WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:
http://wrap.warwick.ac.uk/133958

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

16

Black Box and Mechanistic Modelling of Electronic Nose Systems

By

James W. T. Yates

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Engineering

University of Warwick, School of Engineering

September 2004

Contents

List of Figures

List of Tables

For Yan

Acknowledgements

There are so many people, without whom this thesis wouldn't be in your hands now. Firstly I'd like to thank my supervisors Dr. Mike Chappell and Prof. Julian Gardner for their kind support throughout my studies. I think everything makes sense now. Cheers!

Then I'd like to thank everyone in the School of Engineering at the University of Warwick. In particular I'd like to thank Prof. Keith Godfrey for my new boss (you know who I mean); Dr. Neil Evans, Dr. Reza Ali, Dr. Graham 'Fatboy' Searle and Mr. Ross 'The Reverend' Folland for stimulating coffee breaks; Frank Courtney and Ian Griffiths for their expertise and helping this poor mathematician to hold a screwdriver properly; Dr. James 'Doctor Daddy' Covington for mumbles, curses, electronics and advice; Everyone in A405b. Thank you.

Then there are all my friends that I don't work with (but probably still long suffering): Rich Norton; Gav Howell; Ade Legouix and Nessa Sainsbury; Andy and Emma Turl; Adrian and Rosie Manise; Rich Howell; Karen Kudar; Jim Berryman; Jez Austin; Glen, Anne and Laura Lane; Diana Smith. Thank you for allowing me to share my thoughts, music, undisciplined hair and unintentional Ozzy Osbourne impressions with you.

I'd especially like to thank my family for their encouragement and patience. See - I will get a proper job now! Thanks Grandpa, Mom, Dad, Mike, Catherine, Pete, Jules, Sylvia, Ka Chai, Lorretta, Tony, Fanny, Wing Fei, Danny, Ethan, Chloe and the rest of the Yates and Cheung clan...

...and thank you Yan for humouring me, teasing me and making me laugh.

James Yates. Coventry, September 2004

Declaration

This thesis, and the material in it, is my own work. It has not been submitted for a degree at any other university.

The following articles have been published, or are in the progress of submission as a result of the work contained within this thesis:

J.W.T Yates, J.W. Gardner, M.J. Chappell and C.S. Dow. Identification of Bacterial Pathogens using Quadrupole Mass Spectrometer Data and Radial Basis Function Neural Networks. *IEE Proceedings: Science, Measurement and Technology.* In Print.

J.W.T. Yates, J.W. Gardner, M.J. Chappell, F. Bolt, L. Beeby and C. Dowson. Chemical sensor screening of blood samples: Robust analysis via data set reduction. *Proceedings: IFAC Modelling and Control in Biomedical Systems, 2003* pp 289-294.

J.W.T. Yates, M.J. Chappell, J.W. Gardner, C.S. Dow, C. Dowson, A. Hamood, F. Bolt and L. Beeby. Data Reduction in Headspace Analysis of Blood and Urine Samples for Robust Bacterial Identification. Revised draft submitted to *Computer Methods and Programs in Medicine.*

J.W.T Yates, M.J. Chappell and J.W. Gardner. Phenomena Based Dynamic Model of Carbon Black-Polymer Composite Sensors. *Proceedings IEEE Sensors conference 2004* Vienna, October 24-27. (M. Vellekoop, Editor)

Abstract

Electronic nose systems have been in existence for around 20 years or more. The ability to mimic the function of the mammalian olfactory system is a very tempting goal. Such devices would offer the possibility of rapid chemical screening of samples. To gain a detailed insight into the operation of such systems it is proposed to carry out a systems modelling analysis. This thesis reports such an analysis using black box and mechanistic models.

The nature and construction of electronic nose systems are discussed. The challenges presented by these systems in order to produce a truly electronic nose are analysed as a prelude to systems modelling. These may be summarised as time and environmental dependent behaviour, information extraction and computer data handling.

Model building in general is investigated. It is recognised that robust parameter estimation is necessary to build good models of electronic nose systems. A number of optimisation algorithms for parameter estimation are proposed and investigated, these being gradient search, genetic algorithms and the support vector method. It is concluded that the support vector method is most robust, although the genetic algorithm approach shows promise for initial parameter value estimation.

A series of investigations are reported that involve the analysis of biomedical samples. These samples are of blood, urine and bacterial cultures. The findings demonstrate that the nature of such samples, such as bacterial content and type, may be accurately identified using an electronic nose system by careful modelling of the system. These findings also highlight the advantages of data set reduction and feature extraction.

A mechanistic model embodying the operating principles of carbon black-polymer sensors is developed. This is validated experimentally and is used to investigate the environmental dependencies of electronic nose systems. These findings demonstrate a clear influence of environmental conditions on the behaviour of carbon black-polymer sensors and these should be considered when designing future electronic nose systems.

The findings in this thesis demonstrate that careful systems modelling and analysis of electronic nose systems allows a greater understanding of such systems.

Chapter 1

Thesis Introduction

In spite of its ubiquity in the human experience, the sense of smell has never attracted the same fascination that some of the other senses have. It is possible that this is due to the lucid picture that sight and sound alone provide. However the chemical nature of the air, delivered to our minds via the sense of smell, can tell a great deal about that which cannot be detected by sight and sound. Take, for example, the sense of taste; smell is inherent to this aspect of experience as the tongue can only detect such coarse concepts as saltiness, sweetness, sourness, and bitterness; in comparison there are thought to be of the order of a thousand types of nasal receptors [6].

These chemical signals that are called smell play a major role in nutrition, territorial recognition and orientation, sexual behaviour and the detection of hazards such as fire or noxious fumes. In higher organisms, such as humans, a highly specialised and sensitive olfactory system has evolved. Here a distinction is made between the odour sensation and the volatile molecule that caused it. Odorant refers to a molecule that bears an odour, the odour is the resulting perception. In the human olfactory system it has been found experimentally that the olfactory receptors in the nasal cavity are typically sensitive to concentrations of the order of parts per billion of odorant molecules. However, our olfactory experience is largely due to mixtures; it has been found that the human nose is poor at identifying individual compounds in a mixture. This is not surprising as an arbitrary mixture of odorant molecules may not necessarily exist in nature.

It can be seen how much the olfactory experience has infiltrated modern human thought by the number of adjectives relating to odours: for example sulphurous, fruity, earthy, and minty [7]. These classifications have been extended and have now been used and evolved over the past century or more by olfactory specialists. These include those in the cosmetic trade, such as perfumers, who have now an extensive vocabulary to describe the sensations that are caused by the odorants they work with. This is also true of wine connoisseurs. The rich language used to describe the 'bouquet' of a wine is well known and often parodied for its hyperbole.

The thesis of the above is: without really realising it, the chemical make up of the atmosphere is extremely important to our perception of the world around us. It will be seen later that a sense of smell is necessary to industry as well as the individual.

1.1 Motivation

Why is it necessary to develop a system that can smell? Quite simply an artificial nose would be sensitive to the volatile chemistry of a sample; it would never tire or fall ill in the way that humans do. This would be extremely useful if it was hypothesised that there is a difference in this chemistry between two types of samples; it will be demonstrated later that this approach can be used to detect pathogens in biological samples. Another reason is that chemically reactive sensors may detect compounds that we ourselves are not sensitive to, and this will become evident later in the thesis. Finally, it will be demonstrated that these systems provide diagnostic tools that are significantly quicker than existing methods.

Systems have been developed which respond to gases. These devices are commonly known as 'electronic noses' due to their obvious emulation of organic systems. The outputs of the sensors in an electronic system are usually of the form of some voltage or current change. This output needs to be interpreted in some way so that it may be made useful. This may be an interpretation of the nature of the sample producing the vapour, or for a greater understanding of the internal mechanisms of the sensor. This method of interpretation must be an algorithm of some sort which maps the system output to some target classification. This then requires some data processing

or modelling of the system.

The characterisation of a system by a mathematical abstraction, a model, has long been an integral part of scientific investigation. The aim is often to predict the input-output behaviour of a system in order to test theories about its internal structure.

Another reason for developing a model is to analyse the output of the system. This is usually in a pattern recognition context where the output is known, but the input is not; this unknown input is the identity of the sample. This is a second type of problem because it is an inversion of the system input-output behaviour. The input needs to be predicted retroactively given the output. As with all inverse problems the questions are: is there a solution to this problem; if there is a solution, then is it unique?

There are two major methods of modelling. The first is to assume a standard form for the input-output behaviour of the model; this is the black box model approach. The alternative approach is to build a model based upon the assumed internal mechanisms of the system. These internal effects may well be experimentally unobservable and only a culminative effect may be measured. This second approach has the potential to determine knowledge, indirectly, about the internal states of the system; or at least with respect to the artifacts within the abstracted model of the real system.

This abstraction of modelling and its reconciliation with experimental data will be a recurring theme. The consequences of model complexity and the necessity to recondition data sets will also be explored.

1.2 Aims and Objectives

The aim of this thesis is to apply systems modelling techniques to electronic nose systems. This will be done in order to gain a greater understanding of the dynamics of such devices. To achieve this it will be necessary to develop modelling methodologies for electronic nose systems. Therefore the objectives of this thesis are to:

- Examine the current techniques for black box and mechanistic modelling.
- Analyse the relationship between model complexity and data quantity and quality.

- Develop data processing techniques for electronic nose systems.
- Investigate parameter identification algorithms.
- Apply these methods to output data from real electronic nose systems.
- Develop a mechanistic model of a specific electronic nose system.
- Develop experiments to validate this model.
- Investigate the environmental dependencies of electronic nose systems.

1.3 Thesis Outline

In this thesis the investigation will concentrate upon the application of modelling techniques to electronic noses; devices which mimic the function of mammalian olfaction. This is in order to characterise better the behaviour of gas sensors, but it has also been decided upon in an attempt to improve the discriminant power of the systems when applied to real world applications. This means taking into account noise and any time dependent characteristics of the system.

These are reasonably complex, modular systems and are composed of a number of specific functioning subsystems. The design of electronic nose systems is, for the purpose of this thesis, based upon that of the mammalian olfactory system. This is examined and current sensor technology is discussed. Sensor based electronic nose systems have, with respect to input-output behaviour, a number of interesting features. These are: time dependency due to changing characteristics of the sensors; concentration dependence, rather than just composition; and cross-sensitivity between sensors which yields redundant data.

A number of different modelling techniques are examined; their advantages and pitfalls are discussed. These techniques are applied in order to model the input-output behaviour of a system. The aim is three-fold: to predict the system behaviour; to understand the internal mechanism of the system; and to account for factors external

to the system ideal such as temperature fluctuations. Of interest then will be the models' ability to provide this information.

The typical design of an electronic nose system is discussed in Chapter 2. First the organic mammalian system is described, and this allows an analogy to be drawn which provides a means of creating an artificial olfactory system. The various means of detecting vapour are surveyed and their particular characteristics are discussed.

The problems with the designs of these sensory systems are identified, often with respect to the organic system, and the extent to which they affect the discriminatory power of electronic noses is considered. These problems range from environmental noise through to the characteristics of sensor responses being time dependent. The ways in which the work detailed in this thesis approaches these challenges is discussed.

The particular methods for modelling the systems are of course very important. In Chapter 3 various modelling techniques are discussed; specifically those that will be applied to real systems analysing biomedical samples later in the thesis. As the thesis title suggests both black box models and mechanistic models, based upon physical knowledge of the system, are discussed. The merits of both approaches are investigated.

The necessity of applying and validating these hypothesised models is recognised. In Chapters 4 and 5 the interactions between data sets of different sizes and models of varying complexity are examined. This is also done with respect to various parameter estimation techniques to judge how robustly they estimate the unknown system parameters. Robustness is taken in the sense of the best candidate model being identified. The results of these investigations are applied later in the work to yield more practical model building.

There are three aspects to parameter estimation: identifiability, estimability and the ability of an optimisation method to locate the feasible neighbourhood of the unknown parameter vector. These criteria are chosen because the reason for modelling is so as to analyse the data by model fits. It is found that greater robustness is obtained if model complexity is controlled by the model input dimension and methods are developed for realising this.

The practical applications of these techniques are then examined in Chapters 6 and 7. Specifically biomedical data are analysed in order to examine rapid laboratory

screening of urine and blood samples. Two electronic nose type devices are investigated. The first device is an Agilent 4440 which is a mass spectrometry device. The second is a C320 unit, produced by Smith Detection (Pasadena, USA), and comprises of 32 cross-sensitive sensor elements. These two devices are inherently different in the data that they generate.

The emphasis of the work changes in the second half of the thesis. A mechanistic model of a particular type of gas sensor is developed. This is done in order to characterise the way that the polymer sensors respond to volatile molecules.

The mechanistic approach is taken in the hope that *a priori* knowledge of the system will enable more faithful modelling. It is shown by making informed simplifying assumptions that the complexity of the resulting model may be minimised. From this, system characteristics are identified to yield a better portrait of the electronic nose system's behaviour.

In Chapter 8 the chemical mechanism within the body of a polymer interacting with a solvent are reviewed. A nonlinear PDE model is derived and is simplified to a linear case so that it may be solved analytically. It will be seen that from this an algebraic expression for the current passing through the sensor may be derived. The model's unknown parameters are also analysed in order to examine the information that may be extracted from them.

It is necessary to validate this model and to this end a series of experiments are detailed in Chapter 9. An experimental rig is designed to characterise the response of gas sensors. Environmental variables are recorded as well in order to understand the effect they have upon the sensor's response: this will be observed in the variation of the free parameters of the model. These parameters are estimated in Chapter 10. The relationship between model parameters and environmental factors is examined empirically.

This is followed in Chapter 11 by conclusions on the work performed.

Chapter 2

Odour, the human nose and its electronic analogue

The chemistry of smell can be extremely complex. The majority of volatile organic compounds (VOCs) are either alcohols or aldehydes [8]. Odorant molecules can range from ethylacetate [9], which indicates bad wine, to 2,3-diethyl-5-methylpyrazine [10], which is one of 30 identified constituents of roasted coffee bean odour. It should be noted that the chemical constituents of the headspace of something such as a foodstuff may contain in the region of 100 chemically distinct VOCs. This is made more complex by the fact that not only do we as humans perceive odour intensity, we also perceive valance [11]; this being the quality of the odour. An odorant can trigger the perception of a pleasant or unpleasant odour. For example, Citric acid is perceived to be "lemony" and "pleasant", whereas valeric acid gives the impression of smelling "rancid"; however, structurally they are similar.

How may this organic system be mimicked effectively? How may it be put to good use? To answer these questions, and to introduce the concept of an electronic nose system, the human olfactory system must be examined; then an analogy may be drawn, to act as a 'blueprint' for an artificial system.

The aim of this chapter is to describe the concept of an electronic nose. Here an electronic nose system will be formally defined as *...an instrument, which comprises of an array of electronic-chemical sensors with partial specificity and an appropriate*

pattern recognition system, capable of recognising simple or complex odours[12]. Further *headspace analysis* is defined as *the analysis of the chemical composition of volatile compounds by such a device.*

2.1 Anatomy and Physiology of the Human Nose

The nasal passage is lined with soft tissue known as the epithelium: the nasal epithelium which lines the majority of the passage and the olfactory epithelium which contains the olfactory receptors. In humans the olfactory epithelium is about 6.5 cm^2 in area and is difficult to distinguish from surrounding tissue due to its similar colour. Odorants reach these receptors in one of two ways (Figure 2.1). The first is by orthonasal transport through the nares by sniffing or breathing, and the second being retronasal transport from the oral cavity. In orthonasal transport a series of bones called the turbinates direct odorants to olfactory receptors at the top of the nasal cavity. This mechanism has been shown to be instrumental in the discriminatory function of the mammalian model [13]. How the olfactory receptors respond to odorants is still not fully understood. Anatomically a receptor cell is a long structure capped by cilia and has neural connections from its base to afford communication with the olfactory bulb (Figure 2.2). This structure is supported on all sides by supporting cells, and basal cells at the base; the site for new cell production. This bulb provides the first tier in processing the signals received from the receptors before they are relayed to the anterior olfactory nucleus, the olfactory tubercle, the prepyriform cortex, the amygdala and on to higher brain centres.

Knowledge of the mechanism by which olfactory receptors react to odorants is still sparse, and this is mainly due to the the large number of receptors and odorants. Importantly there are very little data available relating molecular structure with perception qualities such as threshold detection level and type; these being commonly referred to as potency and quality. The mechanism appears to be a 'lock and key' mechanism whereby an odorant molecule binds onto the cilia of a particular type of receptor causing a change in electrical potential across the receptor. About 300 olfactory binding proteins have been identified to date, this is accounts for around 3% of

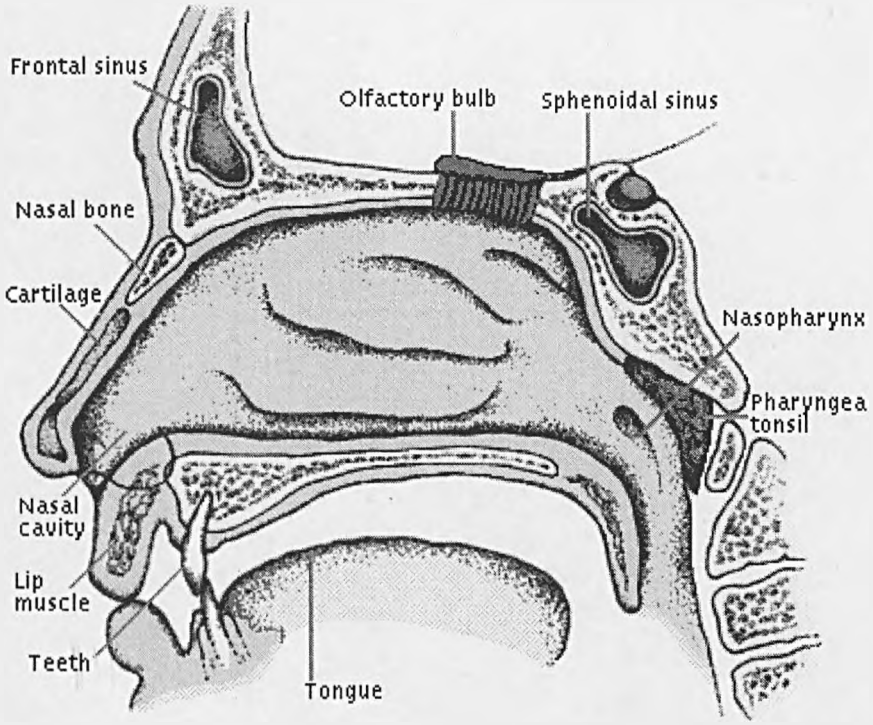


Figure 2.1: Slice through the human skull showing the nasal passages [1]

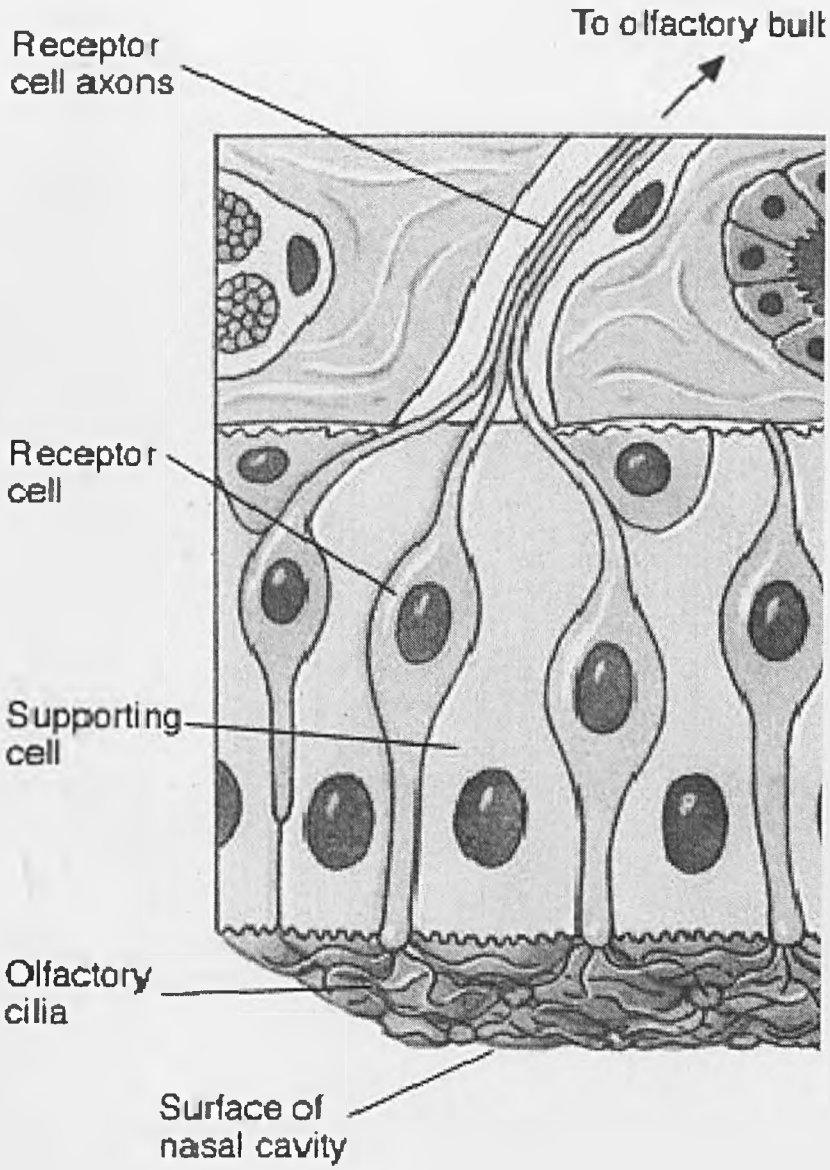


Figure 2.2: Detail of olfactory receptors [2].

the human genome.

