm SO_2 shown before/after convolution with instrument function

Step S4 for the synthetic data is illustrated in the following figure 5.1.9; the equivalent differential from figure 5.1.6 is also shown on the same plot for comparison.

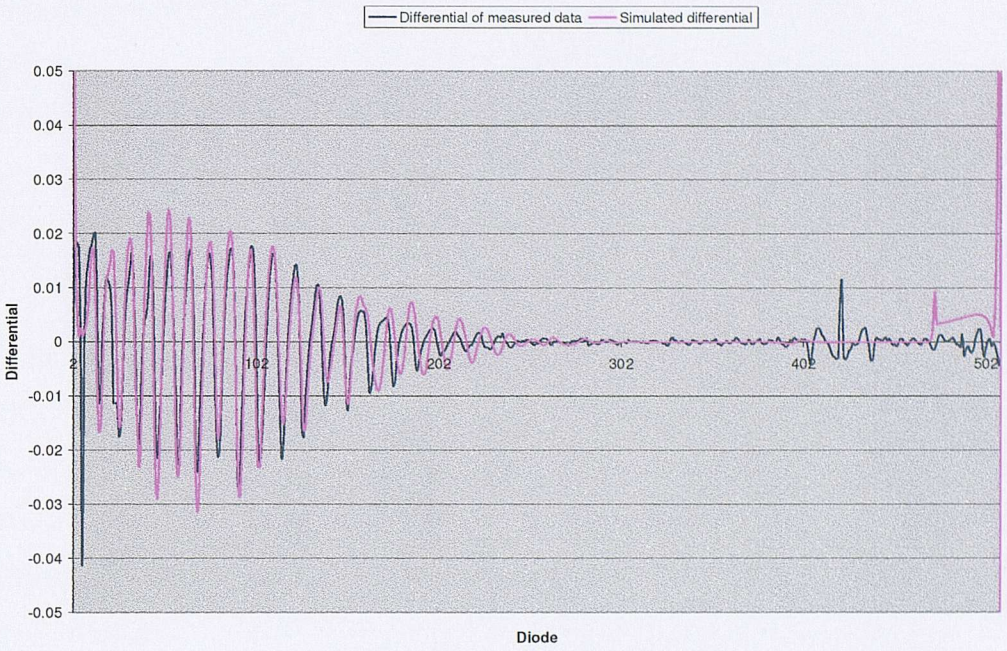


Figure 5.1.9: Synthetic differential for 6ppm SO_2 signal over 0.5m shown over differential for data from 29/07/02. RealDiff is the differential of the measured data, synthDiff is the differential of the simulated data NB. plot is prior to wavelength calibration.

Figure 5.1.8 illustrates steps S2 and S3 from figure 5.1.4. The area under each curve (between diodes 10 to 250) is the same (to around 0.04%) which is as one would expect for a convolution of this type [66].

The differentials shown in figure 5.1.9 show good agreement between around the 10^{th} diode up to the 160^{th} diode, this may more easily be seen by referring to the same plot after the wavelength calibration has been applied, figure 5.1.12. There is a background under part of the signal which is not compensated for, giving rise to the changing difference in heights of the spectra. However, there is excellent agreement in the middle of this range. The lower 10 diodes on the array have almost no light input compared with the rest (see fig 5.1.5) and thus the noise plays an increasingly apparent role. There is a wavelength shift corresponding to non ideal behaviour of the dispersion at around diode 158. The features in the real data differential above diode 400 are due

to the diode array as these are also visible in the raw data set shown in figure 4.4.5 where there were no added absorbers present.

The wavelength scale and dispersion for the array is discussed in the next section as there is a clear inconsistency with the default values (wavelength range divided by number of diodes) shown in figure 5.1.9.

5.1.3 Wavelength Calibration

The monochromator is controlled by a program running on the same laptop that runs the ADC data acquisition hardware and is connected by an RS232 cable. The standard operation of this form of device [63] is to set the grating position and unplug power from the spectrometer unit. This prevents the spectrometer from resetting the grating position. The same grating position can be used for multiple experiments. There is however an error on the initial position of the grating [57] and thus the wavelength in the centre of the PDA is not known to more than around 0.1nm.

Also the dispersion across the PDA is not uniform [57] and thus the entire PDA must be checked to calibrate the wavelength on any given diode. The optical set-up inside the spectrometer is illustrated in figure 3.1.1 [67].

The actual set-up inside the GUSTO unit is an example of the Ebert-Fastie type, though the same theory applies to both this and the more general Czerny-Turner as the former is a special case of the latter with the same mirror being used as both the collimator and camera.

The linear dispersion (measured in nm/mm), may be written in the usual form [67]:

$$p = \frac{d\cos(\beta)}{mr} \qquad (5.1.2)$$

where p = the linear dispersion, d = number of lines per mm on grating, β = angle between light coming off grating and the grating normal, m = the order of the light imaged at the exit slit and r = focal length of the spectrometer.

For our system d=1/(1200) per mm, m=1 and r=110mm while beta is 52° for a central wavelength of 230nm and varies over the length of the array by ±1.3°.

Hence the wavelength boundary between any given diode pair may be calculated by the following method.

$$\begin{aligned} \lambda(diode[i], diode[i-1]) &= 230 - l \times \sum_{n=0}^{255-i} p\,|_{255-n} \qquad for(0 < i < 255) \\ \lambda(diode[i], diode[i+1]) &= 230 + l \times \sum_{n=255}^{i} p\,|_{n} \qquad for(255 < i < 511) \end{aligned} \tag{5.1.3}$$

Where *diode[i]* = diode i, *λ(diode[i],diode[i+1])* = wavelength between diode i and i+1, l is the width of a single diode in mm and $p|_n$ is the linear dispersion for the array measured at diode n as given by 5.1.2.

The difference between the real wavelength between any given diode pair, given by equation 5.1.3, and the simple approximation of a linear fit to equation 5.1.3 is shown below in figure 5.1.11 as a fraction of one SO_2 peak (10 diodes from fig 5.1.9). Thus a value of 100% for a given diode indicates that the application of the approximation for the dispersion on the PDA results in a difference at this diode equal to 1 complete period of the peaks in the SO_2 absorption cross section.

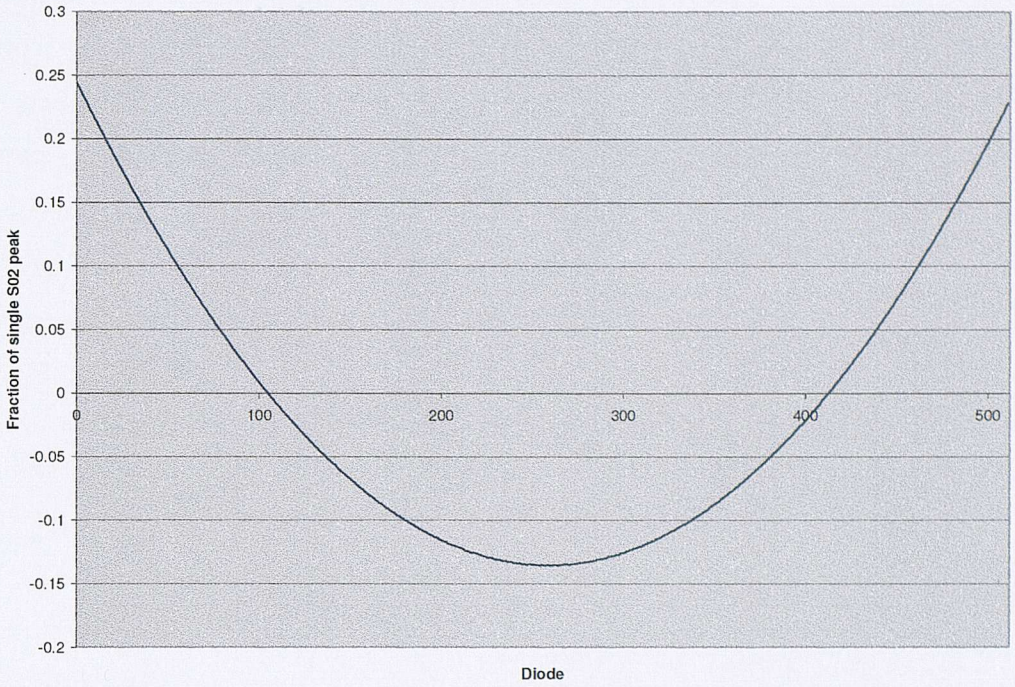


Figure 5.1.11: Difference between ideal and linear approximation for dispersion as fraction of single SO_2 peak for each diode in array