It seems that combinations of odorants are more common than single species. This is probably as a consequence of evolution. It has also been noted that very volatile substances have much higher threshold detection levels than those that evaporate less readily. Thus odour analysis in the human olfactory system involves the processing of signals from receptors sensitive to different molecules. To complicate matters further, it has been suggested [6] that a specific receptor may respond to several different odorants and more than this that a single odorant type will bind to different receptor types.

Various types of receptors, sensitive to various airborne compounds are expressed in four different regions of the epithelium, and there are estimated to be a total of between 10^6 and 10^8 of these in humans [14]. These olfactory nerves have an average lifespan of thirty days and are then replaced. In the context of the pattern recognition problems discussed below, this neurogenesis, which is driven from below by the basal cells, poses interesting challenges for the identification of odorants. The characteristics of the detection system change with time and therefore processing has to adapt as well. This important aspect of the mammalian olfactory system is very difficult to emulate with artificial systems.

2.2 Mass Spectrometry and Gas Chromatography

One method of analysing the molecular composition of vapour is to separate the molecules in a sample by mass, which is the method used in mass spectrometry. The most common form of mass spectrometer is the quadrapole mass spectrometer.

This type of device uses varying potential differences to differentiate between different masses of molecule. The sample gas is ionised into a plasma by use of a beam of electrons. The flow of plasma is focused and made to run between four rods, parallel to their major axis. The rods are arranged in pairs of opposing rods. These pairs have an applied A.C. potential which is π out of phase between the pairs. This potential affects the flight of the charged particles and so only those of a certain mass to charge ratio may pass through an aperture and be detected by some ion detector. Different masses are allowed through by altering the potential between the rods. In this way

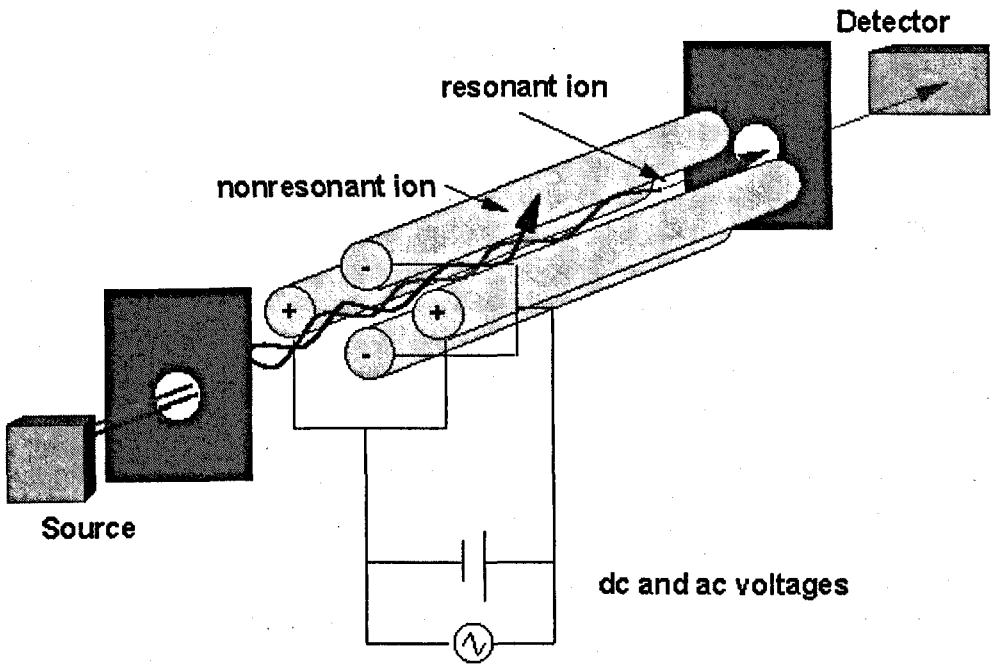


Figure 2.3: Schematic of a Quadrapole lens unit [3].

molecules with different mass to charge ratios may be differentiated. This process is illustrated in Figure 2.3.

There have been a number of studies seeking to mimic the human nose by using a mass spectrometer. Examples include foodstuff [15] and beverage [16] analysis as well as the detection of bacterial infection [17]. The systems can suffer from aliasing of molecular masses when presented by complex mixtures; the behaviour of such systems will be observed in later chapters.

Data processing is intrinsically the same for each channel of these devices. Each channel output is an intensity count for either time since sample introduction (gas chromatography) or a specific molecular mass (mass spectrometry). Thus this positive intensity number may be used as inputs of a model for discriminant analysis.

Gas chromatography is a variant of molecular separation in that it relies on diffusion and flow rates which will largely be dependent on molecular structure. A gas

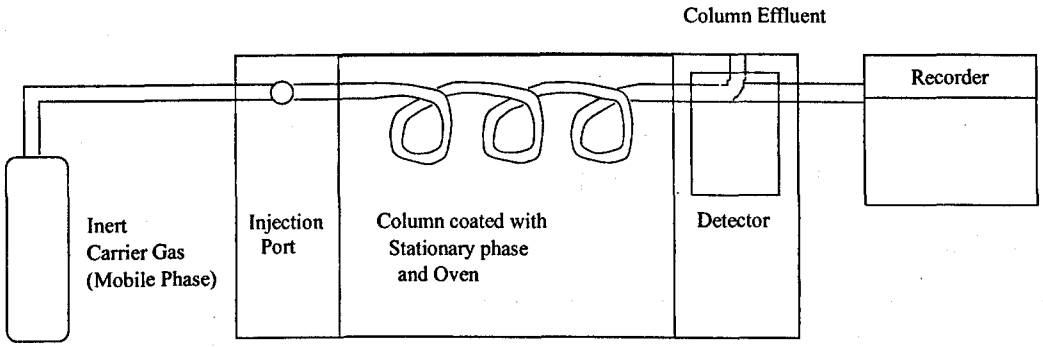


Figure 2.4: Schematic of the Gas Chromatograph Principle [4].

chromatograph consists of a long capillary tube which is packed or coated with a porous medium that forms the stationary phase of the capillary; the mobile phase being the carrier gas (Figure 2.4). Some, usually inert, gas such as helium or nitrogen flows along the tube and this carrier gas is the mobile phase.

Gas chromatography relies on three effects to separate different species of molecule: the rate that each species passes along the column, which is driven by a carrier gas flow, secondly the fact that there are paths of different lengths through the porous packing material which are available to different molecules, finally that the molecules move between the stationary and mobile phases at different rates. This arrangement affords remarkable resolution due to the varying rates that these above processes take for different molecular species [18].

The food industry is also interested in analysing foodstuffs via the volatiles they emit. As mentioned above, specific molecules are markers for rancidity etc. The quality of dried herbs [19] in relation to the method of drying has been investigated using gas chromatography. This allowed the researchers to draw positive conclusions about new methods of herb preservation. Alcoholic beverages [20] [21] have also been analysed in order to ascertain their perceived quality.

Another application for gas chromatographs has been the monitoring of environmental toxins. In [22], the presence of volatile organic compounds (VOCs) was monitored in a factory environment. There have also been a number of novel applications in animal behaviour such as [23], whereby the technology was used to isolate a scent which

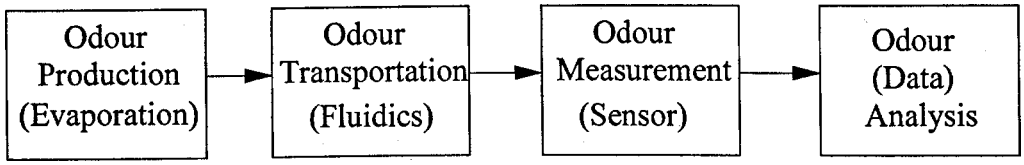


Figure 2.5: Schematic of a typical electronic nose system.

mice use to determine another's age. The paper industry [24] has seen applications to production quality.

There are numerous examples of gas chromatography used in a headspace analysis role. Many are medical in nature, from noninvasive heart monitoring via patient breath [25] to identifying bacteria [26] [27]. These underline the amount of investigation that has been performed with respect to medical applications of headspace analysis.

Gas chromatographs have also been incorporated with mass spectrometers to produce GC-MS. These units produce a time series of mass spectra for more detailed analysis. These have been used for food quality [28] and environmental safety [29].

2.3 Artificial Olfaction: The Electronic Nose.

Over the years there have been a number of attempts to mimic the human olfactory system. In fact the structure of the organic nose at a modular level has been used as a direct analogy to artificial endeavours(Figure 2.5). Thus the olfactory system is broken down into the following components: an odour delivery system to replace the nasal passages and lungs; a sensory array to meet the requirements set by the function of olfactory receptors; electrical measurement and transmission to convert and preprocess the sensor signals into electrical signals; data processing to extract information about odorants which the array has encountered to replace the brain.

2.3.1 Sensors

To construct a device to mimic the human nose, the common approach [30] is to have an array of sensors that are sensitive to odorants. The requirements are that they should be sensitive to levels of concentration of the order of parts per million down to parts per billion and that there should be some cross-sensitivity built into the array. Such sensitivity is required in order for the system to compare favourably with its biological counterpart. Cross-sensitivity is vital to the design of such systems and is a slight departure from the organic model, though, as discussed above, cross-sensitivity is manifested in the epithelium as well. If an array was designed that was faithful to the human epithelium it would require thousands of different, chemically specific sensors. This would generate both a manufacturing problem and, in the data analysis stage, a computational one. The consequences of cross-sensitivity will be addressed in a later chapter.

It is desirable that an array of similar devices may be constructed such that each individual sensor is sensitive to a different range of odorants. The reaction by which the particular sensor works should be reversible and reproducible; this is so that robust models may be built using the data. With present technology it would be unnecessarily difficult to construct an array with a comparable number of different receptors(1×10^7) as the human nose due to both power consumption [30] and the physical size of the array.

Electronic Nose sensors work on a similar principle to their biological analogue. Their function as a sensor can be roughly divided into two: a chemically reactive surface that has some property altered when it comes into contact with an odorant; and a method of measuring this change, be it conductance, a field effect transistor (FET) device, which utilises changing work functions, or a surface acoustic wave (SAW) microbalance. A linear relationship between concentration and sensor response is thought desirable and up until the advent of SAW devices this was not considered feasible [30]. Linearising circuits or preprocessing techniques [31] have been used in the past to modify the output in such circumstances.

Absorption of an odorant into a sensor may be divided into four heat dependent

processes [32]. Physisorption, that of non-chemical bonding like water in a sponge, is a low temperature energy process. This type of reaction normally results in a change of mass or dielectric constant which is reversible when chemisorption is not possible. Chemisorption (chemical bonding) is also a low temperature, energy driven process. Here a change may also be observed in electrical or optical properties.

Entropy driven, high temperature effects are categorised as surface defects and bulk effects. The most intrinsic of these being donor-type oxygen vacancies in oxides, for example tin oxide sensors. These result from lattice defects on the surface of the oxide sensor. Bulk effects can play an essential part in sensor operation. However, at low temperatures they are undesirable as they can result in nonreversible reactions that ultimately lead to sensor drift.

The scale of devices is important, both for power consumption and for reaction times. Practicalities dictate a response time of a few seconds, so it is desirable to produce these sensors on a small scale. It is also essential that the sensor quickly 'recovers' its baseline after the exposure halts [30].

Tin Oxide Sensors have been commercially available for many years. The surface of the oxide is oxygen deficient [33]. Electrons are removed from the conduction band by the absorption of oxygen molecules. The concentration of oxygen species on the surface of the oxide body is related to the reducing or oxidising properties of the gas detected. Taguchi sensors, being the most common commercially available sensors, consist of sintered tin oxide films containing a precious metal such as palladium or platinum. These tin oxide/catalyst sensors have been found to react more readily with combustible gases, although the mechanism is still not fully understood. There has been some investigation into this mechanism [34] [35], but the results of these studies have yet to be conclusive.

One disadvantage with tin oxide sensors is the high running temperature, which is in the region of 300 to 600°C. Alternatively, conducting polymers (electroconducting conjugated polymers) have intrinsic electrical conduction at room temperature. This is due to the polymer backbone comprising of alternate single and double bonds which result in charge mobility. Such materials as poly(acetylene), poly(pyrrole), poly(thiophene), and poly(aniline) exhibit these properties [36] and as such are candidate polymers for

gas sensors. In their pure state these materials are only semiconducting: conductance changes of several magnitudes are achieved by doping.

Deposition of polymer onto a recipient surface is normally brought about through an electrochemical reaction. This electropolymerisation is an extremely attractive technique as it allows controlled deposition of thick films [36]. As would be expected the particular polymer used [37], dopant [38] and resulting geometry [39] [40] all affect the dynamic response of a sensor when challenged with some vapour. Due to this it is possible to *a priori* dictate the sensitivity and selectivity of the sensors by controlling the fabrication process [41].

Nonconducting polymers may be used as gas sensors as well. There are two approaches, either to make the polymer conduct or to detect mass uptake as a solvent diffuses into the bulk of the polymer. Such materials are used as the coatings found in gas chromatographs described earlier.

The first method uses fine carbon or metal powders mixed in with a polymer in solution. If there is sufficient conductant by volume, when the polymer dries the composite material will conduct. The minimum volume fraction required is defined as the *threshold percolation fraction.* The resulting composite is deposited onto electrodes by spraying a microdot. The resulting composite material changes its conductance in response to some vapours; these diffuse into the body of the polymer and cause a swelling effect which causes the volume fraction of the conductant to fall.

The advantage of this type of sensor is that a much larger number of polymers become available in comparison with conducting polymers. Typical materials that are used include poly(4-vinylphenol), poly(vinyl-acetate), and poly(ethylene oxide). For a list of other polymers see [42]. As above, the approach is to construct an array of these sensors so that, through the chemical diversity present, a sensor array capable of discerning between a number of different odorants is produced. A full survey of the chemical mechanism of this class of sensors can be found in a later chapter where an analytical model is discussed.

Surface acoustic wave devices may be used in conjunction with polymers as well. In these devices a surface wave is transmitted along the sensor from transmitter to receiver. When the surface absorbs molecules from the environment it increases the

density of the medium transmitting the wave: altering the time it takes the wave to travel. This time delay is very small and so the sensor actually outputs a frequency change which may be measured with an accuracy of 0.0001% [43]. These devices become more sensitive the higher their resonant frequency [44] and so these types of devices typically operate at around 50MHz [45].

The exact chemical mechanism taking place in any of the above sensors is important. The rapidity of the reaction of interest is dictated by the binding energies involved, which can lead to undesirable properties if not correctly balanced. If the binding energy is very low then the reaction is very slow. However, if the binding energy is too high then the reaction becomes irreversible. Dynamic studies of these reactions will reveal their nature, and whether they might be used in this type of application.

2.3.2 Data processing

For the above sensors to be integrated within some system in a useful way the response of the sensor or array of sensors must be measured. It is reasonable that, first, the response should be logged in real time; this may be the resistance of the sensor or frequency of the output, depending upon the type of sensor. The sample rate might be of the order seconds or even tenths of seconds, for example, depending upon the time scale on which the sensors respond: moving from baseline to some steady state for the perturbed system.

The responses of these sensors as a time series of values are of very little obvious use if an input-output relationship is required. The organic nose analogy is thus continued to 'the brain'. A method for processing the dynamic response is sought with a number of desirable properties, the main being that the resulting data are manageable for pattern recognition and that maximum use is made of the information in the response regarding the sample.

The data resulting from logging the dynamic response from an array of sensors can be large, a computer file of tens of megabytes may result and will contain, as will be noted in a later chapter, redundancy, together with information on the chemistry that maybe of no interest as far as the application is concerned. Methods are therefore

necessary to look for features of the responses that distinguish between different types of samples. This is difficult without prior knowledge of how the sensors respond. Therefore a number of *ad hoc* methods are used for the processing of the response data ready for the application.

The most common technique for data analysis is to use the steady-state of the sensor after sample introduction. More specifically there is the absolute change and relative (or fractional) [42] [46] change which may be defined respectively as follows:

$$R_{ABS} = R_{max} - R_0 \tag{2.1}$$

$$R_{REL} = \frac{R_{max} - R_0}{R_0} \tag{2.2}$$

where R_0 is the baseline and R_{max} is the steady-state value of the resistive sensor; these are illustrated in Figure 2.6 which is a generic sensor response. This method is popular due to the fact that the static response of many sensors is proportional to the concentration of the analyte and it reduces sensor drift (see Section 2.5). However proportionality will be different for different analytes.

The other main data processing approach is to analyse the response time series. It is reasonable to suppose that dynamic analysis will yield as much information as the static measures above because static response can be retrieved from the dynamic; the converse is not true.

There have been attempts using black box modelling and mechanistic models (see below and next chapter for full definitions) to analyse time series.

2.4 Applications of Electronic Noses

These devices have a great potential for odour analysis in many different industries. This is because they are machines. In many case, especially the food industry, analysis is carried out using a human panel. The fundamental flaw with this type of method is that the panel is subject to the normal human frailties such as fatigue, sickness and subjective perception. The statistical design of such panel studies is such as to eliminate subjectivity, but at a great financial cost. Thus an automatic system is very attractive.

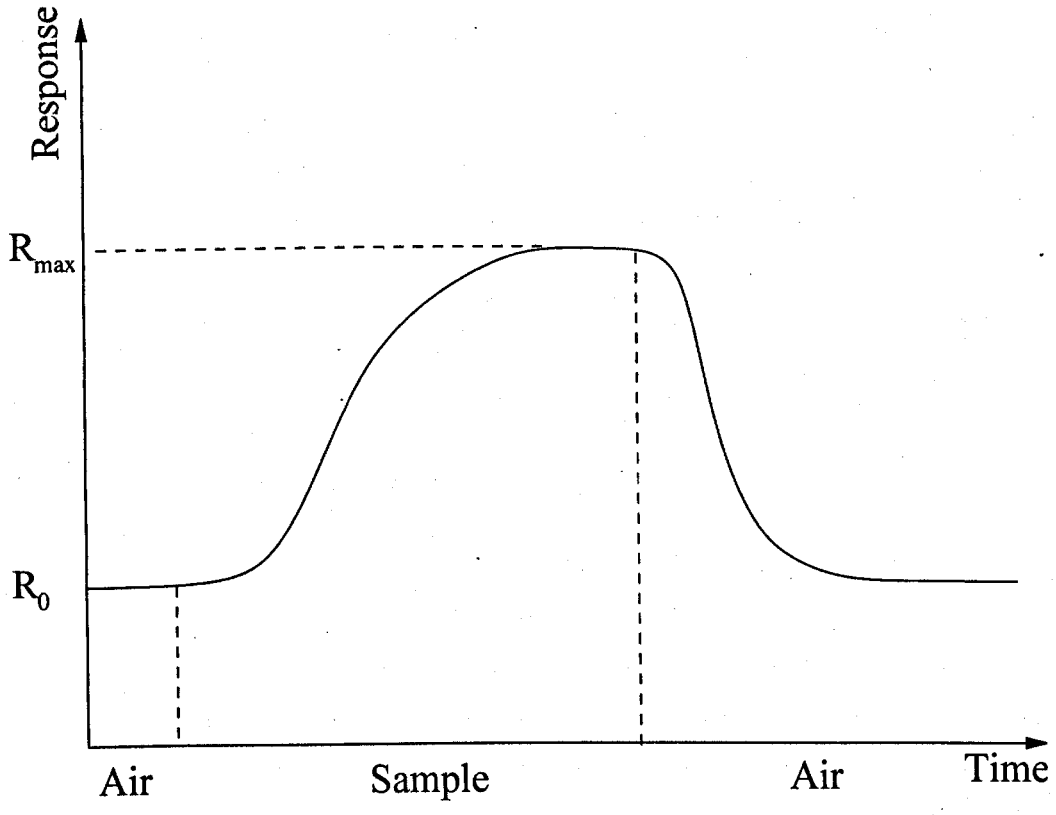


Figure 2.6: Illustration of the concepts used in steady state measure of responses.

To put into context the sensitivity and reproducibility required of the devices described above, a survey follows of the applications of headspace analysis. There are two major areas which require rapid, non destructive analysis: the food industry and the medical profession. Reproducibility will be seen to be a major factor in the context of medical applications. Electronic nose systems need to be robust in order to make consistently good diagnoses.

2.4.1 Food Industry

Since taste and food perception are largely driven by a human's sense of smell, the food industry is very interested in standardised measurements of this sense. Traditionally, companies have employed panels of testers for product development. These groups grade the smell and taste of foodstuffs against a number of categories that represent different aspects of perception. These may be such things as 'fruity', 'floral', 'oily', 'musty', and 'sour' [47]. This process can be costly or impractical when there are a large number of samples to be surveyed.

The challenges inherent for electronic noses in this particular area of olfaction result from the fact that all of these applications are based upon human perception [48], which is subjective. Attempts have been made to assess the ability of electronic nose systems to predict human percepts (the impression in the mind given by a sensory perception) [47], the conclusion being that while existing technology is effective for single compound samples, even binary mixtures are beyond the scope of current systems. The problem is that two chemically similar compounds may be perceived completely differently by humans, besides the consequences of genetic variety on perception.

Nevertheless some attempts have been made to emulate human perception in such diverse areas as milk [49], fish freshness [50], and olive oil [51]. The encouraging and positive conclusions of these studies demonstrate that there is an ever-increasing range of applications and potential applications of electronic noses in the food industry.

2.4.2 Medical

The importance of odours in medicine has been recognised since ancient Chinese medicine. Modern research into the medical applications of electronic noses has concentrated on rapid diagnosis via biological samples: samples such as eye swabs, blood and urine have all been considered. This has been the case due to the observation that, traditionally, a diagnosis is obtained several days after the production of a sample. This is due to the time it takes for laboratories to culture any pathogens that might be present. The situation may be improved by using some form of artificial olfaction unit that can analyse a sample in a matter of minutes.

In [52] a report is made of a study to detect human skin odour. It has been noted that certain ailments such as diabetes and some hepatic diseases result in an unpleasant odour. The study investigated the possibility of developing non-invasive diagnosis techniques based upon gas sensor technology. It was concluded that current sensor technology could detect odour changes in the test subjects' skins.

The type and metabolic state of bacteria have been successfully detected as demonstrated in [53] [54]. These results are extremely important given that pathogen drug resistance and toxin release are often cycle dependent. What is more astonishing is the sensitivity of the equipment necessary given the background chemical noise in such situations.

2.5 Challenges for Electronic noses

The above has demonstrated the technologies and applications associated with electronic noses. However there are still a great number of challenges for chemical analysis using electronic nose systems, especially those using resistive sensors. By examining these challenges the context of this thesis may be seen.

2.5.1 Sensor degradation

Electronic noses use chemical sensors and as such they react with whatever they are challenged. This results in two problems for the necessary repeatability of the equip-

ment; the first is degradation of the sensor structure; the second is sensor drift due to 'poisoning'. It is this drift due to odorant molecules becoming trapped in the body of a sensor that causes problems for the long term stable use of electronic noses and consistent reproduction of measurements. Any molecules that are not desorbed will alter the baseline and have the result of such static processing as discussed above.

The sensor output is required to be repeatable so that pattern recognition and modelling techniques may be applied as a way of analysing and elucidating important information about a substance being sampled. Any drift will result in previously constructed models becoming less valid as the sensors are exposed to more samples. These partially irreversible reactions are due to high binding energies for certain substances [44] which also create highly specific sensors. Thus there is a balance to be struck.

The challenge here is how can the sensors or data processing algorithm be designed in such a way as to eliminate the drift problem.

2.5.2 Empirical Nature of Black Box Modelling

The modelling and pattern recognition techniques discussed in this thesis are inherently empirical, they require data from real experiments; the models are optimised with respect to these data. As mentioned above these data may be time dependent, but there is also noise and its interaction with parameter estimation to be considered.

Electronic nose systems have two sources of noise. They are typically electrical systems and will be subject to electrical noise resulting from the mains and equipment around them. More importantly from a modelling point of view is the variation in the signal created by 'chemical noise'. The information content of the chemistry of a sample has been alluded to, but not quantified; this is a very difficult thing to assess. However, there may be compounds in samples that vary in some way, but which are not interlinked with the processes that are required to be observed. To illustrate this it will be seen that a bodily fluid sample will contain compounds relating to diet, illness elsewhere in the body, medication, sex and hormonal state, besides, perhaps, the presence or not of bacteria. The problem arises when this superfluous information

has a much stronger signal than that which is of interest. Hence ways of avoiding this situation must be investigated.

The problem with the multiarray paradigm is the large output dimension of the system which requires a great deal of data to fit even a linear model to the system's input/output behaviour. These data are often very difficult to obtain. This is particularly so in the case of biological samples where the 'training' set of samples must be classified by labourious laboratory techniques in order to carry out supervised pattern recognition. The challenge is thus to identify as simple a model type as possible that will successfully, according to some criteria, mimic the behaviour of the system.

2.5.3 Physical Constraints

The final set of challenges relates to the effects of the environment on the system. The systems in question reside in and interact with the environment. This makes the production of an all encompassing model extremely difficult. The first problem is that it is very difficult to manufacture a group of sensors that have identical characteristics. The geometric, and quite often chemical, properties of the sensors will vary. Thus a model will have to be tuned to each sensor. This is a problem similar to the problem of neurogenesis; the model has to be retuned each time that a sensor is replaced. This is an extreme example of sensor degradation.

The second problem is the variation of characteristics due to temperature and humidity changes. These both affect the chemical mechanisms within the active components of the sensor. A greater understanding of this would result in better predictive and discriminative models.

The final effect will be observed later in the thesis; the effect of gas flow rate upon the response. This effects the rapidity and the intensity of the resistive sensor's response.

2.6 Conclusion

The title of this thesis refers to the analysis of data from electronic noses. Therefore the discourse will be concerned almost exclusively with model, rather than sensor, design. It will be investigated how careful modelling techniques can improve and build upon the previous work discussed above. Data used in the studies will be of a biological nature: this is due to a long running collaboration between the School of Engineering and the Department of Biological Sciences, at the University of Warwick. Two types of electronic nose will be investigated. These are the mass spectrometer based headspace analysers and those systems based upon the resistive polymer sensors.

The two basic approaches of modelling will be considered, compared, and contrasted: black box modelling via assuming some standard input/output structure and white box modelling, often known as mechanistic modelling, whereby some previous knowledge or assumed internal mechanisms are 'built' into a model of the system. These methods are examined as a way of analysing the system's time series and predicting future input/output behaviour and classifying unknown samples.

Parameter estimation (model identification) is also considered. Ways of reducing model complexity are investigated in order to make parameter estimation a robust and well-posed problem. Methods include dimension reduction of data sets, model order reduction, assumptions of the smoothness of the model being sought and modelling using physical knowledge of the system.

Although the data analysed are of biological origin from electronic nose systems the approach taken in this thesis investigates the interaction between theory and experiment. This process is intrinsic within the training epochs of a neural network as it is with mechanistic modelling and validation of a category of gas sensor. A recognition is made of the fact that theoretical models are only as powerful as their experimental verification and in fact the converse is true: data do not become informative until some decision is made as to how to analyse them.

There are a number of challenges for these systems as discussed above. Black box and mechanistic modelling techniques need to be investigated in order to improve our understanding of these systems. The time and environmental dependencies of such

systems need to be identified.

In summary: information extraction from electronic nose data has been discussed. By information extraction is meant constructing a model whereby a data set may be divided along some predetermined classifications. The aims of this thesis are therefore:

- Investigate ways of gaining the full potential in terms of information extraction from electronic nose data.
- Investigate black box and mechanistic modelling techniques and thereby gain a greater understanding of the behaviour of electronic nose systems.
- Investigate the environmental dependencies of electronic nose systems.

Chapter 3

Black Box and Mechanistic Models.

The rationale behind the use of gas sensors is to ascertain the nature of some sample via its 'odour'. It is apparent that some form of pattern recognition is necessary for this task. The intrinsic assumption here is that the data with which the algorithm is provided have some difference, or pattern from which the 'type' of each datum may be identified.

A data set is a set of experimental measurements. Each datum is a particular sample. Consider the binary classification problem. In *supervised learning* a training set of samples will be supplied with a classification label for each sample (quite often 1 and -1). The algorithm will then attempt to find a model amongst a candidate set which best predicts each sample's class.

Models, as a method of information extraction, as well as knowledge validation, have featured highly in electronic nose research. There are a number of different areas of emphasis: sensor response characterisation, sample analysis and identification, information transfer.

Sensor characterisation is the non-parametric study of the analyte and environmentally dependent response of the sensor, usually the steady state response. Analyte dependence normally concerns the relationship between the same analyte at different concentrations and the resulting sensor response. It has been observed that resistive sensors, such as metal oxide sensors [55] [56] [30], have a nonlinear concentration dependence. This makes analysis difficult because data sets are often dominated by needless

concentration information.

Environmental factors that have been investigated include ambient temperature and that of the sensor body and humidity of the air. These have been demonstrated in several studies to have an appreciable effect upon the response of polymer type sensors. These investigations have tended to take a more nonparameteric approach [57]; general trend analysis has been carried out without any reference to mechanism; this being the philosophy of black box modelling. The bearing this has on electronic nose applications will be discussed in Chapters 7 and 8.

In the context of this thesis these data will be used for model building. This process is called system identification, That is, a particular model from a candidate set is identified from the experimental data. The model is chosen using some criteria of 'fitness'. The aim of the generation of this model will be pattern recognition and, in Chapter 8 and onwards, to gain an understanding of the sensor dynamics.

In pattern recognition it is hoped that some features of the data may be elicited that help distinguish between two or more different types of data. These features can be very difficult to ascertain. In the case of data from Electronic Noses, there are a great number of processes that take place between a sample introduced as the input to a system and the recording of data in a form that a computer can manipulate. Briefly there is:

- Odour production [58].
- Delivery to the sensors [59][60].
- Electrical/chemical reactions with these sensors [61].
- Observation of these responses by analogue and digital electronics.
- A treatment of the data ready for pattern recognition*.

Black box models used in pattern recognition assume two things; firstly that there is a detectable difference between the different classes of data, and secondly, a good

*This cannot be defined any further without losing generality. It may be that some kind of 'Static data' such as the change in conductivity are required, but this is really a preprocessing issue that is discussed in several parts of this thesis.

approximation of the indicator function is realisable within the class of the candidate black box models. That is that the phenomenon is such that it can be modelled and the form of model necessary is assumed to be known.

The first assumption is necessary as otherwise there is no way to continue with the modelling process. Of course, it is the primary hypothesis that needs to be tested, but stopping there will yield nothing more. The second type of assumption allows the search for a pattern to begin. Essentially it is a case of an assumed form of the pattern; the order of the polynomial in regression, whether two data clusters are linearly separable, etc. It is this assumption that allows pattern recognition techniques to exist, in particular artificial Neural Networks.

Let us define a system. A system is a dynamic object which has some input-output behaviour. The system behaves in some predetermined way (here only deterministic systems are considered) to an input stimulus. The system may be viewed as in a black-box scheme:

- An input which may be some control, stimulus or initial conditions.
- The dynamics represented by equations or some algorithm
- The output which is a real time measurement of a subset of the internal dynamics.

The mathematical modelling of a system may be divided into two distinct stages: modelling, where a tentative family of models is produced, in which it is hypothesised the behaviour of the system may be predicted; identification follows in which the particular model is identified, normally via parameter estimation. Modelling is the treatment of the problem of representing *a priori* structure information based on facts, assumptions, and hypotheses [62].

The identification stage involves the minimisation of the differences between the model and the observed behaviour of the system. This optimisation normally includes some iterative process which converges to a 'best' solution. There are a number of ways of approaching this. It is usual to derive some penalty function which combines the need for the model to fit the experimental data with some sense of 'least complex is

best'. This rephrasing of 'Occam's Razor'† is an important part of modelling, as shall be seen in the next two chapters, especially when dealing with relatively small data sets.

There are a number of criteria for the 'best' model of a system. 'Best' might be the smallest error, least complex model or one which is arrived at in the shortest time. The standard definition is that model with the largest *expected* success rate; this will be the model that balances complexity with error. It shall seen below that these ideas are incorporated into penalty functions for parameter estimation.

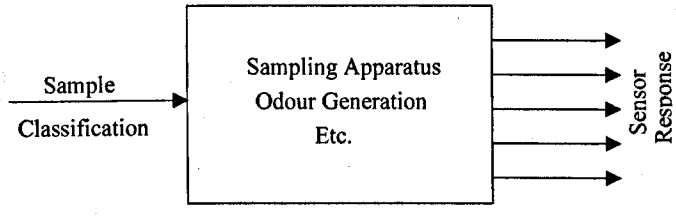
It is apparent that during the parameter estimation phase there is an optimisation process taking place against some form of error function. There are two parts embodied in the above three paragraphs, namely model and system identification. Model identification involves the production of some class of models and the actual system is identified by parameter estimation.

3.1 Forward and Inverse Models

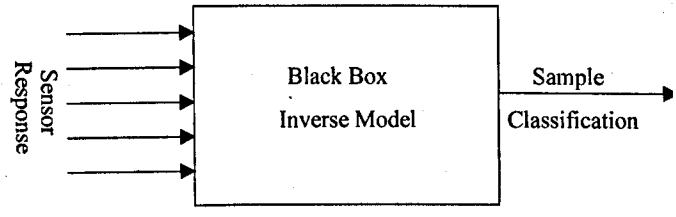
In a standard model, the experimental input-output behaviour would be compared with a hypothesised family of models, usually this family is a parameterised set of functions. This is the forward model concept (Figure 3.1(a)); given a particular input the output may be predicted. Pattern recognition normally requires the converse; given the output, what was the input? This is the inverse model (Figure 3.1(b)). The advantages of using inverse models are discussed in a number of papers including [63]. These models are *black box* because the model tells us very little about the internal processes of the system.

Looking at mechanistic models, such as differential equation systems, the input to the system is often embedded within the model solution in the form of parameters that need to be identified. Due to their implicit nature there is no standard way to invert the equation, gaining the input in terms of the predicted output. The only way

†Attributed to William of Ockham (c1285-1349), mediaeval philosopher: 'one should not increase, beyond what is necessary, the number of entities required to explain anything'. Known more recently as the principle of parsimony.



(a) Schematic of causal relationship between the sample classification and the sensor response.



(b) Schematic of the inverse model.

Figure 3.1: Schematic of forward and inverse models.

to 'invert' the model would be to identify the parameters of the forward model with respect to some dynamic response and to use these parameters as a way of representing the input. This is reasonable enough if there is some prior knowledge of the input, e.g. an impulse of a certain magnitude, sinusoidal or polynomial.

It is apparent, however, that the use of black box models allows this problem to be avoided. A black box model can be chosen such that the inverted input-output behaviour of the system may be modelled and identified in the experimental data. The resulting parameterised model may then be used to retrospectively predict the input.

3.2 Standard Forms of Black Box Models

Here, black box models are considered from a pattern recognition point of view. By this it is specifically meant that there is some data set with some classification label from a finite set of labels attributed to each datum. Hence, the inverse model problem is considered; this is because, in the original experiment, the sample classification was the input and the sensor response was the output. For pattern classification it is required that the output of the model is the classification of the sample.

Before discussing design and implementation, it is necessary to lay down some basic notation and terminology. For the sake of clarity, consider a dichotomy problem, as the principles laid down here can be extended to any finite number of classes; the majority of the problems in this study will be of this type.

Here the discrimination problem is formalised in such terms that black box model theory can be applied to it. Given a set of data $\{\mathbf{x}_i, d_i\}_{i=1}^N$, where the $\mathbf{x}_i$ are measurement vectors (electronic nose system output for example) and the d_i are the required responses (category of sample introduced to electronic nose); use this *training set* to categorise new measurements. This problem can be stated in terms of the *Interpolation problem*[64] :

Definition 3.2.1 Interpolation Problem for a dichotomy: *Given a data set* $\{\mathbf{x}_i, d_i\}_{i=1}^N$, *find a function* $F : \Re^N \longmapsto \{-1, +1\}$ *such that*

$$F(\mathbf{x}_i) = d_i \tag{3.1}$$

There are a number of different models that may be used in this context, each with their own set of inbuilt assumptions. In this thesis a number of different forms of deterministic models will be considered. Deterministic models are used under the assumption that the output is uniquely determined by the system input, stochastic models will not be considered.

3.2.1 Auto-Regressive, eXogenous input models

A typical type of black box model is one which employs the shift operator [65]. This type may take into account previous inputs and outputs. Due to this they are often written as transfer function models.