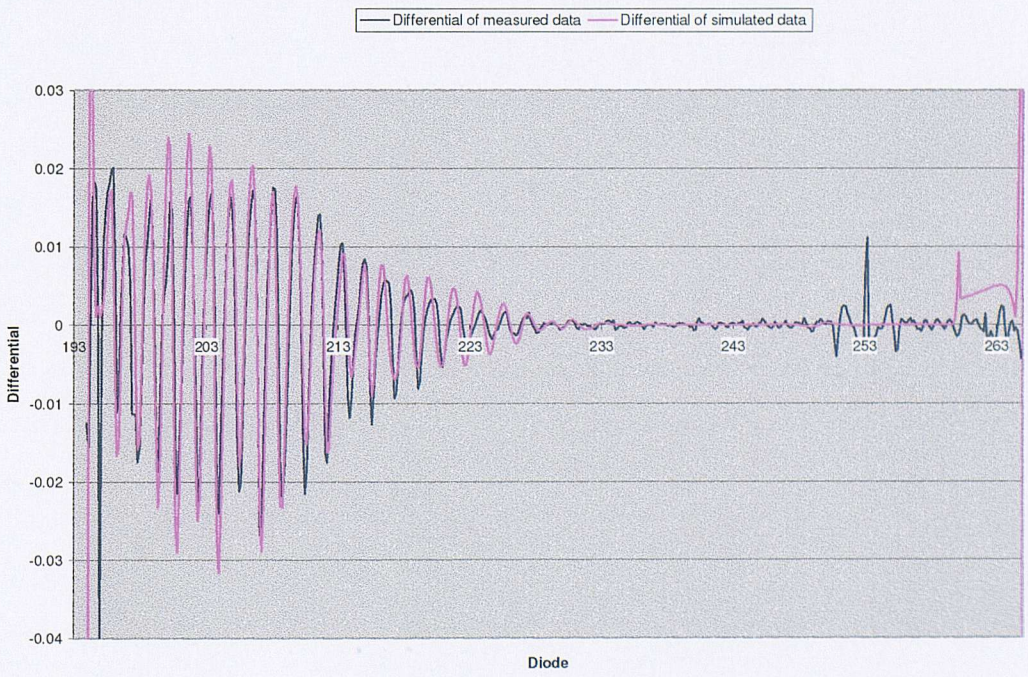


Figure 5.1.12: Same plot as figure 5.1.9 but after wavelength calibration has been applied

The errors present in any given spectrum on the wavelength axis relative to the features under study for SO_2 can be seen from figure 5.1.1 to be of the order of 10%. Thus the more accurate method of equation 5.1.3 was used to calculate the wavelength along the array. This effect is illustrated in figure 5.1.12 (compare to the default wavelength calibration shown in fig 5.1.9). The agreement between the measured and simulated differentials is better than one diode over the full array with the more accurate dispersion calculation.

5.1.4 Quantitative Data Retrieval

If a given differential signal is plotted against wavelength in wave number units then, assuming it to be biased such that all the peaks are just above the axis, the area under the peaks will be equal to the area under another differential which has been measured with the same instrument function and corresponds to the same original number density. To implement this method of cross-comparison between differential signals a routine was written which finds the lower minimum (of each distinct peak) of the differential and linearly interpolates between them, thus adding the resultant to the original spectra yields a differential suitably biased above the axis. The area is then calculated assuming trapeziums of width one diode, with the x-axis being converted from nm to cm^{-1}. This then forms a value by which different biased differentials may be compared.

The areas within the window region of diodes 48 to 152 calculated for all original concentrations of Sulphur Dioxide between 1 and 20000 ppb over a 15 m open path using the method of figure 5.1.4 are shown in figure 5.1.13.

For the lower region 0 to 1000 ppb shown in figure 5.1.13 a linear fit with gradient 0.349 is accurate to 0.2 percent. As GUSTO was originally intended to observe urban pollution levels where the government guidelines table 1.3.1 and 1.3.2 [36] require less than 100ppb for the majority of the year, the linear approximation may be used.

The previously shown figures for the absorption cell have all been over the 0.5 m path of the cell. Within the range we are working the scaling is linear, thus the equivalent

signal over 15m path would be for a concentration 30 times lower, i.e. 6ppm over 0.5 m open path will look the same as 180ppb over 15m path.

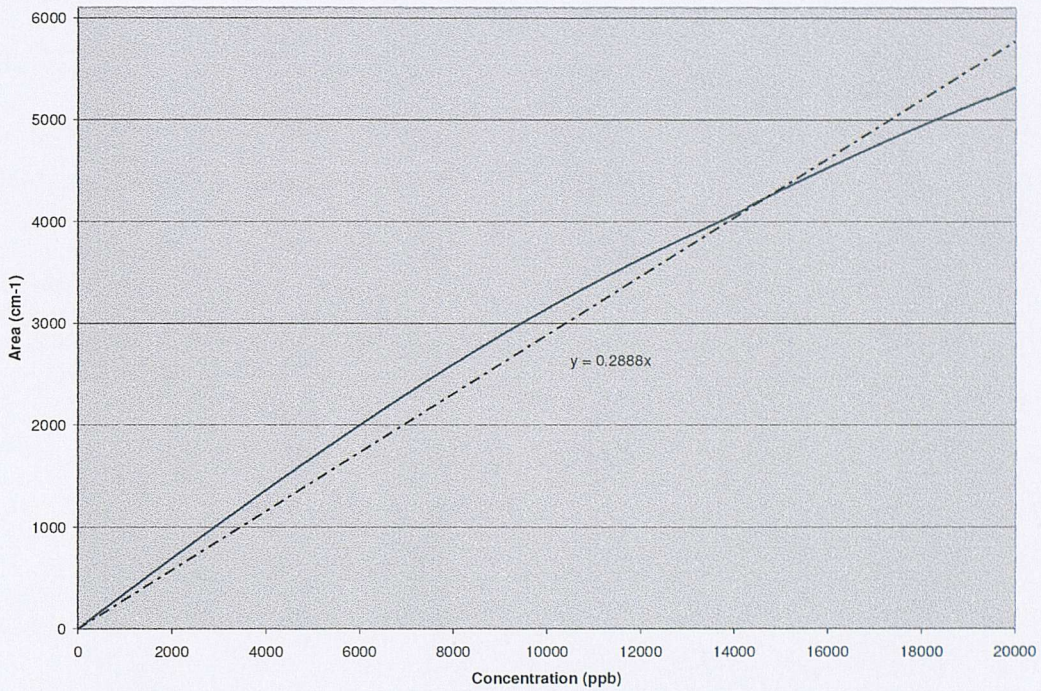


Figure 5.1.13: Area within window region (diodes 48-152) for concentrations of SO_2 between 1 and 20000 ppb, dashed line is linear best fit with equation shown. There is one data point per ppb.

Sulphur dioxide is not the only absorber present within the GUSTO wavelength range (see figure 2.3.6) and thus any absorbance not part of a smooth background will not be removed from the differential signal. Thus one would expect any given peak to have an area corresponding to a slightly different absorber number density. This variation of number densities will not be systematic over the entire window range as SO_2 has unique spectral structure. Thus the standard deviation of these variations from the mean retrieved number density may be taken as a measure of the error on any one signal. However within the chosen window region of diodes 48 to 152 there are nine distinct peaks and as the mean scales with the root of the number of statistics used we would expect the upper and lower limits on the retrieved concentrations to be given by the following equation 5.1.4:

$$\sigma(n) \approx \frac{1}{\sqrt{N}} \times \left\{ \frac{\sum_{peaks} abs\left(n[A_{measured}] - n[A'_{synthetic}]\right)}{N} \right\} \qquad (5.1.4)$$

Where $\sigma(n)$ = error on retrieved number density of n, $n(A)$ = *number density* for area of a single peak, $n(A')$ =*number density* for best fit area of single peak and N= number of distinct peaks within window region.

The implementation of both the bias routine to the differentials and the area/error calculation are illustrated in figure 5.1.14.

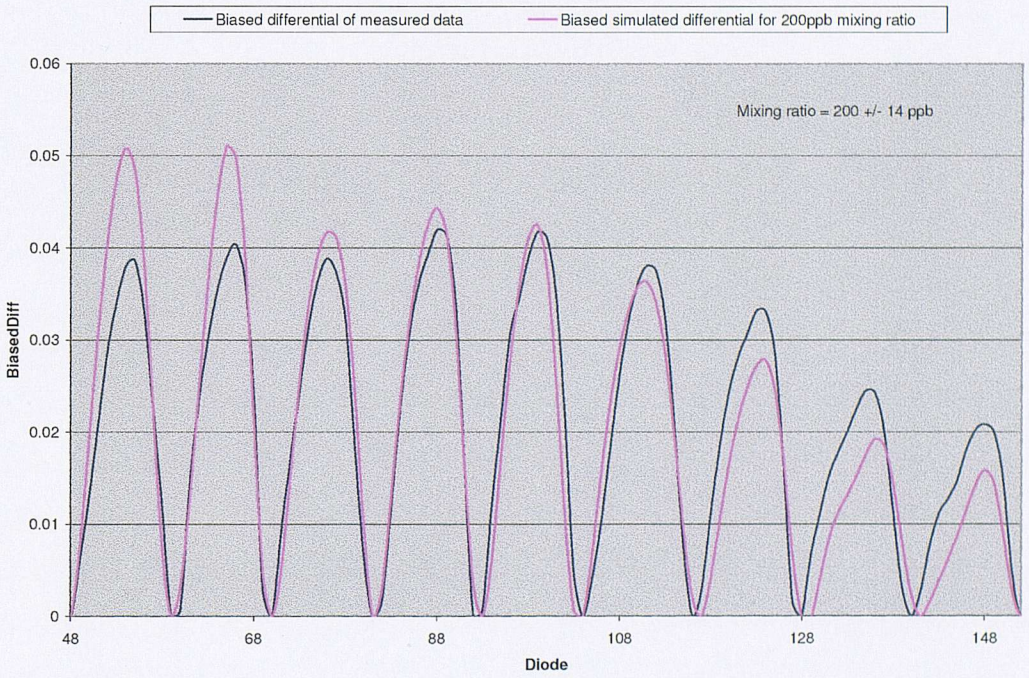


Figure 5.1.14: Biased differentials for SO_2 data from 29/07/02, error calculated from equation 5.1.4. Shown are both biased differentials for measured and simulated data.

The data shown above in figure 5.1.14 was taken in the cell over a 0.5 m path, however the concentration added was scaled up to give a signal equivalent to that seen over a 15m path such as that for the outside system. The cell has an internal volume of 3500 cm^3 which if filled by a syringe containing pure gas would result in 10 ppb corresponding to an input volume of around 0.035 mm^3, which is virtually impossible to either obtain or contain directly within a normal gastight syringe. However this value is

30 times smaller than the volume needed for a 15 m path length and the gas may first be diffused into a secondary vessel of known volume and thus diffused gas of known concentration may be extracted.

In order to validate the quantitative data retrieval method it was decided that three independent methods to calculate the amount of gas within the absorption cell would be used. These are first, estimation from the amount of gas injected with the syringe, which turns out to be governed in part by the dead-volume of the needle/syringe combination. Secondly by using the real background instead of the DUVAS method, i.e. standard spectroscopy, which is possible in the lab as a real background may be taken. Lastly one can use the full DUVAS method in the laboratory, which if validated, may then be used independently on real data from outside.

The standard spectroscopic method for calculating absorber number densities is to fit synthetic cross section data to measured absorbance data with the background signal still present, i.e. equivalent to DUVAS but using a 'real' background covering the entire peak areas. In order to implement this, real background signals were taken for all the laboratory data immediately before any pollutant was introduced into the cell. Thus using these measured backgrounds will only include contributions to the signal from the pollutant under study, further it has the distinct advantage of automatically compensating for the diode-to-diode response as this will be the same with and without the pollutant present (assuming low concentrations of pollutant). This process is illustrated below for the data taken on 29/07, figure 5.1.15 shows the original background (measured at the same time as the data) with the equivalent data (same as plot 5.1.5 but with the real background and no fit), figure 5.1.16 shows the absorbance (-ln(signal/background)) after signal has been normalised such that zero corresponds to zero input, also shown are the simulated data for 6ppm.

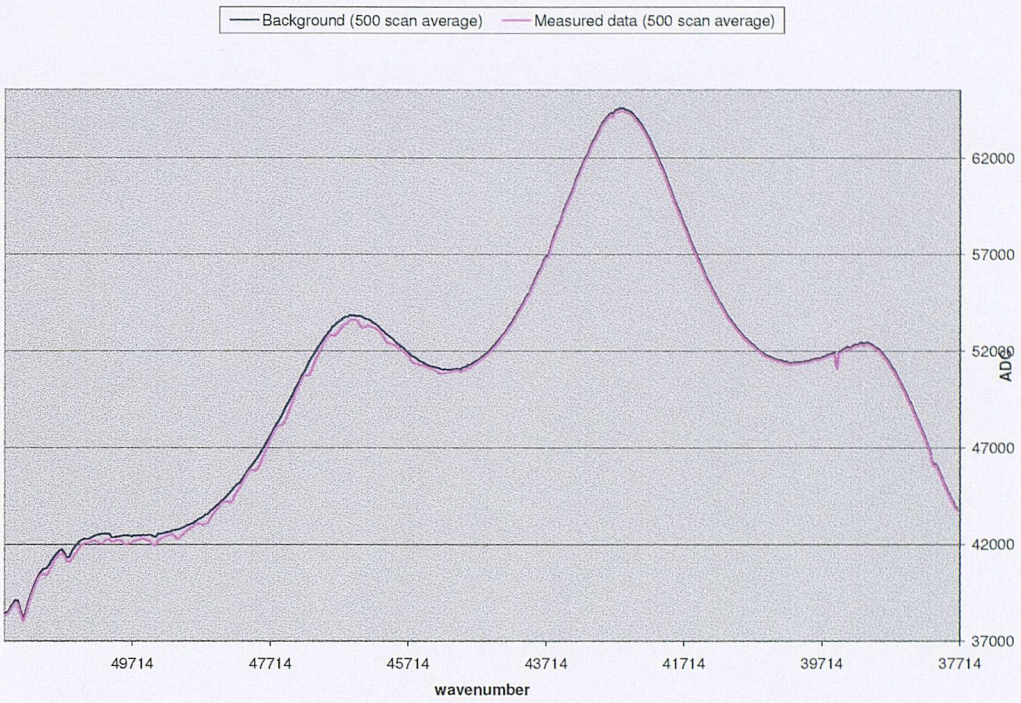


Figure 5.1.15: Measured background with SO_2 data.

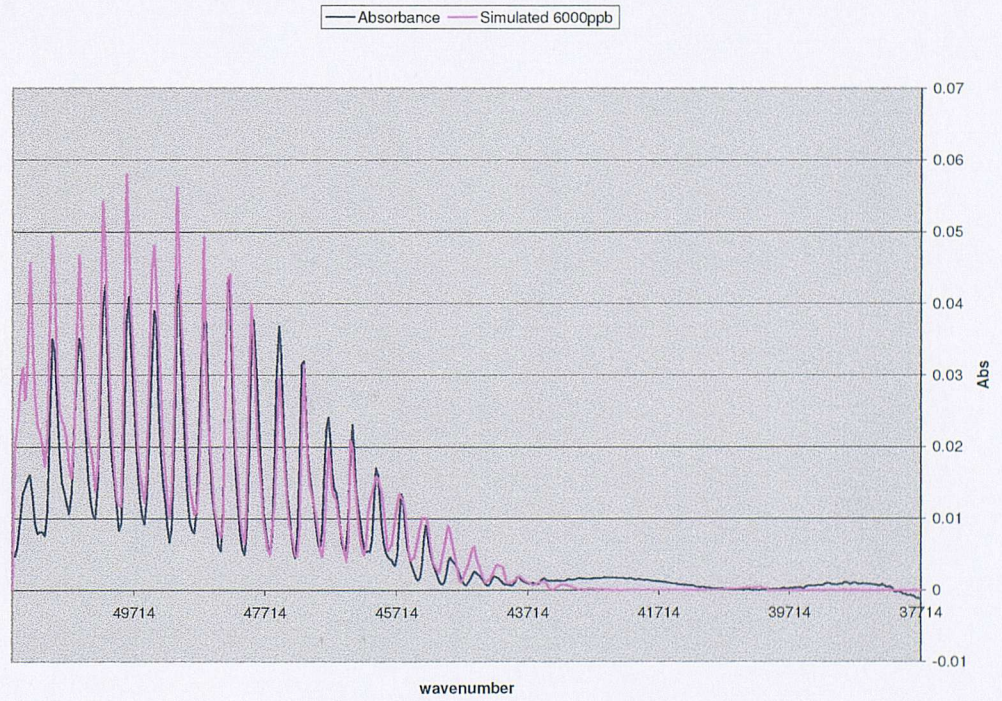


Figure 5.1.16: Absorbance and simulated signal for 6ppm data

The DUVAS technique was also applied to all three data sets taken in the lab; this is illustrated in figure 5.1.17 which is the data from the experiment using DUVAS and area matching for peaks within the chosen region of 47500 to 49000 wavenumbers (diodes 70-116).

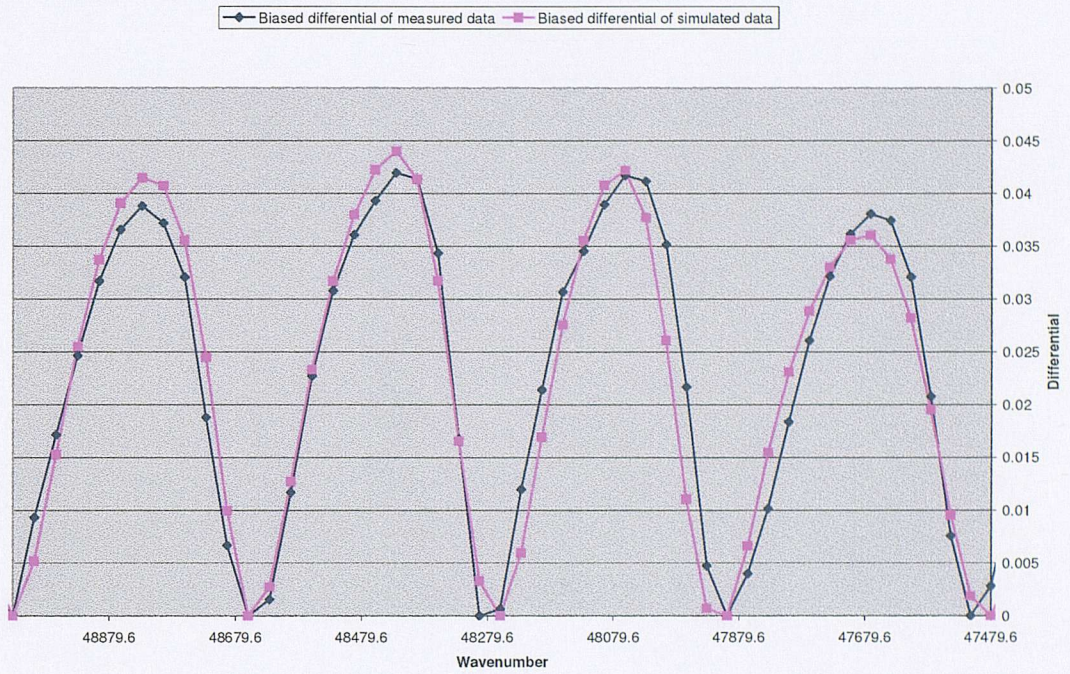


Figure 5.1.17: DUVAS method for data taken 29/07/02

5.2 Results

Application of the methods outlined in the above quantitative retrieval sections yields actual values with corresponding errors to the various experiments with and without the cell. For concentration data retrieved from the laboratory the results will be quoted as mixing ratio (usually ppm); this is for synthetic absorbance and differentials over a path length of 0.5m. For data taken outside, it was decided to use the cell as the signals needed were of order 100 ppb SO_2 varying by tens of ppb (so as to allow a range of signals within the government guidelines to be seen), while the natural signals present were less than 4 ppb [16]. Although the cell was used, the concentrations will be

quoted as 'effective concentrations', which are calculated over the full 15m path. The relationship between these two different methods of presenting the results is that the data for an experiment of concentration x (units) is identical to the data for an experiment of effective concentration (x/30) units as the same physical amount of gas would be present in both cases and the optics/system is identical in both cases.