An important class of black box predictive models are linear systems of the general form

$$y(t) = \Phi^{\top}(t)\theta. \tag{3.2}$$

Here

$$\Phi(t) = \begin{bmatrix} \phi_1(t) \\ \vdots \\ \phi_n(t) \end{bmatrix} \tag{3.3}$$

is the input vector, $y(t)$ is the output and θ is the parameter vector representing the parameterisation of the system.

As with any proposed model it is normal to identify the parameters by fitting experimental data to a model of the form

$$y(t) = \Phi^{\top}(t)\hat{\theta} + e(t) \tag{3.4}$$

where $\hat{\theta}$ is the parameter giving the best fit to the experimental data and $e(t)$ is the error.

This fit is normally obtained by optimising a least squares cost function. This is optimisation by setting

$$e(t) = y(t) - \Phi^{\top}(t)\theta \tag{3.5}$$

and

$$e = [e(1), e(2), \ldots, e(m)] \tag{3.6}$$

for some m. The cost function

$$V(\theta) = \|e\|^2. \tag{3.7}$$

is minimised to find a solution.

In the case of a linear system the parameters are estimated using Least Squares Regression

$$\hat{\theta} = \left(\Phi^\top \Phi\right)^{-1} \Phi^\top y \tag{3.8}$$

A particular linear model defining the input(U)-output(Y) relationship in a black box model is an ARX (Auto Regressive eXtra input) model.

$$\mathrm{ARX}(q, r) : Y[n] = \sum_{i=1}^{q} \alpha_i Y[n-i] + \sum_{k=0}^{r} \gamma_k U[n-k] + e[n] \tag{3.9}$$

or in matrix notation

$$A(z)Y[n] = B(z)U[n] + e[n] \tag{3.10}$$

where $A(z)$ and $B(z)$ are polynomial matrices,

$$A(z) = 1 - \alpha_1 z^{-1} - \ldots - \alpha_q z^{-q} \tag{3.11}$$

$$B(z) = 1 + \gamma_1 z^{-1} + \ldots + \gamma_r z^{-r} \tag{3.12}$$

the orders of which are defined by q and r in Equation 3.9 above and the dimensions are dependent on the number of inputs and outputs. The α_i and γ_i are coefficients whose values are to be determined. The variable z represents the time series of inputs and outputs; the indice denotes the position in the time series. Here the autoregressive part is defined by the matrix A, and the exogenous part is defined by the matrix B. Due to this, Finite Impulse Response (FIR) models are considered by setting the order of A to one.

This form of model is particularly attractive to work with because it defines a linear regressive scheme. As such the minimisation of least squares is a well studied problem [66]. There are a number of ways of identifying parameters, such as gradient search and genetic algorithms.

The model assumes a number of things. Firstly that the system is linear in behaviour. Secondly, that the system is deterministic and is only determined by a finite history of input-output responses, i.e the system is finite and causal. Finally that the noise is zero biased and white in nature. This last assumption is necessary for least squares fitting to be an unbiased estimate of the maximum likelihood error.

3.2.2 Radial Basis Function Neural Networks

Not all systems exhibit a linear input-output behaviour. It is apparent that often the relationship between input and output, or the 'boundaries' between different data sets, will be nonlinear, especially given the reactive nature of some chemical substances. A generic, nonlinear, black box model is required to account for such systems. A good candidate is a model using Radial Basis Functions (RBFs); these incorporate nonlinearity but are linear in the free parameters, thus retaining many of the advantages of ARX models.

Radial Basis Function neural networks use a weighted sum system for interneural connections, but here the hidden layer is composed of a number of nonlinear functions with some interesting properties. These relate to parameter estimation, functional smoothness and separation of patterns.

To understand the definitions of objects related to RBFs, it is necessary to understand their origins. A justification for them results from Cover's Theorem [67]:

Theorem 3.2.1 Cover's Theorem *A complex pattern-classification problem cast in a high-dimensional space nonlinearly is more likely to be linearly separable than in a low-dimensional space.*

To illustrate this, let the input space be of dimension n_0 and let the target space

be of dimension n_1 where $n_1 \gg n_0$. Then a non-linear map is denoted as follows:

$$\Phi : \mathbb{R}^{n_0} \longmapsto \mathbb{R}^{n_1} \tag{3.13}$$

$$\Phi(\mathbf{x}) = [\Phi_1(\mathbf{x}), \ldots, \Phi_{n_1}(\mathbf{x})]^\top \tag{3.14}$$

where

$$\Phi_i : \mathbb{R}^{n_0} \longmapsto \mathbb{R} \tag{3.15}$$

The Φ_is are referred to as *hidden functions*, the reason for which will become apparent shortly, is that these do not need to be known explicitly. Following Cover's theorem, define separability as follows:

Definition 3.2.2 Φ-separability *[68]*

A dichotomy S_1, S_2 of a set $S \subset \mathbb{R}^{n_0}$ is said to be Φ-separable if there exists a vector $\mathbf{w} \in \mathbb{R}^{n_1}$ such that

$$\mathbf{w}^\top \Phi(\mathbf{x}) > 0 \qquad \mathbf{x} \in S_1 \tag{3.16}$$

$$\mathbf{w}^\top \Phi(\mathbf{x}) < 0 \qquad \mathbf{x} \in S_2 \tag{3.17}$$

Figure 3.2 shows this as a network where $\mathbf{w} = (w_1, w_2, \ldots, w_{n_1})$ is manifested as the output weights of the network, or as a function:

$$F(\mathbf{x}) = \sum_{i=1}^{n_1} w_i \Phi_i(x) \tag{3.18}$$

Essentially there are two 'ingredients' to Theorem 3.2.1:

- A nonlinear mapping of the input space by *hidden* functions.
- A 'hidden' space of much higher dimensionality than the input space.

In reality, arbitrary families of functions cannot be used for the following reasons. The evaluation of the weight vector $\mathbf{w}$ may become computationally difficult. The functions may amplify noise or disregard important variations in the data. Some *a*

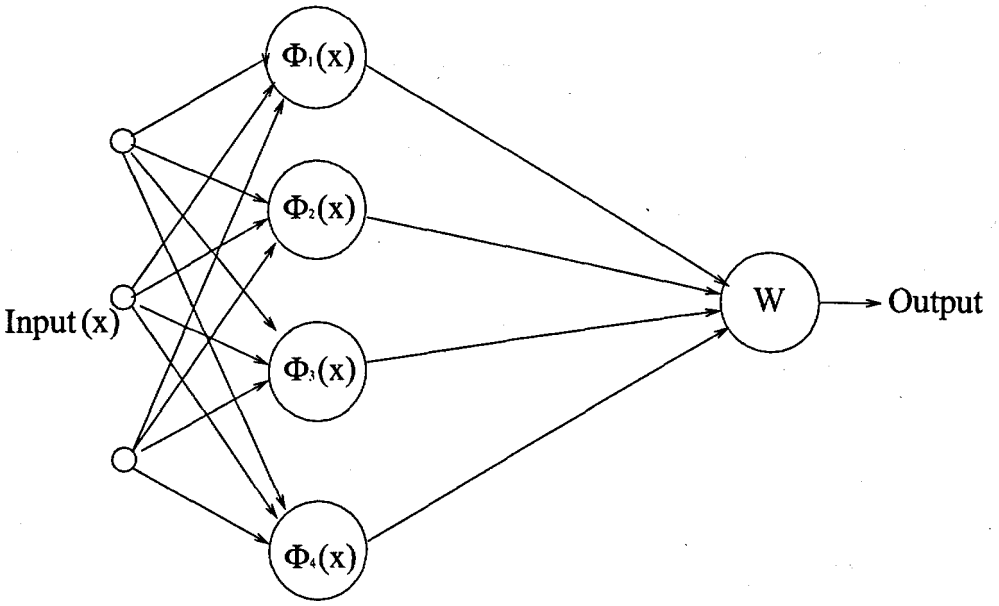


Figure 3.2: Diagram of network produced by Φ-separability

priori conditions have to be assumed about the solution. The most common assumptions are smoothness of the Φ_i's which controls the complexity of the resulting function F(x).

The weights are chosen to minimise an error function with a penalty on the curvature of the resulting function F; this will yield what is called a *regularised* model. In terms of the above, we define a *regularised* risk functional, weighted by λ, thus:

$$R_{reg}[F] = R_{emp} + \frac{\lambda}{2}\|PF\|^2 = \frac{1}{N}\Sigma_{i=1}^{N} c(F(\mathbf{x}_i), y_i) + \frac{\lambda}{2}\|PF\|^2 \qquad (3.19)$$

Here $c(.,.)$ is a cost functional that measures the the function F's deviation from the observed response y_i; that is the empirical error, or R_{emp}. Typically, the error squared would be used. P is a linear differential operator and embodies the smoothness constraints on the curvature. Effectively, RBFs, and their networks, are solutions to this constrained optimisation problem [69]:

Definition 3.2.3 Radial Basis Function. *A function, which minimises the regularisation constraint (3.19),* $G : \Re^n \longmapsto \Re$ *of the form:*

$$G(\mathbf{x}) = \Phi(\|\mathbf{x} - \mathbf{x}_i\|) \qquad (3.20)$$

is called a radial basis function.

Further:

Definition 3.2.4 Centre. *The vector* $\mathbf{x}_i$ *in equation (3.20) is called a centre.*

The solution is identified in terms of its weight parameters, $\mathbf{w}$. There is a wide range of functions that meet these criteria.

- **Multivariate Gaussian functions** $e^{\frac{-x^2}{2\sigma^2}}$
- **Multiquadratics** $\mathbf{x}^\top Q\mathbf{x}$
- **Multipolynomials** $\sum a(i,j,n,m)x_i^n x_j^m$

Gaussian functions are used in this thesis for two reasons. The first is the probabilistic shape of the function which can be used to isolate clusters. The second reason is that they are easy to visualise. This will become important in Chapter 4 when parameter estimation is considered.

3.2.3 Some considerations

The problem of estimating parameters in practice is that the estimation problem may be over-determined. Having a basis member for every data point is not very efficient. Given a set of input-outputs $\{\mathbf{x}_i, d_i\}_{i=1}^N$, the parameters required are found by inverting the matrix with entries $\{\Phi(\|\mathbf{x}_i - \mathbf{x}_j\|)\}_{i,j=1}^N$. Computationally this grows proportionally to N^3. It is possible that $N = 500$ but n_0=10, resulting in a grossly over-fitted curve that takes a significant amount of time to evaluate.

Secondly, by using every data point, no consideration is given to the quality of the data. Even with smoothness constraints, outliers will have an effect. Finally it would be very difficult to extract any rules from the result. A far better approach is to either:

- **Pick centres that are not necessarily data points.** Use an algorithm that chooses where best to place centres to express the distribution of the data.
- **Use data points that define boundaries of classes.** As the ultimate aim is to construct hypersurfaces between the dichotomy classes, use points from both classes that are "close" to this surface.

3.2.4 Support Vectors

The Support Vector algorithm selects a relatively small set of data points as centres, based on a complexity criterion. In the case of a linearly separable dichotomy, $\{\mathcal{H}_1, \mathcal{H}_2\}$, the algorithm attempts to fit a hyperplane of the form:

$$W^{\top}\mathbf{x} + \mathbf{b} = 0 \tag{3.21}$$

between the two classes.

Definition 3.2.5 Support Vector. *A vector* $\mathbf{x}_i$, *in the data set, satisfying:*

$$W^{\top}\mathbf{x_i} + b = 1 \tag{3.22}$$

where $W^{\top} = (w_1, w_2, \ldots, w_3)$, *or,*

$$W^{\top}\mathbf{x_i} + b = -1 \tag{3.23}$$

is called a ***support vector.***

Hence, data may be classified by which 'side' of the hyperplane each point lies. This is well defined once an orientation for the normal to the plane has been chosen. This produces a binary classification machine, the extension to a general number of classes is left for discussion in Section 3.5. The practicalities of finding such a hyperplane are now examined. Further, as applying Cover's theorem is of interest, this theory is extended to include a nonlinear mapping in Section 3.3.

Firstly place a cost on the size of the weighting vector in an attempt to reduce the complexity of the resulting hyperplane. This follows the theory presented in [70] where it is derived that the complexity of a linear neural network is proportional to the size of the node weights. It is also required, at the very least, that the hyperplane categorises correctly as many of the training data as possible. Therefore the problem of finding a hyperplane is formulated in terms of a *quadratic programming problem.* Let the training set be $\{\mathbf{x}_i, d_i\}_{i=1}^{N}$. Here $d_i \in [-1, 1]$ and it is required that

$$d_i(\mathbf{W}^{\top}\mathbf{x} + b) = 1 - \xi_i \tag{3.24}$$

The variables $\{\xi_i\}_{i=1}^N$ are referred to as *slack variables*. They measure the deviation of the data point from ideal conditions for separability. Hence the weights and the bias of the required hyperplane are the solutions to:

$$\arg(\min_W(\phi(\mathbf{W},\xi))) = \frac{1}{2}\mathbf{W}^\top\mathbf{W} + C\Sigma_{i=1}^N\xi_i \tag{3.25}$$

such that condition (3.24) is satisfied, with $\xi_i > 0$, for all $1 \leq i \leq N$. C is a weighting on the penalty incurred by the slack variables. To obtain the solution, the Lagrange multiplier technique is used to convert the system to a *dual* problem.

3.2.5 A short excursion: dual problems

Definition 3.2.6 Dual problem *The dual problem to a min/max problem is the max/min problem of the Lagrangian, considering it as a function of the Lagrange multipliers.*

By taking taking the conditions for a local (or hopefully global) stationary point of the Lagrangian function the following optimisation problem is obtained:

Consider

$$\mathcal{Q}(\alpha) = \Sigma_{i=1}^N\alpha_i - \frac{1}{2}\Sigma_{i=1}^N\Sigma_{j=1}^N\alpha_i\alpha_j d_i d_j \mathbf{x}_i^\top\mathbf{x}_j \tag{3.26}$$

subject to the constraints

$$\Sigma_{i=1}^N\alpha_i d_i = 0 \tag{3.27}$$

$$0 \leq \alpha_i \leq C \tag{3.28}$$

where $\{\alpha_i\}_{i=1}^N$ are the Lagrange multipliers.

The following relationship is also obtained:

$$\mathbf{W} = \Sigma_{i=1}^N\alpha_i d_i \mathbf{x}_i \tag{3.29}$$

By using the constraint (3.27) this can be reduced to an N-1 dimensional problem and solve it using an appropriate numerical method, such as the QUADPROG routine in MATLAB.

The consequence of this [‡] is that many of the α_i's become zero. The data points corresponding to the non-zero α_i's are thus the support vectors, that is the corresponding slack variable in (3.24) is zero. Hence this technique is called the Support Vector Machine (SVM).

3.3 Applying Nonlinear Maps to Support Vector Machines

So far pattern recognition has been considered in the original *input space*. However, following Cover's theorem it would be interesting to use a nonlinear transformation into a *feature space*. Consider a general map $\phi : \Re^{n_0} \longmapsto \Re^{n_1}$. That is, our input space is of dimension n_0 and our feature space is of dimension n_1. To solve the hyperplane problem in the feature space, simply replace all the '$\mathbf{x}_i$' vectors by '$\phi(\mathbf{x}_i)$' in Equation (3.26). This yields:

$$\mathcal{Q}(\alpha) = \Sigma_{i=1}^{N}\alpha_i - \frac{1}{2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\alpha_i\alpha_j d_i d_j \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) \tag{3.30}$$

subject to the constraints

$$\Sigma_{i=1}^{N}\alpha_i d_i = 0 \tag{3.31}$$

$$0 \leq \alpha_i \leq C \tag{3.32}$$

The relation

$$\mathbf{W} = \Sigma_{i=1}^{N}\alpha_i d_i \phi(\mathbf{x}_i) \tag{3.33}$$

is also obtained.

It still is difficult, in practice, to apply this to families of RBFs when using the mappings ϕ. It is desirable to use RBFs because they have 'nice' properties such as smoothness and rotational symmetry. Looking closely at the expressions above, it can

[‡]This is due to the solution to a constrained optimisation problem always lying on a boundary representing one or more of the constraints.

be seen that we only require the inner product of the functions ϕ, which from now on is written using the $\langle , \rangle$ notation. To develop this further, the concept of a kernel [71] is introduced.

3.3.1 Kernels

Definition 3.3.1 Inner-Product Kernel. *A function* $K : \Re^{n_0} \times \Re^{n_0} \longmapsto \Re$ *is an inner product kernel if it can be written in the form*

$$K(\mathbf{x}, \mathbf{x}_i) = \langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle \tag{3.34}$$

where $\phi : \Re^{n_0} \longmapsto \Re^{n_1}$

Thus, it can be seen that inner-product kernels can be used in SVMs as (3.33) implies

$$W^\top \phi(\mathbf{x}) + b_0 = \Sigma_{i=1}^{N} \alpha_i d_i \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}) + b_0 \tag{3.35}$$

where b_0 is a parameter called the *bias*. This can then be evaluated for $x \in \Re^{n_o}$.

RBFs are such kernels [69] and so can replace the inner product $\phi(x_i)^\top \phi(x_j)$ in (3.30) to give the Lagrangian:

$$\mathcal{Q}(\alpha) = \Sigma_{i=1}^{N} \alpha_i - \frac{1}{2} \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} \alpha_i \alpha_j d_i d_j \Phi(\|\mathbf{x}_i - \mathbf{x}_j\|) \tag{3.36}$$

Thus with the set of minimising Lagrangian multipliers, $\{\alpha_i\}_{i=1}^{N}$, the neural network, expressed as a function, becomes.

$$\langle W, \phi(\mathbf{x}) \rangle + b_0 = \Sigma_{i=1}^{N} \alpha_i d_i \Phi(\|\mathbf{x}_i - \mathbf{x}\|) + b_0 \tag{3.37}$$

RBF networks encompass a number of assumptions, chief amongst them being proximity implies similarity and the system is nonlinear. They also assume that smoothness, manifesting low complexity, is good. This is controlled by the width parameter; setting this parameter is of the utmost importance when implementing this family of models (see Chapters 4 and 5).

It has been suggested in [63] that RBFs might prove useful for time series analysis. A model that encompasses the history of the system, such as ARX type models do, while allowing nonlinear behaviour would be very useful.

3.3.2 Non-linear ARX type model using Gaussian Kernels.

The theory for nonlinear black box models may also be found in articles such as [72] and [73]. The acronym NARMAX was coined for use with a generic nonlinear form of ARMAX model. Here a novel way of combining the concept of the ARX type model with RBFs, which was developed by the author, is demonstrated. This results in a model whose parameters may be identified using the Support Vector Method, the NARX, the advantage being that a model is generated that is endowed with robust parameter estimation.

If the form of Equation (3.9) is considered then it may be thought of in the Many-In-Many-Out (MISO) case as being

$$\alpha_1 y(n) + \alpha_2 y(n-1) + \ldots + \alpha_{q+1} y(n-q) = \gamma_1 . u(n) + \gamma_2 . u(n-1) + \ldots \\ \ldots + \gamma_{r+1} . u(n-r) + e(n) \quad (3.38)$$

Here the input as a vector of the form

$$u(.) = \begin{bmatrix} u_1(.) \\ u_2(.) \\ \vdots \\ u_d(.) \end{bmatrix} \quad (3.39)$$

By rearranging (3.38) the following is obtained:

$$\alpha_1 y(n) = \gamma_1 . u(n) + \gamma_2 . u(n-1) + \ldots \\ \ldots + \gamma_{r+1} . u(n-r) - \alpha_2 y(n-1) - \ldots - \alpha_{q+1} y(n-q) + e(n) \quad (3.40)$$

or

$$\alpha_1 y(n) = [\gamma_1, \ldots, \gamma_{r+1}, -\alpha_2, \ldots, -\alpha_{q+1}] . \begin{bmatrix} u(n) \\ u(n-1) \\ \vdots \\ u(n-r) \\ y(n-1) \\ \vdots \\ y(n-q) \end{bmatrix} + e(n) \quad (3.41)$$

This allows us to place the model in the context of a Support Vector Machine using Gaussian kernels. The dot products above can be replaced with kernel products of the form

$$< x, y >= exp(-\frac{\|x - y\|^2}{2\sigma^2}) \tag{3.42}$$

Where x and y are two vectors of the same dimension and σ is commonly referred to as the width parameter.