The implementation of the method of estimating the amount of gas input into the cell by means of the physical volume inserted with the syringe proved to be governed by the dead volume. If an amount of gas x cm^3 is taken out of a container of pure gas with a syringe that has a dead volume of y cm^3 then the volume of pure gas actually input into the next vessel if the full x cm^3 is then ejected from the syringe is given by the following equation 5.2.1

$$actual\ volume\ transfered = \frac{x^2}{(x+y)} \qquad (5.2.1)$$

If a concentration is first injected into a secondary vessel then the final concentration will depend strongly on how much of the remaining pure gas within the syringe dead volume diffuses out of the syringe before it is filled again with dilute gas. In the case of the laboratory cell data one experiment (A) was done with a secondary vessel with the syringe being inserted into a natural rubber block between the injection into the secondary vessel and the taking out of diffuse gas from said vessel, 0.25cm^3 was taken out in each case (volume of secondary vessel = 4200cm^3, volume of cell = 3300cm^3, thus 1cm^3 of pure gas into cell is 300ppm, in this case the dead volume of syringe = 0.1cm^3) thus giving an estimate of 15ppm. The error on this is highly dependent on the accuracy with which the syringe was used, and the dead volume could be calculated, each of which was of order 10% thus the quoted error on this measurement is 20%. In a further experiment (B), also done in this way, the syringe was left open during the 5 minutes interval between filling the secondary vessel and then filling from the secondary vessel, thus the concentration is only known to be less than the maximum value (full dead volume of pure gas) which is 29ppm. The final experiment (C) was a direct injection of pure gas into the cell, 0.05cm^3 was used with a syringe of dead

volume 0.1cm^3, thus giving from equation 5.2.1 a mixing ratio of 5ppm with the same quoted accuracy of ± 20%.

5.2.1 Laboratory Absorption Cell Results

Calculation of the areas under the curves for each analysis method together with combining the differences between each peaks predicted concentration as in equation 5.1.4 for the errors, yields the following results shown in table 5.2.1 for each of the three experiments with each of the three methods.

Experiment	Mixing ratio using non DUVAS method	Mixing ratio with DUVAS method	Pollutant added
A	16.5ppm ± 5%	17.2ppm ± 2%	15ppm ± 20%
B	2.45ppm ± 5%	2.3ppm ± 2%	<29ppm
C	6.1ppm ± 5%	5.9ppm ± 2%	5ppm ± 20%

Table 5.2.1: Summary of laboratory cell measurements

A total of 7 experiments were done with the syringe although only those experiments that can be compared with another method are included in the above table. The remaining experiments were to test the real time ability of the system to link its mixing ratio output to a remote display computer. An example plot of vapour pressure benzene background and data is shown below in figure 5.2.1 together with its corresponding absorbance in figure 5.2.2 (calculated with the real background). Application of the bias routine to shift the differential above the x-axis also proved to be difficult given the nature of the benzene cross sections (e.g. see figure 2.3.8).

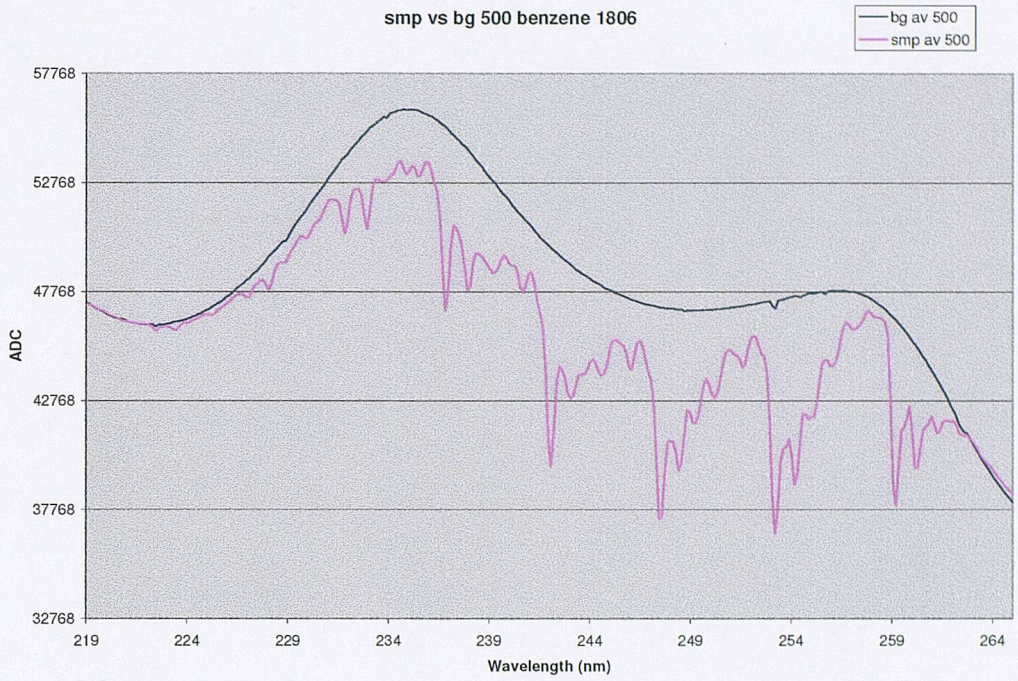


Figure 5.2.1: Benzene background and data for 18/06/02. NB. background drift not accounted for.

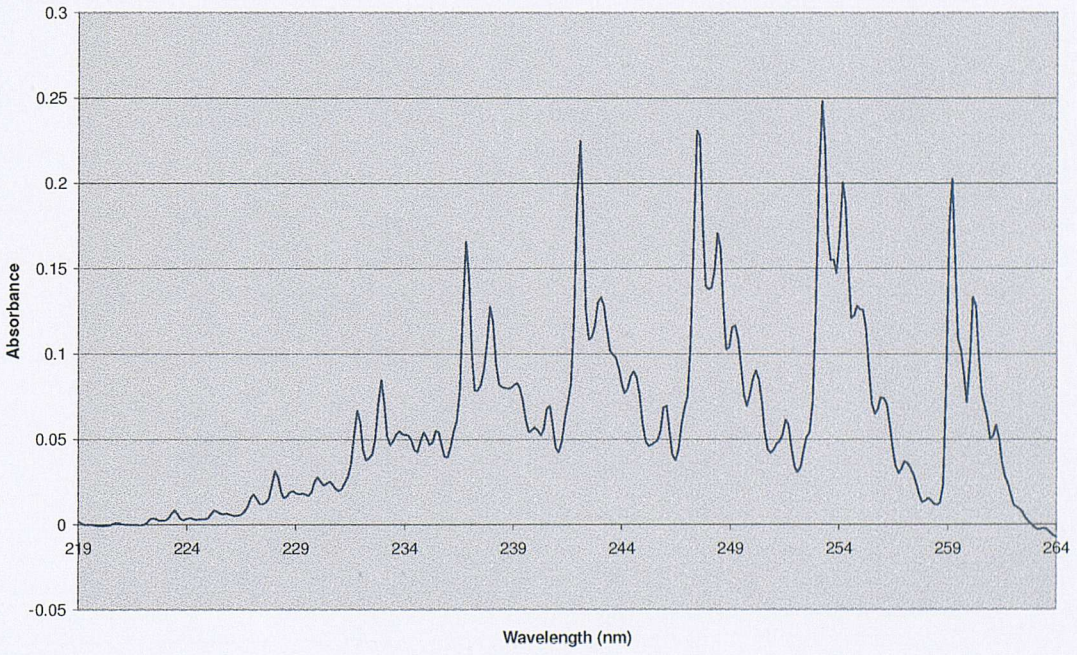


Figure 5.2.2: Absorption for data shown in figure 5.2.1

The background drift was not compensated for in the above benzene data due to the fact that concentrations low enough to yield signals similar to effective mixing ratios of order 5ppb (see table 1.3.1) proved difficult. The drift of the background below the level of the measured data explains how the absorption goes negative for high diode numbers in figure 5.2.2.

5.2.2 Results from Outside

It was decided that the cell should be taken outside and used as a means of introducing time varying concentrations. The aim was to replicate conditions of effective mixing ratios of around 500 ppb and change this over time by a few hundred ppb. This would then demonstrate both the ability of the GUSTO instrument to measure changes in pollutant concentration but also demonstrate the second by second aspect of the data retrieval algorithms, while using the optical set-up and data acquisition hardware from outside as opposed to the laboratory equivalent.

The data from outside proved to be dominated by an extremely low photon count, corresponding to around 0.1 of the full range, this then leads directly to a low SNR (see discussion in section 3.3). An example of the data taken for the time varying SO_2 experiment conducted on the 30th August 2002 is shown in figure 5.2.3.

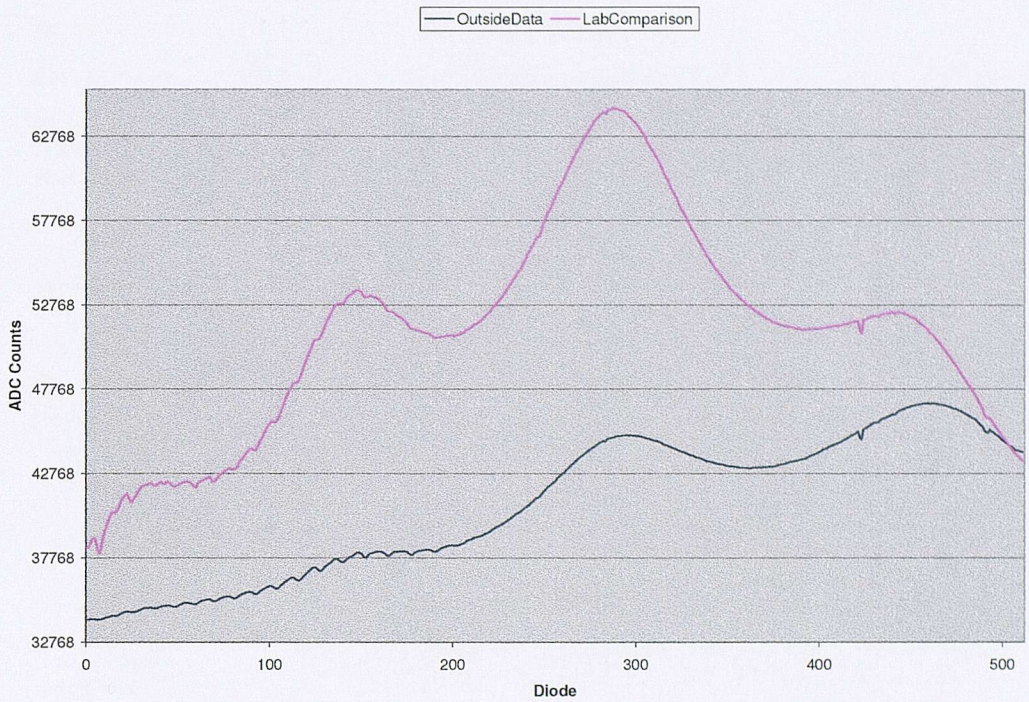


Figure 5.2.3: Example of shape of data obtained with outside system configuration, laboratory shape is shown for comparison

Clearly the magnitude and shape of the signal shown above is such as to reduce the SNR within the SO_2 range (c.f. figure 4.1.5). The experiment was run for 15 minutes. During the first 4 minutes the sample was injected into the cell and measured it in static conditions (in fact some SO_2 does escape from the cell, as confirmed by the fact that one of the laboratory cells left overnight with 2 ppm contained less than 200 ppb after 24hrs). The three septa were then removed such that gas could escape. The first and last parts of this data set (25 scan averages) as illustrated below with accompanying residual in figure 5.2.5. Shown in blue is the biased differential at the beginning of the experiment, while shown in purple is the same biased differential but for the end of the experiment. The difference between these is shown in yellow, illustrating the change in differentials over the course of the experiment.

This shows a decrease over the 15 minute period, which when turned into an effective mixing ratio by the usual DUVAS method gives the following figure 5.2.5 for the experiment duration.

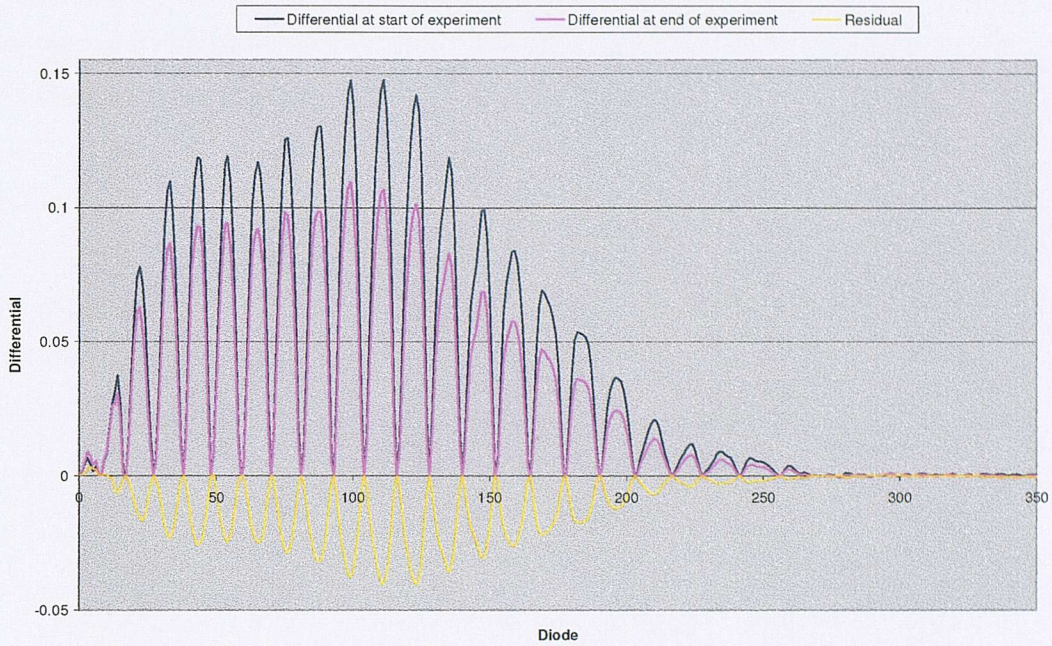


Figure 5.2.4: Start and end of data for time varying SO_2 experiment 30/08/02

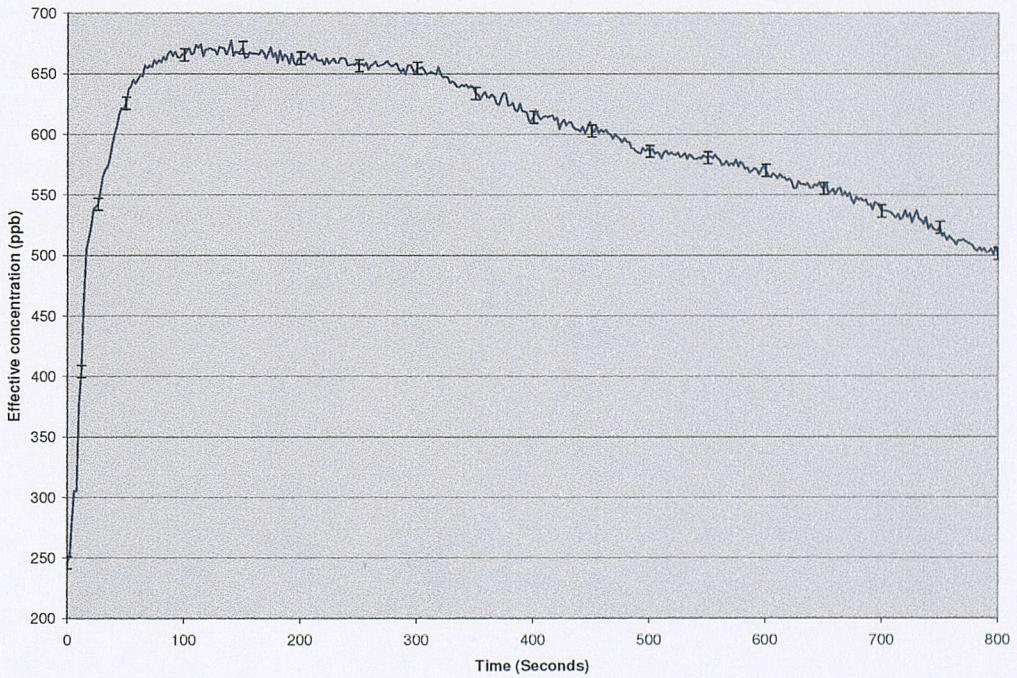


Figure 5.2.5: Effective mixing ratio as a function of time for outside measurement 30/08/02

The data shown above in figure 5.2.4 is for every scan and the error bars are the root mean square variation for the data set. The change from one (set of 5 scan average) to another is illustrated below in figure 5.2.6, the residual has been multiplied by a factor of 10 to allow it to be seen.