Hence what is obtained is an RBF network whose input is the column vector on the right hand side of equation (3.41) and the target output is $y(t)$. The parameters α_i and γ_j are found using the SVM algorithm provided in the previous section.

3.4 Multi-Layer Perceptron

This class of neural network uses linear weights and a smooth activation function for its nodes [68]. The typical architecture is of an input layer, one or more hidden layers, and an output layer (see Figure 3.3). Neural connections are again defined by weighted sums. Training is via the error back-propagation algorithm [74]. During each iteration of the algorithm there are two passes through the network. In the first *forward* pass the weights are fixed and an activation pattern is passed to the input layer. The signal propagates through the network, resulting in some output. This output generates an error due to there being a target output. In the *back* pass the error is used, along with some learning rule, to adjust the weights so that the network output comes to resemble the target output.

The error generated for the back pass is commonly the mean squared error. A gradient descent method is derived from this error and used to adjust the weights. Adjustment may occur after each training example or after all of the training set has been used; this is called an epoch.

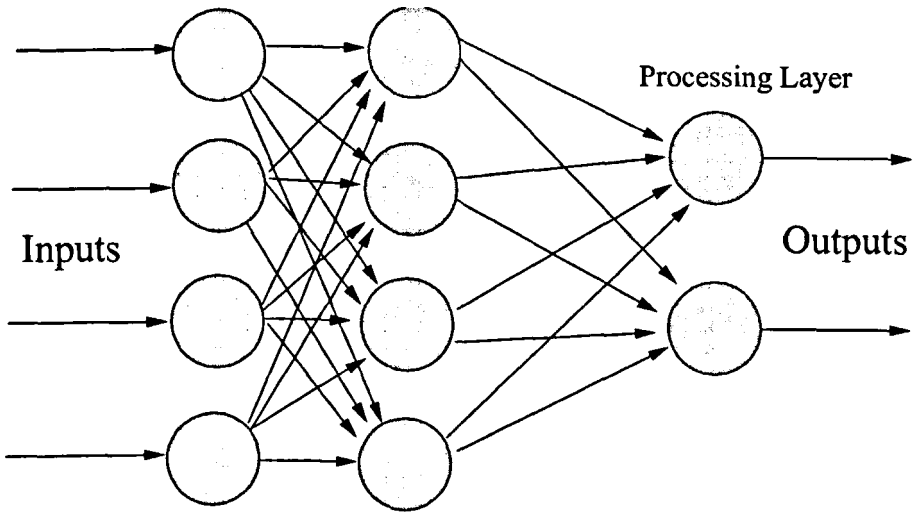


Figure 3.3: Schematic of a fully connected Multilayer Perceptron Architecture with a processing layer.

3.5 Applicability to Electronic Noses

Sample analysis and identification are two of the prime aims of electronic nose research. Here models are used in a pattern recognition context, and so are commonly parametric. However a survey of the literature reveals that only a minority of papers are concerned with the parametric mode of modelling of electronic noses. Usually the statistical/black box model analysis is concerned with the static, steady state response of sensors. Mainly it is statistical analysis in the form of PCA for data visualisation [75] which only demonstrates the potential for classification. However, as PCA results in a linear projection, it only demonstrates whether a linear model might be successful in separating the various classes.

A number of applications of fuzzy logic to discriminant analysis have been reported [54] [76] though it is often unclear what kind of fuzzy system has been used. These methods do perform better than their linear counterparts in the presence of curvilinear clusters, these being common with nonlinear sensor responses.

Multiclass Radial Basis Neural Networks may be implemented in a number of ways, such as: real number tags; binary output; dedicated class identifiers. In the first method

each class is attributed a number, for example 1,2 3, and 4. A single RBF NN is then trained to give the correct output for each training sample. However it would seem better to use the SVM theory discussed above to ensure robust parameter estimation. A binary output might be used where $N = \log_2 M$ networks are implemented for M data classes. Each class is then given a binary string label and the networks are trained to each output a bit of this sequence. This is a plausible as it has been shown [77] that any data set is separable using the SVM algorithm. The final method which tends to be found in the literature is to have $N = M$ neural networks: each of which identify a separate class positively. The last methods two are examples of committee machines [68].

As for processing the input of the model, in [78] a number of interesting trends are reported. The first is the extraction of parameters from the dynamic response of a sensor. Sorption (maximum rate of change), desorption (maximum negative rate of change), divergence (total step change) and area under the response curve were used as outputs for each sensor. This was a move forward in data processing. More than this is the use of genetic algorithms (see Chapter 5) for 2 processing layer multiperceptron neural network parameters.

Another use of modelling is for 'information transfer'. The problem of information transfer between electronic nose devices has existed as long as the devices; however it is only recently that the problem has been approached. The basis of the problem is very similar to that of drift and sensor poisoning. The problem, as alluded to in the previous chapter, is that the manufacture of devices is not consistent enough for reproducible responses between different sensors of the same original design. Several approaches have been applied, using standard odours to compare sensor responses. One method [79] uses the sensor responses to construct a transformation to apply to data resulting from one array of sensors to another of the same design. Several different maps have been considered though all are empirical. Hence there is plenty of scope for new map applications.

3.6 Mechanistic Models

A mechanistic model is one which embodies some 'physical' knowledge of the system considered, such as chemical reactions. This is most commonly achieved by an appropriate system of differential equations: a common assumption being time and space invariance of the evolution of the model leading to specification of the derivatives only. By their nature these types of models often result from the necessity to test some scientific theory in order to understand a system more fully.

Mechanistic models have an advantage over black-box models in that the initial modelling assumptions may help to reduce the complexity of the model, with regard to the number of free parameters that require estimation; the model is 'tailored' to a specific set of systems which the assumptions characterise. A black-box model, including the transfer function types considered in Section 3.2.1, have to be flexible enough to describe a large number of different systems' behaviours. This may lead to such problems as over-fitting (see next chapter).

Another advantage of mechanistic models is the 'meaning' of the system parameters. The parameters in a mechanistic model have some abstract significance attached to them as they denote the previous knowledge gathered at the model building stage. This allows the complexity of model to be reduced. Various aspects of the model may be ordered with respect to the contribution they make to the phenomenological behaviour of the system. Knowledge of the parameters' significance may allow bounds to be placed upon the possible values they may take; this has been demonstrated to influence the complexity of a model [80]. The physical meaning of parameters also allows validation of the model against previously experimentally observed values.

Mechanistic models still require validation and this is normally achieved by parameter identification from experimental data. The validation test is normally of the form that the model can be made to 'fit' the system's response to some input. A comparison is made of the model's responses to the observed system behaviour. Requirements are normally made about repeatability of the parameter estimation. The requirement that a specific behaviour should be represented by a unique set of parameters. These criteria will become important in Chapter 10 when a mechanistic model is validated.

3.6.1 Application to Electronic Noses

There are still few studies of the physical mechanisms of electronic noses or the time-dependent sensor outputs from mass spectrometers. The justification for this kind of investigation is the increased information content of the sensor transients. There is evidence [81] [82] that this is the case. Importantly [81], a dynamic model allows the output of the system to be modified to account for short term drift factors such as sensor temperature, if parameters were shown to be time or environmentally dependent.

The dynamics of tin oxide resistive sensors have been modelled using reaction-diffusion equations [83]. It was assumed that the bonding reaction occurred much quicker than diffusion. This adsorption model was coupled with a conductance model for common electrode configurations to provide a prediction for the observed conductance transients. Simulations were provided and an analytical approximation was derived which corresponded well with the simulations. The bonding aspect of the model was extended to a nonlinear model [61] to allow for the nonlinear phenomena which had been observed experimentally. It was noted that the transients of a gas sensor could be classified by the relative rates of reaction and diffusion. In both cases, though simulations predicted a general trend that agreed with experimental observations, there was no validation via parameter estimation.

The class of conducting polymer resistive sensors has also been investigated [84]. This too was a reaction-diffusion model coupled with a proposed conduction expression. Again an extension to account for nonlinear isotherms was made [85].

Parameter estimation to identify these types of models requires modification to the original expression [86]. A time delay is incorporated to allow for a delay between sample introduction and a response in the transient conduction. However there was some simplification as, instead of the full tin oxide expression that had been derived previously, a two exponential approximation was applied. This particular article contains two points of extreme interest which resulted from the analysis of model parameters,via neural network models, to identify samples. The first is the removal of correlating sensor data; the second is that the most important parameters in the model were identified, with respect to the amount of information they contained about the

nature of a sample. These were the exponential time constants. This was measured via the success rate of the neural network models identified using the parameter values.

It can be concluded that there is evidence that an analysis of the transients of gas sensors will yield more detail on the nature of a sample to which they have reacted. There is an opportunity to understand more fully the environmental effects on gas sensors and the way that repeated exposures affect the response of sensors.

3.7 Parameter Estimation Needs

To couple the theoretical model types discussed above with experimental data requires parameter estimation. Usually this is referred to as 'curve fitting' where some prescribed error function is minimised.

If an algebraic expression which predicts the output, given a certain input, is available then quite often it is a case of adjusting the parameters in this expression so that the model curve is as close to that prescribed by the data set as possible. Error is often measured as a sum of the square errors: often a penalty clause is added for constraints on parameter size.

If the model output is in some implicit form, such as a system of differential equations, then other methods are necessary. Numerical integrators are one method of producing model responses based upon a given parameter set. Parameter estimation in these cases is an iterative process starting from some reasonable start values. A number of parameters are varied and the model resimulated with these values; the best variants, those that reduce the fitted error, are chosen to be new parameter values. This process is continued until some stopping condition is reached, such as a maximum number of iterations, or a desirable error measurement is reached, or minimised.

All of these methods rely upon the data provided and the complexity of the models. Factors such as noise, sparseness or otherwise of the data set, and whether the package can handle that particular type of model all affect the validity of the identification. These factors will be investigated in the next two chapters.

3.8 Conclusion

In the above the various approaches to modelling that are available to analyse electronic nose systems have been examined. It must be noted that there are two ways of approaching the problem of modelling a system's behaviour: black box and mechanistic modelling. Each have their advantages and disadvantages, as discussed, and these should be borne in mind when at the model building stage.

Of particular importance is the *reason* for building a model, is it predictive, is it discriminatory, is it to test a hypothesis about the internal structure of the system? The assumptions made for a particular model need to be examined. This is especially so if the model fails validation; the assumptions may be re-evaluated, modified and the model redesigned.

It is noted that there are two aspects to building a model; the production of a tentative, usually parameterised, model class and the identification of the particular model by parameter estimation. If a third stage model building is taken to be validation with a previously unseen data set, then it can be seen that these three stages are very much interlinked. These ideas will be developed further in the next chapter.

There is also the fundamental question of choosing which particular input-output behaviour it is desirable to model. There may be a whole plethora of factors affecting a system, but which are necessary for an understanding of the system? The application of electronic noses is mainly in categorising a sample from its odour and so it is asked what aspects of a system it is necessary to understand in order to bring this about.

These ideas will be explored in the course of the rest of the thesis.

Chapter 4

Model Identification I: Parameter Estimation and Sparse Data Sets.

Black box models, in particular neural networks, are trained, or tuned, using finite data sets. The aim is to find an algorithmic representation of the system considered. Numeric schemes are normally used and these algorithms can be seen as inductive techniques [87]. A number of observations of the phenomena of interest are made and the algorithm produces a prediction, or hypothesis. They are inductive for the main part because of the iterative schemes they normally employ. A first guess is produced, its fitness as a model is assessed and it is adjusted accordingly.

Neural Networks represent a parameterised family of functions. It is these parameters that need to be estimated using some optimisation algorithm, coupled with experimentally observed data. The problem is that the estimates are made using a finite data set. These data provide empirical information for a hypothesised model which links input and output behaviour of the system. This can result in two problems that come under the broad umbrella of over-fitting; the first is that any function that interpolates the points will appear to be a 'good' estimation of the function; secondly a large amount of what is 'learned' might be noise. It is because of this finite sampling that the process is considered as a probabilistic one.

For this reason the aspects of model building regarding how stable the training algorithm and the resulting Neural Network are with respect to the probability structure of

the systems output are analysed. In other words; how much does the identified model change with respect to perturbations of the training data; how is the discriminant accuracy of the model affected by overlap of the data classes; how must the complexity of the model change to deal with this overlap?

The last point is important in as much as it has been demonstrated [5] [88] that the data needed to find a good approximation model increases rapidly with the complexity of the chosen model family; by complexity it is meant the number of free parameters. It is fairly evident that a more sophisticated model is necessary to approximate highly convoluted class boundaries. For example, some nonlinear model using splines may be considered instead of a linear model. It is to this end that parent distribution bounds on the complexity of the model are investigated to assess how parameter estimation algorithms are affected by data class overlap.

Electronic nose systems output data which are often of a high dimension, sparse and noisy. It is for these reasons that the effects of noise and complexity are analysed. An understanding of these effects will aid the interpretation of the results in Chapters 6 and 7 where black box techniques are used to analyse experimental data from electronic nose systems.

In the previous chapter the use of nonlinear black box models was examined, specifically RBFs. It is with the application of these functions, and the resulting networks, that this chapter concentrates on. Here a type of RBF, which are rotationally invariant nonlinear functions, and so normally incorporate some width parameter σ, are considered:

$$K_\sigma(x_i, x_j) = K\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{4.1}$$

It will be observed in Chapters 6 and 7 that real medical applications produce data with a great deal of noise (both as interference and 'chemical noise' as was described in Chapter 2), presenting significant challenges for discriminant model design. If electronic noses are to prove successful under such conditions, the effects of noise on discriminant models must be understood.

In this Chapter, these effects will be investigated via statistical modelling and numeric experiments carried out by the author. As in all good experiments the noise will

be controlled. The effects of controlled noise need to be observed so that, when a particular model fails to fulfil the criteria set it with respect to a real system, model failure may be understood. The extremes of model building need to be understood so that modelling techniques are used to analyse systems that they are capable of synthesising.

It will be noted how the data requirements for parameter estimation grow with the complexity of the model. Complexity grows with the input dimension of the model. Thus methods will be examined for reducing the dimension of the input data. This is justified in terms of the robustness of parameter estimation and correlating signals.

4.1 Accuracy of a classifier

The empirical accuracy of an artificial neural network is well known, it is the number of correct classifications over the total number of data points presented. However this is empirical and so this is extended to a definition that includes the expected accuracy (that is if an infinite number of tests were carried out) of the network:

Definition 4.1.1 Accuracy of a Classifier *The accuracy of a binary classifier neural network* $f(x) : X \longmapsto Y$ *on a subset of* $X \times Y$ *where* $Y = \{-1, 1\}$, *such that the ordered pair* (x, y) *represents the input and target outputs ,is defined as follows.*

$$Accuracy[f] = 1 - \frac{1}{2}\mathbb{E}[\|f(x) - y\|] \tag{4.2}$$

Note (that by the law of large numbers, and assuming an ergodic source), the empirical accuracy will converge to (4.2) point wise on the function space as the empirical data set grows in size [77].

This measure of accuracy of a model can be extended to a weighted error dependent upon the application. Note the factor of a half is included as a misclassification results in an error of 2. This is presented in the probabilistic models of this chapter.

4.2 Bounds on Performance

Of primary interest is the interaction between the empirical data, the training algorithm and the model family. That is, of interest is the information provided by the data to

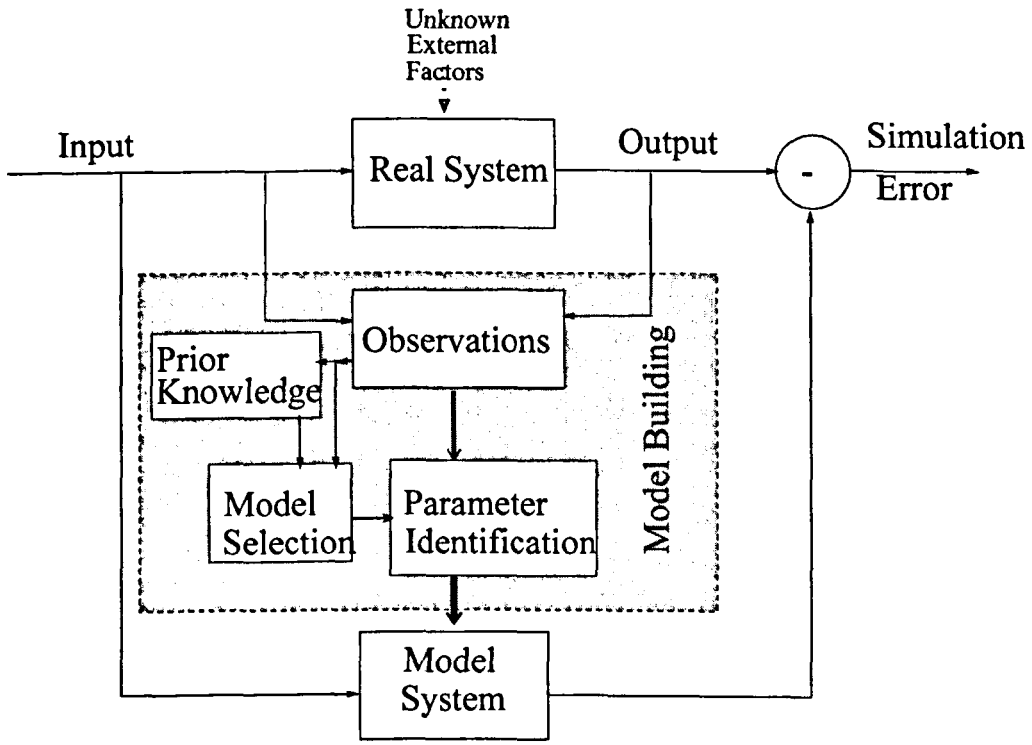


Figure 4.1: Schematic of the model building process. Here the experiments on the real system, model building and model validation are represented.

the model design process, the stability of the system identification algorithm and the generalisation ability of the model. The model building process is shown in Figure 4.1. By resolving power of the data it is meant the amount of information the data provide to discriminate between two or more classes. This is not readily measurable, but it is obvious that what is measured and how many samples of these measurements are taken both contribute to the 'quality' or resolving power of the data set. These include such things as the amount of noise, outliers and the amount of information that has 'filtered' through the system from the original sample. In the context of electronic noses it is taken to mean that, initially, the classification of a sample is absolute, it is either one thing or another. However, this classification is abstracted by sampling the headspace via gas sensors and data logging equipment.

There are theoretical, as well as practical, considerations for the resolving power of

the data. In this section the interaction between the complexity of the network and the size of the training set is considered. The 'learning problem' is formalised and ways that the accuracy may be theoretically bounded are reviewed.

The object of training a neural network using a sample of 'typical' data is to estimate the underlying discriminatory function that it is hypothesised lies inside our concept class. A concept class is a parameterised class of functions considered for identification. In [5], and elsewhere, the required discriminant function is referred to as the *target function.* It is the hypothetical indicator that discriminates perfectly between samples. However, as it shall be seen, this is not a well defined function in general and hence is not attainable. Normally the concept class can only be assumed to be a very large class of indicators.

In using a neural network the search for a model is restricted to a particular family of parameterised functions, which is defined to be the *Hypothesis Class.* This is a matter of convenience as the list of free parameters uniquely determines the function. This bijection between real vectors and the hypothesis class allows us to train, store and analyse neural networks in an efficient manner. It also allows the complexity of the neural network that is sought to be adjusted [77].

Consider a nesting of parameterised families, sorted by complexity. This nesting is denoted by $\mathbb{H}_n$, where the number of free parameters is proportional to n: for example, in a neural network n might denote the number of hidden nodes. The nesting might be denoted by:

$$\mathbb{H}_1 \subseteq \mathbb{H}_2 \subseteq \ldots \subseteq \mathbb{H}_n \subseteq \ldots \tag{4.3}$$

By considering a general training regime using least squares, the error, $I[.]$, may be written [5]

$$I[f] \equiv \mathbb{E}[(y - f(x))^2] = \mathbb{E}[(f_0(x) - f(x))^2] + \mathbb{E}[(y - f_0(x))^2] \tag{4.4}$$

where $f_0(x)$ is called the *regression function,* defined by

$$f_0(x) \equiv \int_Y yP(y|x)dy \tag{4.5}$$

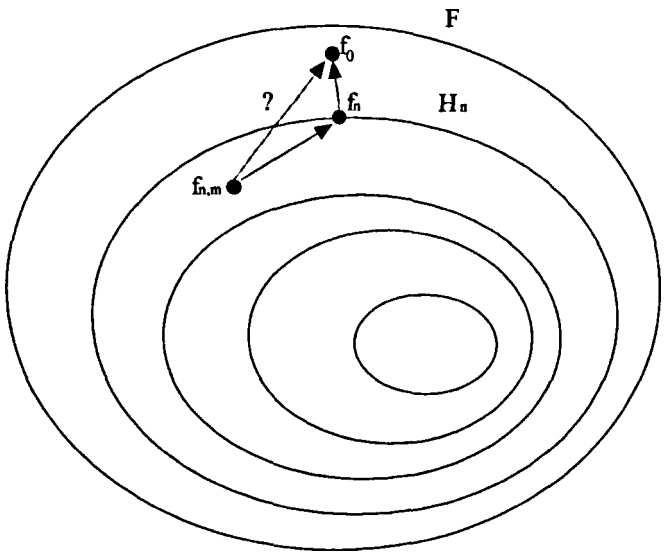


Figure 4.2: Representation of the learning problem. The boundaries represent the nested classes of functions. The outer-most is F, the concept class. Reproduced from [5]

where $P(y|x)$ is the conditional probability of a sample having label y given the measurement x.

From (4.4) it is obvious that the regression function minimises the expected error; it is the well defined function which resembles the input-output behaviour of the system the closest. However it can also be seen that the resulting error will be equal to $\mathbb{E}[(y - f_0(x))^2]$. This is the *variance* of the response, whereas the first term in (4.4) is the *bias* of the function. The variance, as will be seen, is a result of additive noise in the system.

Two properties are desirable: firstly that as n is increased, and so the potential complexity of the network, it is hoped that the hypothesis class will come 'close' to $f_0(.)$. Secondly, it would be desirable that as the number of training points is increased, m, in a fixed hypothesis class, then the empirical minimiser will converge to the best approximation in the class. If we call the empirical minimiser $f_{n,m} \in \mathbb{H}_n$ then these concepts are illustrated in Figure 4.2.

In [5] the three errors shown on this diagram were formalised in the following way:

Definition 4.2.1 Generalisation Error *The error between the learner hypothesis and*

the target function, found by minimising the empirical error, is called the ***Generalisation Error.***

Definition 4.2.2 Approximation Error *The error between the best hypothesis h_n and the target function is called the* ***Approximation Error.***

Definition 4.2.3 Estimation Error *The error between the learner and the best hypothesis is called the* ***Estimation Error.***

Hence the estimation error is due to the finiteness of the training set, and the approximation error is due to the assumptions built into the neural network topology.

Definition 4.2.4 Sample Complexity *The amount of data needed to minimise the estimation error is defined to be the* ***Sample Complexity;*** *there is no exact measure of this except that the number of data samples required to minimise the error will grow with the complexity of the model.*

It would be expected, as the training set is enlarged, that the estimation error is reduced. However, it is only known that the empirical risk is 'close' to the expected risk for each given function. In other words, the convergence is not necessarily uniform. Thus the fact that the empirical risk converges, by the law of large numbers, to the expected risk does not guarantee that the minimiser of the empirical risk converges to that of the expected risk. However it is shown in [5] that under certain conditions, that is by controlling the complexity of the model class with respect to the amount of empirical data available, this convergence is guaranteed.

An immediate consequence of this is that the size of the training set must grow more quickly than the complexity of the hypothesis class. There is a relationship between these two quantities which provides an upper bound on the generalisation error. It is thought that the complexity of model necessary is greatly affected by 'overlap' of the two classes being considered-the greater the overlap, the more refined the model needs to be. This will be observed when considering algorithm independent performance with regard to parameter estimation.

The difficulty of building a classifier stems from the fact that the training regime is an ill-posed interpolation problem. Any functions that interpolate the training points

will have an empirical risk of zero, for example in [89] it is shown that Gaussian kernels are universal approximators. Hence, as the size of the training set is increased, an ever more convoluted function may be produced that fails to encompass the probabilistic structure of the parent distribution. This is due to the ill-determined nature of the problem. There is an equivalence relation on the space of real functions, where two functions are equivalent if they are equal in value at each of the training points. To avoid this an assumption about the class of functions considered is made. This also provides the framework for producing a training algorithm and ensuring unique parameter values for the network.

For RBF Neural Networks the assumptions are mainly concerned with smoothness [69] and complexity (see previous chapter), via the number of tunable parameters. Built into this is the implicit assumption that the regression function is realisable in this class of functions. Of course, with kernels any function is realisable, but it can only be estimated empirically.

The width parameters and the weights interact to set the smoothness. What is of great importance for the complexity of such neural networks is the size of the weights as demonstrated in [80].

4.3 Noise and Drift

It is necessary to consider how various pattern recognition techniques are affected by noise and sensor drift. By noise is meant the part of the signal that is due to environmental factors or information in the sample that is of no interest with respect to the system model design. This is assumed to result in some distribution mixture model.

From (4.4) it is apparent that the minimum error achievable is that when the regression function is used. How is the accuracy of this function affected? From its definition in (4.5), error occurs in regions where the distribution functions of the two classes overlap.

There is a specific measure of overlap already in existence- the signal to noise ratio. The relationship between this and the regression function's theoretical accuracy will

now be examined.

4.3.1 Mutual Subset-hood

The probability distributions of the classes, $P(x|y)$, can be thought of as membership functions as in fuzzy logic. Here, how much these membership functions overlap is considered.

A membership measure is a fuzzy measure of 'how much' an element belongs to a fuzzy set. A fuzzy set is a set with such a measure upon it.

In the cases considered here, $P(x|y)$ is the probability measure of how likely ('how much') it is that the measurement x would occur given that it resulted from a sample of class y.

First some measures on membership functions are introduced:

Definition 4.3.1 Cardinality of A Fuzzy Set. *The cardinality, $c(A)$, of a fuzzy set A with membership function $a(x)$ is defined as.*

$$c(A) = \int_A a(x)dx \tag{4.6}$$

Definition 4.3.2 Intersection *The intersection of two fuzzy sets is defined by the membership function:*

$$(A \cap B)(x) = min\{a(x), b(x)\} \tag{4.7}$$

Definition 4.3.3 Union *The union of two fuzzy sets is defined by the membership function:*

$$(A \cup B)(x) = max\{a(x), b(x)\} \tag{4.8}$$

Conceptually a measure is desired which measures the similarity of the two sets, this is the mutual subsethood and is denoted $E(A, B)$. That is to say the mutual subsethood should measure:

$$E(A, B) = \text{Degree}(A = B) = \text{Degree}(A \subseteq B \text{and} B \subseteq A) \tag{4.9}$$

where *Degree*(.) is a conceptual measure of the similarity of two sets.

The derivation of $E(.,.)$ is omitted but the mutual subset-hood measure is defined as:

Definition 4.3.4 Mutual subset-hood *The Mutual subset-hood of two fuzzy sets A and B is defined as*

$$E(A,B) = \frac{c(A \cap B)}{c(A \cup B)} \tag{4.10}$$

Consider now a mixture model of two normal distributions. In Figure 4.3 a mixture of $P_{\mathbb{C}_1} = N(-5,2)$, which is labelled $(\mathbb{C}_1)$, and $P_{\mathbb{C}_2} = N(3,3)$, labelled $(\mathbb{C}_2)$ is shown. These distributions are defined as:

$$\begin{aligned} P(x|x \in \mathbb{C}_1) &= P_{\mathbb{C}_1}(x) \\ P(x|x \in \mathbb{C}_2) &= P_{\mathbb{C}_2}(x) \end{aligned} \tag{4.11}$$

It is assumed that these two distributions are known, and so the regression function is known. It is also assume that the two classes are labelled using binary labels $\{-1,1\}$. In area A, members of $\mathbb{C}_1$ will be classified correctly. This is also true for $\mathbb{C}_2$ in Region D. However in region B, elements of $\mathbb{C}_2$ will yield a squared error of 4, as will elements of $\mathbb{C}_1$ in region C. Hence it can be seen from 4.4 that the total squared error of the regression function will be

$$\mathbb{E}[(y - f_0(x))^2] = 4\int_B dP_{\mathbb{C}_2}(x) + 4\int_C dP_{\mathbb{C}_1}(x) \tag{4.12}$$

$$= 4c(A \cap B). \tag{4.13}$$

Hence we also find that

$$\text{Accuracy}[f_0(x)] = 1 - \frac{\mathbb{E}[(y - f_0(x))^2]}{4} \tag{4.14}$$

$$\leq 1 - \frac{1}{4}\frac{c(\mathbb{C}_1 \cap \mathbb{C}_2)}{c(\mathbb{C}_1 \cup \mathbb{C}_2)} \tag{4.15}$$

$$= 1 - \frac{1}{4}E(\mathbb{C}_1, \mathbb{C}_2). \tag{4.16}$$

Thus an upper bound on the accuracy of the regression function, and so the maximum discriminant accuracy that can be obtained when discriminating between two

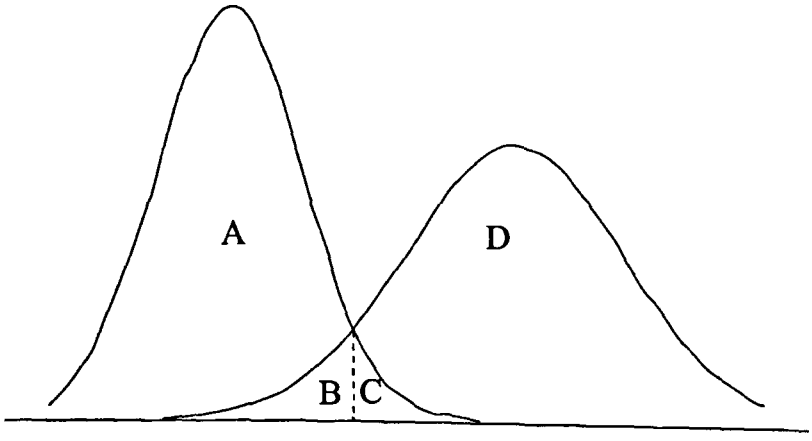


Figure 4.3: A mixture model of two normal distributions

set, is set by the mutual subsethood of the two classes. The larger the subsethood, the lower the accuracy of the regression function.

It is with this observation that the mutual subsethoods of two different distributions are considered.

4.3.2 Informal Concepts

In the case where the membership functions are probability distributions then $c(A) = c(B) = 1$. Hence the mutual subset-hood measure reduces to

$$E(A,B) = \frac{c(A \cap B)}{2(1 - c(A \cap B))} \tag{4.17}$$

Considering the case of unimodal distributions, it is apparent that $c(A \cap B)$ is a manifestation of the signal to noise ratio of the mixture model. Hence the mutual subset-hood is as well. This will be examined in the next section.

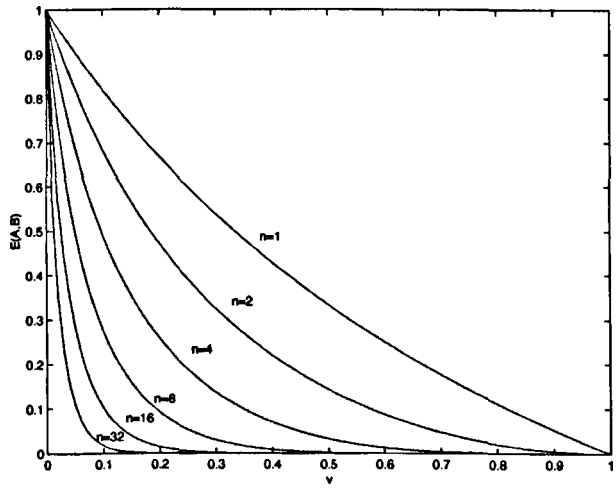


Figure 4.4: Mutual Subsethood for different dimension n and class separation λ.

4.3.3 Two examples of Mutual Subset-hood

To illustrate the above points consider two mixture models- two uniform distributions and two normal distributions. The extent to which overlap can bound the theoretical accuracy of any pattern recognition algorithm is demonstrated.

In an n-dimensional uniform distribution, points are distributed with equal probability on the set $[0,1]^n$. Consider the case where $n = 32$. Two classes are mixed; one distributed on $[0,1]^n$ and the other on $[\nu, 1+\nu]^n$. The parameter ν is referred to as the *separation parameter.* The mutual subsethood is thus

$$E(A,B) = \frac{(1-\nu)^n}{2-2(1-\nu)^n} \tag{4.18}$$

A plot showing the influence of ν and the dimension, n, is shown in Figure 4.4. The first thing to note is that this is a monotonically decreasing function of ν on the interval [0,1]. The second is that it is also a monotonically decreasing function of the dimension n. This alludes to Cover's theorem of separability in higher dimensions.

For a discussion of the effect on a normal mixture model the dimension is restricted to one in order to produce an intelligible expression. The classes are again labelled $\mathbb{C}_1 \sim N(\xi_1, \sigma)$ and $\mathbb{C}_2 \sim N(\xi_2, \sigma)$. It is assumed that the variance of the two classes is equal, and is of value σ. Further, it is assumed that for the means, $\xi_1 \leq \xi_2$. This is so

as to make the resulting expression clear enough to glean some insight into the effect noise has on the overlap of two classes. A similar derivation is to be found in [90], but there the bell shaped Gaussian distributions have not been normalised to have area under the curve equal to 1.

Clearly the boundary of regions B and C in Figure 4.3, where one class becomes more probable than the other, is at $x = \frac{1}{2}(\xi_1 + \xi_2)$. Hence it is deduced that

$$E(A,B) = \frac{1 - \text{erf}[\frac{\xi_2 - \xi_1}{2\sqrt{2}\sigma}]}{1 + \text{erf}[\frac{\xi_2 - \xi_1}{2\sqrt{2}\sigma}]} \tag{4.19}$$

where

$$\text{erf}[x] = \frac{2}{\pi}\int_0^x e^{-r^2} dr \tag{4.20}$$

It is noted that $\frac{\xi_2 - \xi_1}{2\sqrt{2}\sigma}$ is half the signal-to-noise ratio, so that the mutual subset-hood is a function of the signal to noise ratio. This is possibly due to the fact that a decrease in the signal-to-noise ratio will result in the two distributions overlapping each other, the result being misclassifications in both classes.

Two things are observed: the first is a manifestation of Cover's theorem which states that a data set becomes 'more separable' in higher dimensions. This may be observed in Figure 4.4. The second is that the modelling task is dictated by the probabilistic nature of the classification problem. The mutual subsethood measure demonstrates that total separation of two data classes may be impossible depending upon the signal to noise ratio. It has also been demonstrated that the theoretical ability of a model to discriminate between two classes, and so the discriminate power of the data themselves, is dictated by the additive noise within the system. It should also be noted that, looking at the intersection of the two data classes in Figure 4.3, the regression function will become less smooth as the signal-to-noise ratio increases. This means that the target function within a given hypothesis class will be consequently less smooth, consequently more complex and so consequently more data is required to identify the model.

It is apparent that there will be some optimal model that may be identified as above, within any parameterised class the target function will be found, given total information of the data distribution. Empirically identified models will converge to this

target in the limit as the data set size tends to infinity (if chosen using the probability distribution of the system).

This empirical identification is carried out using some computer based algorithms.

4.4 Tuning Methods

As discussed above, Neural Network parameter tuning is an optimisation problem. The parameters in the model are optimised against the empirical error.

A kernel map, for instance a Gaussian curve, is of the form

$$G_\sigma(\mathbf{x}, \mathbf{c}) = G\left(\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2\sigma^2}\right). \tag{4.21}$$

We call σ the *width* parameter, and in a Gaussian curve it corresponds to the standard deviation. The parameter $\mathbf{c}$ is the *centre* of the function, and it can be seen in the Gaussian case that this corresponds to the mean vector. As is the convention, the parameters that describe the connections between nodes in the network are simply *weights*. An RBF Neural Network may be represented by:

$$f(x) = \text{Sign}(\sum_{i=1}^{M} w_i G_{\sigma_i}(\mathbf{x}, \mathbf{c}_i)). \tag{4.22}$$

One of the methods in the literature is the Support Vector Method, as discussed in the previous chapter. This uses a least squares error, with a penalty on large weights. However, both the width parameter and the maximum number of centres are fixed before optimisation. This represents a restriction on the class of functions that is allowed. The centres are chosen from the training data. The width parameter is identical for each node.

It is obvious though, from Section 4.2 that if the other parameters are made tunable then a larger family of functions will be realizable. It must be remembered that this will increase the sample complexity of the problem. At the moment there is no estimate of how data demands grow if the width parameter and centres are allowed to be adjustable. The different tuning methods are compared with respect to accuracy, data needs and speed in the next chapter.

4.4.1 Back-propagation and Support Vector Machines

The Back-propagation algorithm was initially developed to tune Multilayer Perceptrons type neural networks. The term back-propagation first appeared in the mid 1980s. It is based upon the gradient descent method to minimise the empirical error on the training set. Each example is presented in turn and the parameter adjusted according to the error. To maximise the use of the training data, it is recycled through 'epochs' of training [68].

It has been noted [91], that the width parameter is very difficult to set with respect to the SVM algorithm and greatly affects the accuracy of the resulting network. Figure 4.5 demonstrates this. Here the empirical accuracy of an SVM trained neural network is shown with relation to the width parameter.

In some cases the width parameter is set for each individual neuron. However, this complicates the problem computationally and so in this section only the case where all the width parameters are set to be equal is considered. There are several suggested methods for estimating the 'correct' value. One of the most compelling uses cluster analysis, as this takes into account the class of each point. It must be pointed out that there is a no water-tight definition of what constitutes a cluster. In [91] it is noted that at a fine level each data point is a cluster, and at a coarse level the entire data set is a single cluster.

Other methods gain their estimate from the distances between the points of the training set. For instance the average distance between points or the distance between the mean points of classes. These *ad hoc* methods are explored to see how they can be utilised to produce a computationally efficient method of setting the width parameter. The results are placed in context by finding the best value by systematically estimating the accuracy of networks for a range of values of σ.