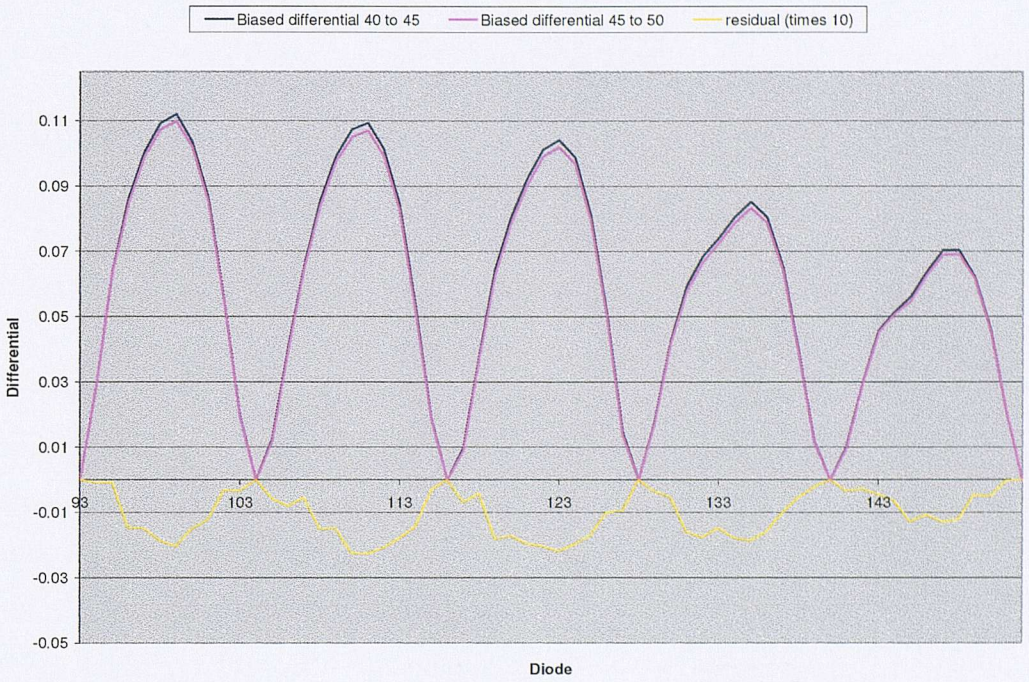


Figure 5.2.6: Biased absorption and residual for two successive 5 scan averages

Thus the outside measurements with the GUSTO system comprised a 15 minute data set varying with time from an initial value of near zero SO_2 mixing ratio to a maximum value of 680ppb ±5 (over the full 15m path) then decreasing by 1ppb around 4 seconds to a value of 500ppb ± 5. This mixing ratio was monitored every two seconds to an accuracy of ±5ppb.

Chapter 6

Conclusions

The aim of the GUSTO project as stated in the introductory chapter in section 1.5.2 is *the production and first use of an inexpensive, robust short path, high throughput DOAS instrument operating in the UV capable of real time concentration measurements of trace gas/gases, with an accuracy of better than ± 10 ppb.* The final unit as outlined by this text has achieved this aim. The unit is indeed inexpensive in comparison to any such similar unit operating in any of the ways summarised by table 1.4.2 capable of the type of measurements outlined in section 1.4.3. It has proven to be extremely robust, with the spectrometer and data acquisition hardware constantly being moved to and from the laboratory/outside unit placement. The only question with respect to robustness is the lifetime of the optics, which required attention[22] around once every six months. At fifteen metres the path length is shorter than other DOAS/DUVAS systems by almost an order of magnitude, and the system takes 'data' in real time, although the retrieval for outside measurements was done in the laboratory. The CPU time involved is less than half of the two seconds available to process data in real time[23]. The first use of the system has been demonstrated both in the laboratory and outside in the target environment, where measurements have been made in the UV on trace gas (SO_2) for changes of around 100 ppb with an accuracy of ±5, within the target of ±10 ppb, on a two second basis.

[22] Recoating of mirror surfaces and general cleaning, keeping input signal above ten thousand ADC counts out of the available thirty thousand.
[23] Figures for data shown in figure 5.2.5, reported by Microsoft Visual C++ on a 1.7GHz Athlon system with 512Mb RAM running Microsoft Windows 2000.

In order to be used for monitoring trace gas concentrations for the purposes of legislative control, the system needs around five times greater sensitivity for SO_2 and three to ten times the SO_2 accuracy for the remaining species shown in figure 2.3.6. Given the low photon count illustrated in figure 5.2.3 and the continued need for effective diode-to-diode response calibration covered in section 3.4 there are a number of logical improvements needed.

The photon count must be increased in order to maximize the full input range of the data acquisition hardware; this may be achieved by the following:

- A more powerful lamp: the existing 30W variant is sufficient to saturate the detector with the optics and path length of the laboratory but not those of the outside unit. Alternatively, more sophisticated optical set-ups could be used (more optimisation and matching of components), as 30W is good for health and safety given the beam diameter and photon loss [4]. A 100W version of the lamp is available[24], which would yield around a three times increase in light output from the system.
- Fewer optical components: the 45° "folding mirrors" shown in figure 4.1.1 not only block out the centre of the parallel beam but also simply fold the light so as to allow circular parabolic mirrors to be used. Eliminating them would reduce the number of optical surfaces by two as well as no longer removing the centre of the beam. This would necessitate moving to elliptical parabolic mirrors, where the difference in cost is less than 10% of the unit cost, yielding a factor of 1.5 increase in the light throughput of the system.
- The full range of the ADC should be used: the introduction of an analogue DC adding circuit to bias the input signal to the ADC by +2.5 Volts would allow the ADC to be switched from its ±5 Volt range down to its 0-5 Volt range. This would thereby be putting the input signal across the full range of all available bits used by the ADC[25]. This would effectively increase the output of the system by a factor of 2.

[24] Noise output is markedly different, spectral shape is similar over the GUSTO range.
[25] An alternative to the introduction of a DC adding circuit is to utilise the differential nature of the ADC input and hold the lower input at a constant 2.5V.

- The PDA could be replaced with a unit possessing a higher quantum efficacy per diode, thus increasing the signal obtained from the existing beam. If the same type of diode were used then more photons would increase the dynamic range, insofar as the unit is photon noise limited.

An order of magnitude increase in the output of the system is therefore possible. If this were implemented along with further study into the degradation of the optical components in polluted urban areas, there is no reason why the signal height need remain an overriding concern to the GUSTO project.

The diode-to-diode response, while low enough with laboratory calibration to allow measurement of SO_2, is still great enough to become the limiting error on any measurements of other species with lower absorption cross sections. The larger signal for the larger wavelengths will somewhat offset this effect for species absorbing at larger wavelengths than SO_2 (which is itself at the shortest wavelengths seen by GUSTO). It is also likely that this effect will still be the limiting error even at larger wavelengths. The effect of this on the system should also be investigated as a function of age of the PDA. This could be partially compensated for by the following:

- The PDA could be replaced with a version possessing lower pixel-to-pixel variation and/or shot noise: cost is clearly an issue as is the averaging of scans. Laboratory calibration, while proven as a technique, has yet to be applied to species other than SO_2.
- The integration time for each scan could be increased: This has a similar effect to averaging, however the noise from the data acquisition hardware would only be applied to the data once, as opposed to once each time a scan that was to be averaged is read.
- One could use an integrating sphere to uniformly illuminate all diodes and thus get a measure of the absolute variation pixel to pixel. Then one could correct for the small angle of incidence by putting light from a standard lamp through the system with all optics in place.

With some of the above implemented in future work, the system could then be applied to species other than SO_2. Benzene data have been included in this text, demonstrating the capability of the system, however more detailed quantitative studies are a logical next step. The problem of disentangling one species from another, both present in the same data set has been addressed during the course of the GUSTO project. However, without quantitative studies from which to form conclusions further discussion is beyond the scope of this text. Initial investigations look encouraging.

There are still many measurements that may be made beyond the initial short integration time, short path data outlined in chapter 5. For example, single species measured over extended time period and multiple species measurements looking at correlations/relative trends. The system could be altered to look at low level concentrations varying with height as well as time, or models could be tested for the ebb and flow of pollutants around real urban centres.

Once multiple species, real time, retrieval of ambient concentrations, within the accuracy required to monitor the targets set out in the governments guidelines summarised in table 1.3.1 have been verified, there is no reason why the GUSTO project should not move to achieve its long term aims set out in section 1.5.2. This could provide a system by which ambient concentrations of numerous trace gases can be monitored, relayed and analysed in order to provide useful information to a diverse range of recipients, across an entire city, in real time and on a second by second basis.

With the addition of IR capability to complement the UV, the possibilities extend, not only to more species of trace gas, but further to applications more localised than multiple geographic locations. There is also no reason why, given appropriate legislation, GUSTO-like units could not one day be a part of everyday life for motorists who require a fast, real-time, *in situ* and accurate description of the pollutant levels of their vehicles. There are many ways in which data taken by units from the GUSTO project may be used. I believe the contribution of the project to the supply of new information has great potential, both within and beyond the types of measurement described in this text.

Appendix

Overview

The following is a detailed discussion of the role of the UK government in urban air pollution control, with emphasis on the origins and content of legalisation together with likely future developments.

A1 Role of Government

In the Air Quality Strategy statement from the UK government [36] (2000) the role of the government in pollution control is defined as follows *"To provide a clear and simple policy framework that will allow and encourage everyone involved in improving air quality to play their parts fully and effectively."* The aforementioned document goes on to outline the broad features of a government strategy e.g. *"The framework should include realistic but challenging objectives. It should include appropriate legislation and offer financial incentives. It should also provide information and raise public awareness"*. The majority of the UK air quality strategy is the implementation of the Rio Declaration on Environment and Development (1992) from the UN Conference on Environment and Development (Rio, 1992) together with the various EC Directives (see later section entitled 'International Influence'). The entire strategy is by definition relating directly to air quality and much of the content, though both relevant and interesting, is beyond the scope of this text. There are however five principles of the nine that underpin the strategy which include the need for scientific understanding and therefore measurement of pollutant concentrations, and these are outlined below.