As finite data sets are dealt with here, the accuracy of the resulting network is always estimated. Figure 4.5 illustrates how the estimated accuracy of a SVM trained neural network changes with the width parameter σ. First, what is meant by the 'accuracy' of a neural network must be defined.

Definition 4.4.1 Empirical Accuracy of a Classifier *The accuracy of a neural network,*

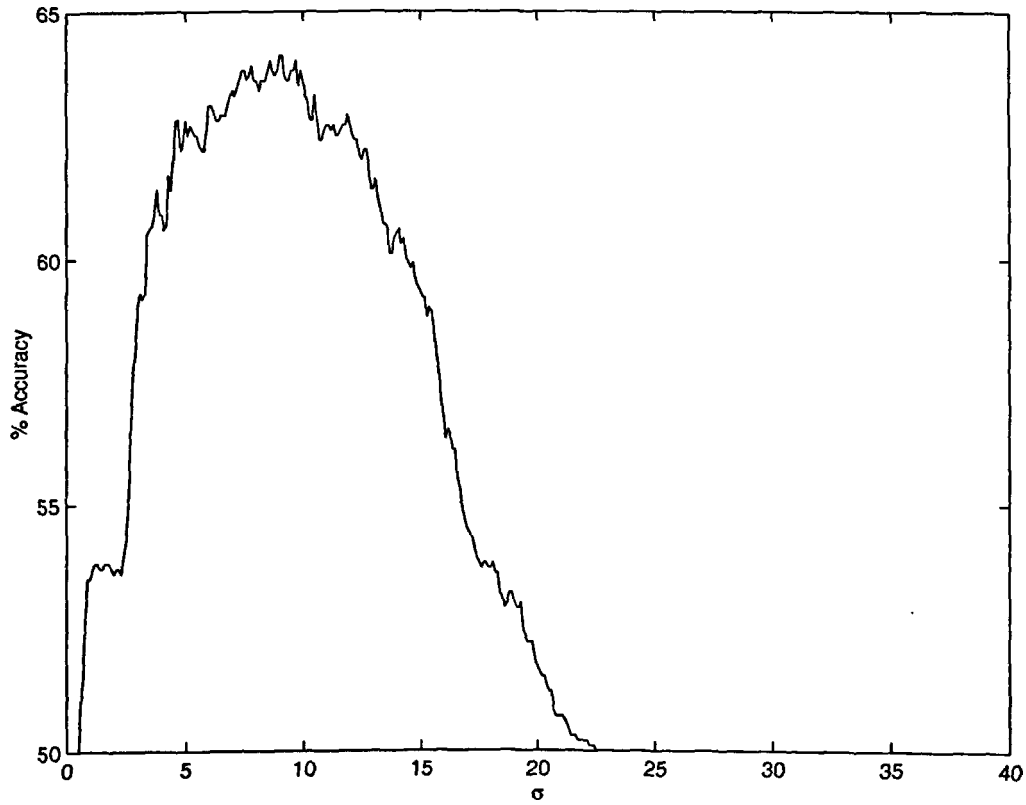


Figure 4.5: Plot of Accuracy against σ for the SVM algorithm.

f(x), based upon a finite data set $(x_i, y_i)_{i=1}^N$ *picked at random is defined as*

$$Accuracy_{emp}(f) = \frac{Number\ of\ Correct\ Classifications}{N} \tag{4.23}$$

It is hope that the empirical accuracy will converge to the true accuracy as the size of the data set grows to infinity.

Using this, what is meant by the optimum value of σ may be defined.

Definition 4.4.2 Optimal σ

$$\sigma_0 = arg\ max_{\sigma \in \mathbb{R}^+ \setminus 0}[Accuracy[f_\sigma]] \tag{4.24}$$

where f_σ is the function representation of the network trained using the support vector method with width parameter σ. The accuracy is the probabilistic expected value, but it is usually estimated using the empirical accuracy over a finite validation set. The law of large numbers dictates that this converges to the true accuracy as the data set grows in size; this empirical optimal σ is defined as follows:

Definition 4.4.3 Empirically Optimal σ *Is defined*

$$\sigma emp = arg\ max_{\sigma \in \mathbb{R}^+ \setminus 0}[Accuracy_{emp}[f_\sigma]] \tag{4.25}$$

There are only empirical means available at the current time to optimise σ. However purely data driven methods do not offer any justification. It may be reasonable to use the spread of training points for Gaussian kernels, this is not so intuitive when applied to other RBF families such as splines. What is required is a model driven technique based on the distribution of the data. This can be applied by estimating the distribution from the training data.

4.4.2 *Ad hoc* methods

An *ad hoc* method is to use the average distance between points. There are two definitions of average distance:

$$\sigma_{emp} = \frac{\Sigma_{i=1}^N \Sigma_{j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|}{N} \tag{4.26}$$

or

$$\sigma_{emp} = \frac{\Sigma_{i=1}^{N}\Sigma_{j=i+1}^{N}\|\mathbf{x}_i - \mathbf{x}_j\|}{\frac{1}{2}N(N-1)} \tag{4.27}$$

Consider a simple classification problem-a binary classification problem, which can be considered as a binary mixture of two distributions. Now, assume further that these two distributions come from the same two-parameter family. Denote the two classes $\mathbb{C}_1$ and $\mathbb{C}_2$ and let their respective distributions be represented by (ζ_1, σ_1) and $(\zeta_2.\sigma_2)$. Let the probability distributions be as in (4.12) and further assume that

$$P(x \in \mathbb{C}_1) = p \tag{4.28}$$

This gives

$$P(x) = pP_{\mathbb{C}_1}(x) + (1-p)P_{\mathbb{C}_2}(x) \tag{4.29}$$

Using this, it is necessary to derive the expected square distance between points. Now

$$\begin{aligned}
\mathbb{E}[\|\mathbf{x}-\mathbf{y}\|^2|\mathbf{x},\mathbf{y} \in \mathbb{C}_1] &= \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \|\mathbf{x}-\mathbf{y}\|^2 dP_{\mathbb{C}_1}(\mathbf{X})dP_{\mathbb{C}_1}(\mathbf{Y}) && (4.30)\\
&\leq \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \|\mathbf{x}-\zeta_1\|^2 + \|\zeta_1-\mathbf{y}\|^2 dP_{\mathbb{C}_1}(\mathbf{X})dP_{\mathbb{C}_1}(\mathbf{Y}) && (4.31)\\
&= 2\mathrm{Var}(P_{\mathbb{C}_1}) && (4.32)
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}[\|\mathbf{x}-\mathbf{y}\|^2|\mathbf{x} \in \mathbb{C}_1, \mathbf{y} \in \mathbb{C}_2] &= \int_{\mathbb{R}^N\times\mathbb{R}^N} \|\mathbf{x}-\mathbf{y}\|^2 dP_{\mathbb{C}_1}(\mathbf{x})dP_{\mathbb{C}_2}(\mathbf{y}) && (4.33)\\
&= \|\xi_1-\xi_2\|^2 + \sigma_1^2 + \sigma_2^2 && (4.34)
\end{aligned}$$

Again notice that the variance is of prime importance. It is apparent then that the distribution of two classes dictates the maximum possible accuracy of a model. It may be concluded that data whose parent distribution exhibits convoluted boundaries between classes, or a great deal of overlapping of classes, represent more complex systems with regard to model building. This complexity results in more parameters to

identify, requiring a greater amount of data than is available. This has been demonstrated in Sections 4.3 and 4.4 to be algorithm independent, and dictated by the real system. Examining the models presented in Chapter 3, it is apparent that a method of controlling the number of free parameters would be to minimise the dimension of the input to the system model. The effects of these *ad hoc* methods will be seen in the next chapter. In the context of this thesis, this means reducing the number of sensor responses in the modelled system.

4.5 Data Set Reduction

The problems of high dimensional data, and the need to reduce the dimension of a data set, have been investigated for some time. Much of the work has come from the 'intrinsic dimension' point of view. This arises from the assumption of some co-dependence between, or constraints on, the observed outputs of a system. This means the hypothesised probability distribution, and so sampled data, lies in some hyper-surface within the output space. Thus a commonly used definition is that the intrinsic dimension of a data set is the number of free parameters in the minimal model that describes the set. However this model is not effectively known. There is also some confusion over this definition, for example in [92] this is taken to mean the subspace that the data lie in. Considering the case of data sitting on the surface of a sphere shows this to be untrue. The dimension of the surface, that is the number of parameters necessary to describe it, will be one less than the dimension of the vector space in which it sits, but the vector subspace in which it sits is the entire space. Thus for generality the data must be considered to reside in a surface and so a subspace is one particular, linear case.

The approach has therefore been to use tools that attempt to approximate this minimal model by making certain assumptions. The approach is to identify a mapping that lowers the dimension of the data set, inline with an assumed form for the surface. For example, Principal Components Analysis (PCA) [93], the Sammon map or some hierarchical cluster analysis [94] have been used for this purpose. PCA looks for directions along which the sampled data are varying. This commonly used algorithm

outputs any number of these directions, or principle components, up to the number of dimensions, or the number of samples, whichever is lower. It is quite common to find that the first three or four principle components describe approximately 95% of the variance in the data [95]. So, for example, assuming linearity, using these new coordinates should encompass the majority of the behaviour of the system modelled.

A number of other techniques have been considered that avoid the need to identify the specific form of the surface in which the data sit and instead attempt to measure some concept of dimension directly. In [92] and [96] a fractal box dimension [97] method is considered that relates the resulting fractal dimension to correlation between the components of the data vector. Another related technique is proposed in [98] and [99] where the covering number of the set is measured in terms of k^{th} nearest neighbour spheres.

In [100] the topological aspects of the problem are explicitly analysed, lower bounds are given for dimension reduction, based upon the assumption of the parent distribution lying in a smooth manifold. It is also explicitly stated that a preservation of topological properties is desirable. The most basic property is that inherent to most pattern recognition(PARC) techniques, which is that close neighbours should remain close under mapping, and remote points shouldn't be mapped onto each other. Thus a continuous map is required.

There have also been studies of sensor selection using genetic algorithms, these being ideal due to the discrete nature of the optimisation problem. The chromosomes are binary coded to represent the presence or absence of a specific sensor. The sensor subset is optimised against some discriminant map. Quite commonly this is the linear discriminant map [101]. Another approach is to use a trial and error algorithm [102]. Each sensor's response is removed in turn from the full array, the one whose removal allows the greatest improvement in discrimination is permanently removed- the process is repeated on the subset until a degradation in the discrimination rate stops the cycle.

However, these genetic algorithms can take some time to run; the individual model identifications for each sensor's removal takes a great deal of computer time. Studies report sensor reduction by about 60% which, for say 32 sensors, represents a great deal of computation. It is argued that the aims of rapid medical diagnosis dictate a stand

alone unit for hospitals; certainly the processors demands should be limited as much as possible. We seek as simple an algorithm as possible to reduce the data set dimension, and to increase the accuracy of the discriminate models.

The Sammon map is a metric preserving map [103]. In fact the algorithm searches for coordinates in a specified lower dimensional space that preserve the inter-point distances. It can be seen that this technique will be most effective in reducing the dimension if the 'intrinsic dimension' of the data is much lower than the space in which they sit. For example data sampled from a system with a four dimensional output may all lie in a three dimensional subspace, hence the Sammon map can find 3D coordinates that retain the inherent metric information.

New variations of these two techniques will be considered. Dimension reduction will be performed before applying four different black box models: two linear, the other two nonlinear. It is apparent in this context that dimension reduction will be beneficial as any interdependence will mean there is not a unique solution for the set of free parameters.

Besides normalisation, two approaches are taken, and both result in a linear projection. However we make a distinction in the following way. One approach is to analyse the information provided by each individual sensor channel of the electronic nose, relative to the rest. Due to this leading to a selection of the most informative channels this is referred to as *Sensor Selection.* The second approach is a more general deferral to the assumption of channel interdependence. A mapping which produces new coordinates, in a lower dimensional space, which are a function of the original channels will fit with this assumption. This is referred to as *feature extraction.*

4.5.1 New set reduction methods

As a consequence of the discussion above, two novel data reduction techniques are discussed. It is assumed that not all sensors supply data containing information relevant to the discriminate task. There may be noise from the environment such as temperature effects etc. There may be information about other aspects of the sample. It is necessary to have a way of selecting some sensors, or functions of sensors, that reduce the dimen-

sion of the input of the discriminant map such that the data set is sufficiently large enough to estimate the free parameters of the resulting reduced dimensional model. Here two different approaches are taken. The first is based upon the observation that any sensor, whose output is dependent upon, or correlated with, another sensor is redundant to some extent. Indeed, if there is a mutual dependence between two sensors then the uncertainty of the output of one will reduce with knowledge of the output of the other. Hence it will have a smaller information content than if it stood alone; the greater the dependence, the less information supplied. An extreme case to consider is that where one sensor output is a function of another sensor. In this case, knowledge of one sensor will allow us to know the output of the other; it does not need to be measured directly. It can then be seen from this that there will not necessarily be a unique parameterisation. Thus, removing a sensor whose output is dependent upon others' outputs will result in a better posed parameter estimation problem where a 'majority' of the information content of the full output has been retained.

The first approach considered uses cross-correlation. This approach seeks to identify a sensor that correlates highly with a number of other sensors, this sensor being used alone to represent the group of correlating sensors. A few such sensors are sought to make the reduced input for the discriminant black box models.

The cross correlation technique, works as follows. The full correlation coefficient matrix is calculated. A corresponding matrix containing the sign of each of the coefficient entries is produced. The sign is encoded as a one for a positive value, a minus one for a negative value and a zero for zero. The columns are then summed and sorted - giving a list of sensors in the order in which they appear to be correlating with other sensors. Sensors then are selected from this list.

The second approach adopts the Sammon map. As it is presented in the majority of the literature the Sammon map is not a map in the traditional sense. It is primarily used as a data visualisation tool, where the algorithm searches for coordinates that preserve interpoint distances in a much lower dimensional space. However, here this algorithm is extended and the algorithm output is used to estimate a linear projection to apply to a previously unseen data set so that the resulting model may be applied to data that were not used to identify the parameters.

A projection is produced from the Sammon map by first applying the algorithm to a subset of the data. This results in a new set of coordinates in a lower dimension. A projection is sought that produces these coordinates, which is a solution to a set of linear simultaneous equations. This projection can then be applied to new data to reduce the dimension whilst preserving metric information.

Formally, define D to be the original data set and D_p to be the coordinates in a lower dimension that result from the Sammon map algorithm (we use the convention that samples are in rows and the outputs are in columns), then we seek a projection P such that

$$PD^{\mathsf{T}} = D_p^{\mathsf{T}}. \tag{4.35}$$

Hence by applying the Penrose-Moore pseudoinverse [104] of D^{T}, $D^{\mathsf{T}+}$, then we obtain an approximation, P_0 for the projection:

$$P_0 = D_p^{\mathsf{T}} D^{\mathsf{T}+} \tag{4.36}$$

This projection P_0, it was observed in practise, produces results very similar to the Sammon mapping. However, unlike the Sammon algorithm, it may be applied to new data to reduce the dimension of the set. The only assumption being that the new data have the same distribution as the set used to produce the projection. This is justifiable as we assume this in applying the black box models.

4.6 Conclusion

A number of phenomena have been observed in the system identification stage of modelling, where data and an assumed model interact to give a specific system model. It has been noted that the complexity of a system, the required system model and the observed data are intimately linked; more complex systems require more data to identify them. This is a potential modelling problem due to this characteristic often not being known prior to data collection.

Previous research has pointed to there being an intrinsic bound on the complexity of a model that may be identified with any given quantity of data. Providing a measure

of complexity is difficult and so the concept is defined relative to the application, often being manifested as a penalty term in the error function used for model identification. This phenomenon will be investigated in the next chapter.

Assuming discrete differences between the samples from which the observed data result, it has been concluded that not only does the additive noise in the real system limit the potential for discriminating between classes; it also increases the size of the data set necessary to successfully identify the system. Indeed, increasing overlap of membership functions requires less smooth maps; hence, a family of maps of greater complexity is required to realize the system's behaviour. This phenomenon will be observed in Chapters 6 and 7, where actual reduction of input dimension of the model system, reducing the complexity, yields a greater successful classification rate. Hence the potential effects of noise should be considered early in the modelling process.

Therefore the following is noted. The input dimension to the system is an important factor in the complexity of a model as is the additive noise present in the system. Interdependence between variables has the potential to simplify a system that looks complex at first examination.

It is finally noted that model reduction, either by temporal order, or input dimension, may serve to increase the long term robustness of the resulting model. New techniques have been suggested and the above theory will be analysed in the next three chapters when they are applied to real world applications such as the screening of biomedical samples.

Chapter 5

Model Identification II: Optimisation algorithms for parameter estimation.

In the previous chapter, some theoretical observations were made on the data needs of system identification, that is the effect of the complexity of a given model. The most important observation was the large amount of data required to estimate the free parameters for a complex system; this complexity increases with the noise in the system.

How parameter estimation algorithms react in practice to noise in the data is of distinct interest. This is especially so when considering the data analysis results that are presented in Chapters 6 and 7. The practical implications of noise and model complexity are considered.

Parameter estimation is normally realised using some form of numerical algorithm. It would be preferable if it was known which algorithms found the best approximation of the theoretical optimal system under consideration. The way that these algorithms react to different data sets is also important.

All system identification in practice is inherently an optimisation process. By requiring a model that 'best fits' the observed data implies the minimisation of such criteria as error between observed and simulated systems and complexity. Hence, in

practice, the algorithms used are based on generic minimisation techniques such as gradient search, Nelder-Meade simplex type searches and genetic algorithms.

In this chapter two novel applications of optimisation techniques to RBF network parameter estimation are prescribed. The performance of these algorithms is compared and contrasted with the support vector method discussed in Chapter 3 which makes use of standard quadratic programming procedures.

5.1 Parameter Estimation Algorithms

There are a number of different methods for finding minima of functions [105]. These may be characterised into three main types: those based upon the gradient of an analytic function such as steepest descent [106], those based upon exploratory searches such as the Nelder-Meads simplex method; genetic algorithms occupies its own class, that of stochastic methods. The first two methods are based upon well analysed rationales. Their advantages as well as their limitations are well known. Primary amongst these limitations is the necessity of a good 'first guess' for the system parameters, that is the starting point for the respective algorithms to iteratively refine. The genetic algorithm is investigated as an alternative: it investigates all reaches of the allowed parameter space and so may identify many alternative solutions.

An RBF neural network of the form in Equation 4.22 is considered in these tuning experiments.

Each method is introduced below and a series of computer experiments are carried out to assess the characteristics of the model optimisation techniques. The aim is to compare these three broad types of optimisation technique in terms of speed, robustness, stability and the accuracy of the resulting tuned model.

5.1.1 Stochastic Gradient

A well established method of finding stationary points is the gradient descent (or ascent) method. This is an iterative algorithm. The cost function is written in terms of the tunable, or free, parameters. The gradient is calculated in terms of these parameters.

At each step parameters are adjusted in the direction of the gradient in the parameter space. The rationale behind this is that the gradient vector is the direction of steepest ascent/descent.

In [107] this technique is applied to time series analysis. Gradient descent is used to minimise the squared error of the predictor. This is estimated, as discussed above, using the empirical error. The empirical error is a random variable for each function within the parameterised family considered, as the estimate is based upon a finite sample from the parent distribution. Hence the method is called a *stochastic gradient learning method.* The technique reported used a sequential training regime, whereby the parameters were adjusted using the error on one training point at a time. The training data were 'recycled' in epochs. The advantage of this technique, over, for example, the support vector method, is that no *a priori* parameter selection is required. The only selections left are the algorithm's step parameters. These dictate by how much the model's parameters are adjusted in proportion to the gradient.