The 'risk assessment' principle of the strategy states "*We must base air quality objectives on an understanding of the relationship between exposure to pollution and its effects on human health and the environment.*" Whilst one can measure the effects of whatever pollution is present on the environment and quantify this effect, this implies a link between pollution levels and environmental effects, which will have little meaning unless the pollution itself can be characterised quantitatively with respect to environmental change.

The 'precautionary principle' comes directly from the Rio Declaration [37] and states "*where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation*". This is essentially to ensure that governments cannot get away with saying that they don't fully understand an environmental question and therefore don't have to do anything about potential problems associated with it. However this principle will mean that governments have to allocate money to all recognised aspects of environmental air pollution and clearly increased scientific knowledge of the specifics associated with these problems will allow these funds to be most efficiently distributed.

The 'using-scientific knowledge' principle of the strategy states that "*When we take decisions, we must anticipate early on where scientific advice or research is needed, and identify sources of high quality information. Where possible, we should review evidence from a wide range of viewpoints.*" Thus scientific research is required in order to provide the information used to facilitate the government's decision on the necessary objectives.

The 'effects-based approach' principle of the strategy states, "*The touchstone for action should be environmental objectives, expressed in terms of environmental quality. This allows the effects on people and the environment to be treated proportionately to the particular risk of harm or damage, using the most appropriate package of measures.*" This can only be achieved if the objectives are possible to implement and it is possible to verify that the implementation has been successful. Also one of the main objectives

is to minimise risk to health and insofar as this risk is associated with a specific exposure, the objectives must be related to specific pollutant exposure limits.

The 'effective international monitoring and enforcement' principle of the strategy is extremely clear in its aims that "*Ratification of international protocols should be followed by national action plans, where appropriate, and reports to the relevant supervisory body. We should monitor and enforce EC legislation rigorously across the Community.*" Not only is the concept of monitoring stressed in the Strategic Air Quality Review [36] but so too is the need for increased understanding of the processes involved in the behaviour of pollutants once they are in the atmosphere. Thus emphasis is also placed on increasing the research into modelling with respect to atmospheric pollution. The enforcement aspects of this type of approach are discussed below.

Legislation aimed at, or bodies created to monitor pollution levels on a local, national or international level are the direct implementation of the 10^{th} principle agreed at the UN Rio Conference [37] which grants individuals appropriate access to information relevant to them with regard to pollution. Legislation outlined in the National Air Quality Strategy Review [36] to enforce pollution controls generally implements the 16^{th} principle of the UN Rio Conference [37], which states that the polluter should pay for the effects any pollution originating from his/her actions/policy. There is heavy emphasis on this principle not being applied disproportionately to restrict fair trade or in any way restrict individuals or companies in matters covered by international law.

A2 Current Legislative Framework

The UK and others report concerns and data through European environmental information network, EIONET [20] to the European Environment Agency, EEA [68]. The EEA then originate plans for legislation and passes these to the European commission who ratify the legislation into directives. These directives are not law, however they are legally binding in that unless member states that do not opt out (only applicable in certain cases, not on environmental limits/caps/targets) they must

themselves pass legislation that implements these directives. Thus the UK government must draft and pass legislation to implement any of its own concerns with regards pollution control (not addressed by the EU) and all EU directives.

Further there are other bodies of which the UK is a member e.g. the United Nations Economic Commission for Europe (UNECE), which brings with it environmental responsibilities and the need for legislation also tailored around the European context.

Finally there is the UN, who put forward declarations that are used to underpin the implementation of legislation, e.g. Rio [37]. In recent years problems have been discussed in quantitative terms, though this has been limited to greenhouse gas emissions, e.g. Kyoto 1997 [69]. The differing aims of so many developed and developing nations makes quantitative unilateral agreement elusive (e.g. Kyoto), however, aims and methods to implement those aims can be agreed, and thus there is a third level to pollution control in the UK.

The worrying aspect of having this type of multi-layered approach to legislation is the possibility that the methods of implementation could be different in each case, leading to a situation where the same thing is covered in multiple and perhaps contradictory ways, together with separate systems for monitoring and feeding back information. This question has been addressed by the EU and the UK government, to which end the EU uses the aims and methods of any UN agreements/principles to underpin its directives while the legislation from the UK government that implements these directives either extends existing legislation/frameworks or entirely replaces them. This standardization of the ambient air quality legislative processes was achieved by the EU directive on 'Ambient air quality assessment and management' [70] which sets out a strategic framework for tackling air quality consistently by setting European-wide limit values for twelve[26] air pollutants in a series of daughter directives. Thus current legislation in the UK is the implementation of these daughter directives that implement the Europe wide strategy.

[26] SO_2, NO_2, particulate matter, lead, CO, benzene, O_3, polyaromatic hydrocarbons, cadmium, arsenic, nicle and mercury. Discussed in section 1.3.4

A3 Future Legislative Developments

As of 2002 only the first 7 of the 12 pollutants mentioned in the EU framework have been covered in detail in their respective daughter directives. This leaves the remaining five, a detailed discussion of which is beyond the scope of this text. For completeness it is worth referring back to tables 1.3.1 and 1.3.2 for the first seven pollutants/existing targets and defining the only non-familiar member of the remaining five which is 'polyaromatic hydrocarbons'. These are a large group of organic compounds occurring as both gases and associated with particles, both of which can be inhaled and enter the lungs. Early studies show some evidence that some of these PAHs may be carcinogenic [36].

The majority of the emphasis for the need for future legislative control is on modelling exposure, cost analysis and new pollutants including the five already mentioned together with increasing emphasis on $PM_{2.5}$, as discussed in section 1.2.6. The present framework is heavily based on targets and the associated monitoring that is necessary to evaluate progress.

Bibliography

[1] The UK National Environmental Technology Centre.
http://www.aeat.co.uk/netcen/index.htm

[2] Committee on the Medical Effects of Air Pollutants (COMEAP). "The Quantification of the Effects of Air Pollution on Health in the United Kingdom" Government Department of Health, London, UK, HMSO (1998).

[3] The World Bank. "Dose-Response Functions and the Health Impacts of Air Pollution", *Environment Dissemination Notes*, (1994).

[4] McCarthy, R. "First Measurements using the GUSTO (Generic UV Sensors Technologies and Observations) Instrument" first year transfer report, Imperial College (1998).

[5] Department of Environment "*Air Quality: Meeting the Challenge*" HMSO (1995).

[6] UK Met Office. "The Great Smog of 1952" HMSO (2000).

[7] Roholm, K. '"The Fog Disaster in the Meuse Valley, 1930': A Fluorine Intoxication". *The Journal of Industrial Hygiene and Toxicology,* (1937) Volume 19, No. 3, pp. 126-137.

[8] Snyder, L P. "'The Death Dealing Smog over Donora, Pennsylvania': Industrial Air Pollution, Public Health Policy, and the Politics of Expertise, 1948-9," *Environmental History Review,* (1994) Volume 18, No. 1, pp. 117-140.

[9] Schwartz, J. "'Air Pollution and Daily Mortality': A Review and Meta Analysis" Environmental *Research,* (1994) Volume 64, No. 1, pp. 36-52.

[10] Wong, C M. Ma, S. Hedley, A J. Lam, T H. "Effect of Air Pollution on Daily Mortality in Hong Kong" *Environmental Health Perspectives,* (2001) Volume 109, No. 4, pp. 335-341.

[11] Samoli, E. Schwartz, J. Wojtyniak, B. Touloumi, G. Spix, C. Balducci, F. Medina, S. Rossi, G. Sunyer, J. Bacharova, L. Anderson, H R. and Katsouyanni, K. "'Investigating Regional Differences in Short-Term Effects of Air Pollution on Daily Mortality in the APHEA Project': A Sensitivity Analysis for Controlling Long-Term Trends and Seasonality"' *Environmental Health Perspectives* (2001) Volume 109, No. 4, pp. 349-355.

[12] Brimblecombe, P. "The Big Smoke" (1987). Methuen.

[13] Pearce, F. "Back to the days of deadly smogs" *Nature* (1992) No. 1850, pp.25-28.

[14] Wilson, R. Spengler, D. "'Particles in Our Air': Concentrations and Health Effects" (1996) Harvard University Press.

[15] Mexico City Air Quality. "Environmental Information System." http://www.tlmsn.imeca.com.mx/tlmsn_valle_de_mexico/mexico.htm

[16] The UK National Air Quality Information Archive. http://www.airquality.co.uk/

[17] The World Gazetteer. "Current Population Figures for Cities, Towns and Administrative Divisions of All Countries" Published online (2002).

[18] Enviromental Protection Agency. "'National Emission Standards for Hazardous Air Pollutants': Revision of Source Category List Under Section 112 of the Clean Air Act" Federal Register (2002) Volume 67, No. 29, pp. 6521-6537.

[19] U.S. Enviromental Protection Agency. http://www.epa.gov/

[20] US, Energy Information Administration. "Annual Energy Outlook 2002 with Projections to 2020" US Gov. (2002).

[21] UK Environment Agency. http://www.environment-agency.gov.uk

[22] U.S. Department of Health and Human Services, Public Health Service "National Toxicology Program. 9th Report on Carcinogens, Revised January 2001" US Gov. (2001).

[23] UK Department for Environment Food and Rural Affairs. www.defra.gov.uk

[24] US. Office of Transportation Technologies, DOE. http://www.ott.doe.gov/

[25] '"Ozone in the United Kingdom:' Fourth Report of the Photochemical Oxidants Review Group" UK, HMSO (1997).

[26] Ravishankara, A. R. Hancock, G. Kawasaki, M. and Matsumi, Y. "Photochemistry of ozone: surprises and recent lessons," *Science* 280, 60-61 (1998).

[27] US, Office of Environmential Health Hazard Assesment and the American Lung Association of California "Air Pollution and Children's Health" US Gov. (2000) California.

[28] B. Trost, J. Stutz and U. Platt. "UV-absorption cross sections of a series of monocyclic aromatic compounds" *Atmospheric Environment*, (1997) Volume 31, No. 23, pp. 3999-4008.

[29] Pollution control department, Air Quality and Noise Management Devision. Thai, Ministry of Science, Technology and Environment. "Health Aspects of Air Pollution" Thai Gov. (2001).

[30] Vedal, S. Air Resources Branch, Ministry of Environment, Lands and Parks "Health Effects of Inhalable Particles: Implications for British Columbia" (1995).

[31] Microsoft Corp. "Automating Office with VC++: Knowledge Base Articles" Microsoft Corp. (2000).

[32] Department of the Environment, Transport and the Regions. "The Air Quality Strategy for England, Scotland, Wales and Northern Ireland, A Consultation Document" HMSO (1999).

[33] Platt U., Perner D. and Pätz H. "Simultaneous measurement of atmospheric CH_2O, O_3, and NO_2 by differential optical absorption" J. *Geophs. Res*, Volume 84, 6329-6335 (1979).

[34] Sigrist, M W. "Air Monitoring by Spectroscopic Techniques" Wiley (1994) London.

[35] UK, National Atmospheric Emissions Inventory. http://www.naei.org.uk/pollutantdetail.php

[36] Department of the Environment, Transport and the Regions "The Air Quality Strategy for England, Scotland, Wales and Northern Ireland, Working Together for Clean Air" HMSO (2000).

[37] UN Conference on Environment and Development "Rio Declaration on Environment and Development" (Rio, 1992).

[38] EMEP (co-operation program for monitoring and evaluation the long-range transmission of air pollutants in Europe). *"Transboundary acidification and eutrophication and ground level ozone in Europe"* Joint CCC & MSC-W Report *(2002)* Europe, published online.

[39] The Airborne Particles Expert Group. "Source Apportionment of Airborne Particulate Matter in the United Kingdom" Report for the Department of the Environment, Transport and the Regions, the Scottish Office, the Welsh Office and the Department of the Environment (Northern Ireland). HMSO (1999)

[40] EEA, European Environment Agency. http://www.eea.eu.int/

[41] Edner, H. Ragnarson, P. Spannare, S and Svanberg, S. "Differential optical absorption spectroscopy (DOAS) system for urban air pollution monitoring" *Applied Optics*, (1993) Volume 32, pp. 327.

[42] Reisinger, AR. Fraser, G.J.Johnston, P.V McKenzie, R.L Matthews, W.A. "Slow-scanning DOAS system for urban air pollution monitoring" In: Proceedings of the XVIII Quadrennial Ozone Symposium, pp. 959-962. Bojkov, R.D.; Visconti, G. (eds.) Parco Scientifico e Technologico d'Abruzzio, L'Aquila, Italy (1998).

[43] Bishop, G A. Starkey, J. Ihlenfeldt, A. Willams, W and Steadman, D. '"IR Long-Path Photometry': A remote sensing tool for automobile emissions" *Analytical Chemistry,* (1989) Volume 61, No. 10, pp. 617-677.

[44] EIONET, European Enviroment Information and Observation Network. http://www.eionet.eu.int/

[46] Toumi, R. "Daily online pollution forecasts for London" http://www.sp.ph.ic.ac.uk/~rtoumi/PANIC/Panic.html

[47] Mellqvist, J. and Rosen, A. "DOAS for Flue Gas Monitoring –I. Temperature Effects In the U.V./Visible Absorption Spectra of NO, NO_2, SO_2 and NH_3" *Quant. Spectrosc. Radiat. Transfer* (1995) Volume 56, No. 2, pp. 178-208.

[48] Brauers, T. Hausmann, M. Brandenburger, U. and Dorn, P. "Improvement of differential optical absorption spectroscopy using multichannel scanning technique" *Appl. Opt.* (1995) Volume 34, No. 21, pp. 4472-4479.

[49] Stutz, J. and Platt, U. "Improving long-path differential optical absorption spectroscopy (DOAS) with a quartz-fiber mode-mixer" *Appl. Optics,* (1997) Volume 36, pp. 1105-1115.

[50] "Numerical Recipes in C" Cambridge University Press, William H. Press, Second Edition. Section 14.8 Savitzky-Golay Smoothing Filters page *650-655,* (2002).

[51] "Numerical Recipes in C" Cambridge University Press, William H. Press, Second Edition. Section 13.5 Digital Filtering in the Time Domain page *558-564,* (2002).

[52] "Numerical Recipes in C" Cambridge University Press, William H. Press, Second Edition. Section 2.3 LU Decomposition and Its Applications page *43-50,* (2002).

[53] ESA Contract No11340/95/NL/CN. "A Study on Cross Sections in the UV and Visible, Final Report and Data" Imperial College, Bremen University, Rutherford Appleton Laboratory and Serco Europe Ltd. (1999).

[54] Martin, P. "Second Academic Prototype of the GUSTO System" first year transfer report, Imperial College (2000). http://www.hep.ph.ic.ac.uk/~martinp/pdf/TRJuly00_PM.PDF

[55] Hassard, J. Toumi, R. "The GUSTO Project" NERC Connect Programme Final Report. (1999).

[56] Hamamatsu Photonics, "C4545 power supply Product Manual" (1999).

[57] CVI Laser Corporation. "Data sheet for the CM110 Compact 1/8 Meter Computer-Controlled Monochromator" (2002).

[58] Hamamatsu Photonics, "NMOS linear image sensor S3901/S3904-F series product data sheet" (1999).

[59] Computer Boards "PCM 16D PCMCIA Product Manual" (2000).

[60] Wagady, M. Awad, S. "Determining ADC Effective number of bits via histogram testing" *IEEE Transactions on Instrumentation and Measurement,* (1991) Volume 40, pp. 770-772.

[61] Renneboog, J. Bossche, V. Schoukens, J. "Dynamic Testing and Diagnostics of A/D Converts" *IEEE Transactions on Circuits and Systems*, (1986) CAS No. 33, pp. 775-785.

[62] Hamamatsu Photonics, "L2D2 L6565 Deuterium Lamp Product Manual" (1999).

[63] Verbal communication with Ray Wrigley, Space and Atmospheric Group, Imperial College (2002).

[64] Set of neutral density filters and micrometer precision translation stage on loan from the Earth Observation Calibration Facility (EOCF) bought for calibration of the Geostationary Earth Radiation Budget (GERB) Project. Imperial College (2002).

[65] MERK Eurolab Laboratory supplies http://www.merck-ltd.co.uk

[66] "Numerical Recipes in C" Cambridge University Press, William H. Press, Second Edition. Section 13 Convolution and Deconvolution Using the FFT page *537-606,* (2002).

[67] Palmer, C. "Diffraction Grating Handbook". (5^{th} Edition), Thermo RGL. New York (2000).

[68] UK, Department of Trade and Industry. "The Energy Report 2000" HMSO (2000).

[69] United Nations, Kyoto conference on climate change. Kyoto, Japan (1997).

[70] Council Directive 96/62/EC of 27 September 1996 on ambient air quality assessment and management.

[71] Encyclopaedia Britannica, Volume 20, pp 840-842 (1963).

[72] Stutz, J. Platt, U. "Numerical analysis and estimation of the statistical error of differential optical absorption spectroscopy measurements with least-squares methods", Applied Optics, 35, 6041-6053, (1996).

[73] Burrows, J. P. Dehn, A. Deters, B. Himmelmann, S. Richter, A. Voigt, S and Orphal, J: "Atmospheric Remote-Sensing Reference Data from GOME: 2. Temperature-Dependent Absorption Cross Sections of O3 in the 231-794 nm Range", Journal of Quantitative Spectroscopy and Radiative Transfer 61, 509-517, (1999).

[74] Pascu, M. L. Moise, N. Staicu, A. "Tunable dye laser applications in environment pollution monitoring" Journal of Molecular Structure, Volume 598, Issue 1, pages 57-64, (2001).

[75] Richards, M. "Study of High Resolution Cross Sections of Nitric Acid Vapour". Imperial College, PhD Thesis (1998).

[76] Ahmad, S. R. Billit, E. M. "Implications of atmospheric attenuation in Raman lidar detection of pollutants", Optics and Laser Tecnology, Volume 23, Issue 3, pages 180-188, (1991).

[77] Richter, P. I. "Air pollution monitoring with LIDAR", TrAC Trends in Analytical Chemistry, Volume 13, Issue 7, pages 267-275, (1994).

[78] Slezak, V. Santiago, G. Peuriot, A. "Photoacustic detection of NO_2 traces with CW pulsed green lasers", Optics and Lasers in Engineering, Volume 40, Issues 1-2, pages 33-41, (2003).

[79] Berden, G. Peeters, R. Meijer, G. "Cavity ring-down spectroscopy: Experimental schemes and applications." Int. Rev. Phys. Chem. 19 pp 565-607, (2000).

[80] Adams, P. J. Seinfeld, J. H. Koch, D. M "Global concentrations of tropospheric sulfate, nitrate, and ammonium aerosol simulated in a general circulation model" J. *Geophys. Res.* 104, 13791-13823, (1999).