However it was felt that when estimating a discriminant function it was necessary to consider the entire training set using a batch algorithm, so that one point can not have an adverse effect on parameter estimation. Thus the algorithm was extended to implement this.

The stochastic gradient method was easily modified by writing down an error function that measured the error over the entire training set. The error function needs to be continuous in the model parameters. Therefore the continuous error function of the form in (4.4) is considered. The empirical error function is written as

$$I_{emp} = \sum_{i=1}^{N} (y_i - f(x_i))^2 \tag{5.1}$$

This is written as a function of the model parameters

$$I_{emp} = I_{emp}(w_1, w_2, \ldots, w_m, \sigma_1, \sigma_2, \ldots, \sigma_m, c_1, c_2, \ldots, c_m) \tag{5.2}$$

where w_i are the network weights, σ_i are the standard deviations of the Gaussian curves and c_i are the centres.

From this and using the fact that

$$\begin{aligned}
\Delta w_j^{(n)} &= w_j^{(n+1)} - w_j^{(n)} \\
&= -\mu_w \frac{\partial I_{emp}}{\partial w_j}|_{\{w_i\},\{\sigma_i\},\{c_i\}} \qquad (5.3) \\
\Delta \sigma_j^{(n)} &= \sigma_j^{(n+1)} - \sigma_j^{(n)} \\
&= -\mu_\sigma \frac{\partial I_{emp}}{\partial \sigma_j}|_{\{w_i\},\{\sigma_i\},\{c_i\}} \qquad (5.4) \\
\Delta c_j^{(n)} &= c_j^{(n+1)} - c_j^{(n)} \\
&= -\mu_c \frac{\partial I_{emp}}{\partial c_j}|_{\{w_i\},\{\sigma_i\},\{c_i\}} \qquad (5.5)
\end{aligned}$$

where μ_i are the step parameters, is obtained

$$\begin{aligned}
\frac{\partial I_{emp}}{\partial w_j} &= -2E. \left[\exp\left[-\frac{\|x_i - c_j\|^2}{\sigma_j^2} \right]_{i=1...N} \right] \\
\frac{\partial I_{emp}}{\partial \sigma_j} &= -2E. \left[\frac{\|x_i - c_j\|^2}{\sigma_j^3} \exp\left[-\frac{\|x_i - c_j\|^2}{\sigma_j^2} \right]_{i=1...N} \right] \qquad (5.6) \\
\frac{\partial I_{emp}}{\partial c_j} &= -2E. \left[\frac{[x_i - c_j]}{\sigma_j^2} \exp\left[-\frac{\|x_i - c_j\|^2}{\sigma_j^2} \right]_{i=1...N} \right]
\end{aligned}$$

Here E is the vector of N elements that represent the square error for each training point.

Thus the algorithm may be written as an iterative scheme

$$\begin{aligned}
w_j^{(n+1)} &= w_j^{(n)} + \mu_w E. \left[\exp\left[-\frac{\|x_i - c_j\|^2}{\sigma_j^2} \right]_{i=1...N} \right] \\
\sigma_j^{(n+1)} &= \sigma_j^{(n)} + \mu_\sigma E. \left[\frac{\|x_i - c_j\|^2}{\sigma_j^3} \exp\left[-\frac{\|x_i - c_j\|^2}{\sigma_j^2} \right]_{i=1...N} \right] \qquad (5.7) \\
c_j^{(n+1)} &= c_j^{(n)} + \mu_c E. \left[\frac{[x_i - c_j]}{\sigma_j^2} \exp\left[-\frac{\|x_i - c_j\|^2}{\sigma_j^2} \right]_{i=1...N} \right]
\end{aligned}$$

There are two more details to be considered. Firstly the choice of network parameters, has been swapped for that of tuning parameters μ_i. Secondly a rule is required to decide when to stop this iterative optimisation process.

The first consideration consists, in itself, of two distinct problems; the step parameters μ_i and the seeding of the initial parameter values. The problem with the stepping parameters is that if too small a value is chosen then the convergence will be very slow with respect to the number of iterations, too large and the algorithm will over-shoot any (local/global) minima it approaches. These problems are likely to be sensitive to an order of magnitude of the μ_i. However, the algorithm may be modified to reduce the stepping parameters if there is over-shoot. This may be detected by an increase in the empirical error. The training data may be used to seed the initial network centre vectors.

Stopping criteria for an optimisation algorithm are difficult to set. It is suggested to allow a sufficient number of iterations such that the empirical error reaches a stationary point, in other words a minimum. Another possibility would be to set a minimum error change between iterations, as is the case in many commercial packages.

It is now apparent why this algorithm is referred to as a *stochastic* gradient search. The cost function being minimised is empirical and so, as discussed above, is a random variable on the family of function $\mathbb{H}_n$. Hence, minimisation yields a good discriminant function only in a probabilistic sense.

5.1.2 Genetic Algorithms

Gradient search is only an example of an algorithm for numerical optimisation. Gradient descent suffers from the possibility of converging to a suboptimal solution. In [108] it is suggested that some type of Genetic Algorithm might provide an improved method for tuning neural network parameters. Genetic Algorithms attempt to avoid the local minima problem by tuning a population of parameterised discriminant functions in parallel. The population is modified subject to a set of rules based upon Darwinian evolution.

Conceptually the algorithm implements a 'survival of the fittest' paradigm. The main aspects of biological evolution are represented within the algorithm. Firstly there is a 'genetic' code, which, in this case, is a binary string representing the model parameters to be estimated. Secondly a method of interpreting the genetic code *genotype*

Figure 5.1: A one 'Gene' section of the Chromosome design.

to the model *phenotype* is necessary so that the *fitness* of the individual may be evaluated. In this case the fitness is based upon usual criteria for a model to be a good representation of a system's input-output behaviour. Finally there must be a method for the mixing together of the characteristics of good individual models, that is, to breed the next generation of individuals from the previous generation.

For the algorithm design, the optimisation problem needs to be coded into what are called chromosomes. A chromosome represents particular values of all the parameters,and so represents a particular tuning of the model. Two main techniques are favoured: In [109] the parameters are coded as binary variables, whereas in [110] the genes are real numbers. The binary technique is appealing because when two chromosomes are mated by gene crossover there is no concern about mixing like with like. In the decimal case, due to representation there is this concern with the order of magnitude that each digit represents. This binary code is developed here and it is thought preferable to use the crossover technique as in [108], where transfer of genes is only allowed between corresponding parameters.

The approach taken here is the common one, to set the fitness function to be the reciprocal of the error.

A 20 bit binary encoding is proposed for each parameter. The conversion algorithm converts the 20 bit binary string to a decimal integer. This is divided through by a factor, dependent upon the number of bits in the representation, to give a real number in the range $[0, 10]$. Further to this, the weights and centre components are off-set by -5 to give values in the range $[-5, 5]$, this is because negative weights are required to reproduce the data class labels and centres may need negative components. This particular encoding is designed specifically for the artificial data test considered later on. A section of the chromosome is illustrated in Figure 5.1.

In the design phase the mutation probability and the crossover algorithm have to be considered. Initially the mutation probability was set to a low value of around one in a million to one in ten million. However, it was found experimentally that one in one thousand was more realistic. This is because the effect of a mutation is sensitive to the position in the chromosome in which it occurs.

This parameter is very much like the step parameters in the stochastic gradient algorithm: too low and there will not be sufficient variation in the gene pool to find the optimal solution; too high and the genetic information will be constantly corrupted. The high mutation rate is required because the binary coding is bit position sensitive. Dependent upon a bit's position in the chromosome, a mutation can have a varying effect upon a particular parameter's value. Also the model may have a varying sensitivity to perturbations of different parameters.

The crossover algorithm has similar subtleties. An early idea in this investigation was to randomly choose a start and stop position on the chromosome for crossover to take place. However it became clear in the course of the investigation that this 'scrambled' the genotypes too much. It was therefore decided to limit the amount of crossover. The algorithm choose a section that is of the same magnitude of size as the bit length of the real parameter encoding. In this case a start position is chosen and a finish point is chosen no more than twenty bits away, this controls the amount that an individuals chromosome is 'jumbled up'.

Hence the design of a genetic algorithm for estimating RBF neural network parameters was developed. This was tested and compared with the stochastic gradient and SVM training algorithms.

5.2 Testing Regime

A testing regime was designed to examine the following points.

- The relationship between training data size, network complexity and final empirical error.
- The robustness of the three optimisation techniques discussed above.

- For the SVM, the relationship between a training set's size and parent distribution and the optimal width parameter and number of centres.

The test data are from an artificial set constructed in 32 dimensions. It consists of a mixture of two uniform distributions with some degree of overlap. The amount of training data available and the maximum number of hidden neurons is varied. The aim is to illustrate the first point above. At the same time we will observe the robustness of the techniques via rate of convergence, premature convergence* and repeat runs to observe stability. Note that the theory above does predict potential for separating the two classes.

The overlap is set at $\lambda = 0.05$ as in Section 4.3.3, which results in a set of extreme overlap. The test is purposefully hard, in fact harder than that posed by the 'real' data examined in later chapters. The reason is that here the aim is to examine data complexity issues and posing a problem that may be eloquently solved with minimal data and model complexity will not illustrate this satisfactorily.

It may be observed in the PCA plot of Figure 5.2 that there is a great deal of overlap in these test data. It is clear that there is no obvious discriminatory line that can be drawn between the clusters.

5.3 Implementation

The tests were designed to compare the stochastic gradient, genetic algorithm and SVM tuning algorithm, all of which are implemented in Mathworks MATLAB v5. The SVM algorithm was implemented by solving the weighted error function (3.30) using the inbuilt QUADPROG quadratic programming optimiser which uses a reflective Newton-Raphson Method [111] to optimise the network parameters. All that was required was to vary the width parameter and the bound on the size of the weights. This algorithm solves a constrained optimisation problem. The two other algorithms were implemented using MATLAB scripts written by the Author.

The stochastic gradient algorithm had its initial step size set by the user. However

*When a numerical technique converges to a suboptimal, or local minima, solution.

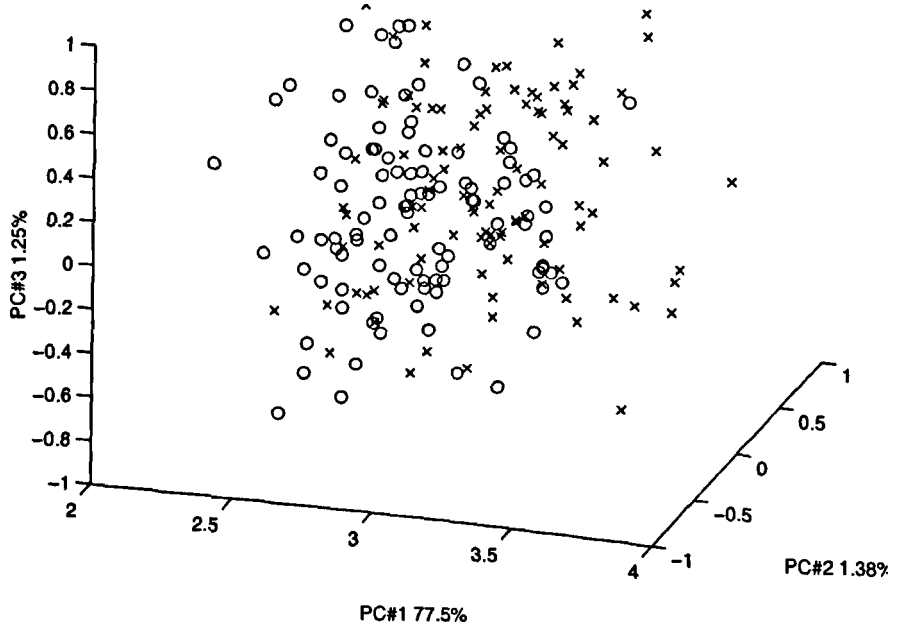


Figure 5.2: PCA plot of a 200 point computer generated test set. Notice the large amount of overlap.

it was found best to 'step out' along the direction of the gradient in increments of one tenth of a step size and judge from the resulting errors which fraction of the step size gave the largest reduction in model error. If there is in fact no improvement in the error then the step size is reduced by one half. This means that as a minimum is approached the step sizes are reduced to prevent overshooting as much as possible.

The genetic algorithm was implemented using MATLAB's standard uniform random number generator. Fitness was based upon the accuracy of the phenotype. MATLAB's standard random number routines were used for the stochastic aspects of the algorithm.

The chromosome crossing algorithm was modified to be slightly more complex than that introduced above. It was observed that, with the gene encoding implemented, there is an inherent symmetry resulting in the genotype to phenotype mapping not being one-to-one. This is because a reordering of the nodes in the hidden layer of the network architecture results in the same model; thus reordering genes results in the same network. This phenotype aliasing becomes a problem because two 'fit' chromosomes in the population may be converging to the same phenotype minima but have different gene orderings. If they are selected to be crossed then the result is that like is not crossed with like. This almost inevitably leads to a degradation in the quality of the offspring.

It was therefore decided to use a selection algorithm based on the Hamming distance between two chromosomes [112]. Using the Hamming distance ensures that two 'close' chromosomes do actually represent converging phenotypes with the same nodal ordering. Thus two individuals were selected and if their Hamming distance was less than 25% of the length of the chromosome string then they were paired off.

The rest of the algorithm was implemented as above. The user provided varying amounts of training data, set the number of nodes, the number of individuals, and the number of generations.

5.4 Results

In this section the results are presented in tabular form, detailing accuracy against the amount of data provided and the complexity of the model, measured via the number of nodes in the hidden layer. The SVM algorithm sets its own complexity so the resulting number of nodes is reported; along with *ad hoc* suggestions from section 4.4.2 for the width parameter and the actual optimal value for this parameter.

5.4.1 Stochastic Gradient

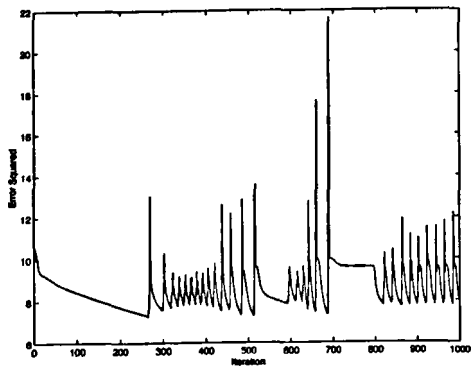
The test data used were those defined by the uniform distribution mixture discussed in Section 4.3.3 with $\lambda = 0.05$. This represents considerable overlap. Training sets of varying sizes were produced, and the accuracy was estimated using a separate set of one thousand points.

Several observations were made when implementing the algorithm without variable step size. The most immediate being the difficulty of setting the step parameters μ_i. Poor choices lead to phenomena like those illustrated (using real output) in Figure 5.3. Here the algorithm heads for some local minimum and seems to orbit about it. This suggests two problems-either the step parameters are too large causing the search algorithm to overshoot, or the gradient of the error function is such that the minimum is unstable. This might suggest a saddle point. Unfortunately reducing the step size often resulted in no convergence at all.

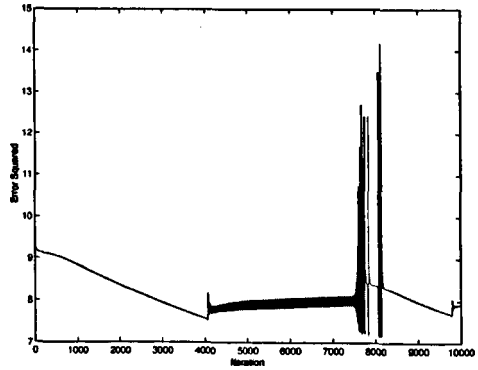
Premature convergence was also observed. This was verified by re-running the algorithm, which generated new start conditions and converged to another solution. In fact in some cases divergence was observed for exactly the same training data and step sizes that also yielded good results. It is obvious then that the algorithm is quite sensitive to the start values of the training points.

By adjusting the algorithm to incorporate a variable step size, it was found that the algorithm became considerably more stable. An initial step size of 1% of the gradient was found to be effective.

The relationship between complexity and training data set size was investigated. The results are shown in Table 5.1. Note that some results are omitted as the algorithm



(a) Example of oscillation around a minimum.



(b) A second example of an unstable minimum.

Figure 5.3: Error against iteration plots for the Stochastic Gradient algorithm. Here the step parameters were badly chosen, so overshoot and oscillation around a local minimum are observed.

took the training data as initial points for the centres. It was found that the choice of starting point, which was randomly generated by the algorithm, was very important to the consequent success of the algorithm.

The algorithm was extremely computer intensive. Some of the more complex runs took up to 5 days to run on a Sun Microsystems server. This length of time offsets the rather good results obtained using this method.

5.4.2 Genetic Algorithm

The behaviour of the implemented genetic algorithm was tracked generation by generation using the individual fitness as a characteristic measure; the accuracy of the best and worst individual in the population and the average. A number of interesting phenomena were observed; destruction of good genotypes; non-emergence of good candidates. These are similar to the cases of oscillation and premature convergence found in the analytic gradient search method.

As mentioned above, destruction of good genotypes was partly due to a symmetry

No. Training Points	Number of Hidden Nodes						
		10	20	30	40	50	60
	20	23%	50%	-	-	-	-
	50	50%	50%	63%	67%	50%	-
	100	66%	67%	69%	69%	69%	69%
	200	50%	50%	68%	67%	68%	66%
	300	50%	53%	67%	65%	69%	69%
	400	50%	50%	67%	70%	70%	69%

Table 5.1: Percentage accuracy of models obtained using the Stochastic Gradient search with respect to training data and complexity of network.

problem inherent to the chromosome implementation. However, there was another problem that was observed-poor fitness functions. It was found that putting the entire population forward for breeding resulted in a fairly large chance that poor genotypes were chosen for the crossover stage. It was therefore decided to put the top 50% of the population forward to this stage.

Non-emergence of good candidates, that is that the evolutionary algorithm fails to find good solutions, is due to two reasons. The first, which cannot be eradicated, is that the initial population contains few chromosome fragments of any value. However it would be hoped that, due to successive generations, something would emerge. This does not always happen. The main factor in the improvement of a base population is the mutation rate. There was found to be a significant problem in setting this parameter, too high and there was a succession of rise and falls in the quality of the population; too low and the population may never progress. It was thus a question of trial and error, but across network models of all complexities it was found to be optimal to use a probability of 0.05 mutations per bit per generation as the mutation rate. This appears to be very high, but this is due to the way the mutation algorithm operates by choosing a handful of bits and randomly choosing to transpose them, this being done to speed up the algorithm. This approximation would result in the actual mutation rate to be much lower.

		Number of Hidden Nodes					
No. Training Points		10	20	30	40	50	60
	20	60%	56%	50%	50%	50%	64%
	50	50%	58%	53%	56%	57%	56%
	100	57%	58%	54%	61%	60%	60%
	200	50%	56%	55%	62%	61%	64%
	300	60%	65%	65%	61%	60%	60%
	400	63%	65%	64%	65%	63%	64%

Table 5.2: Percentage accuracy of models obtained using the Genetic Algorithm Method with respect to training data and network complexity.

The results for the complexity/empirical data comparison is shown in Table 5.2. The algorithm took up to two days to optimise parameters for one model. The complexity/data relationship is clear to see.

It can be observed that the genetic algorithm is not as successful as the gradient search, except, notably, on lower order models. It appears that this algorithm has difficulty coping with more complex optimisation problems. However it was much quicker than the gradient search code; the genetic algorithm proves useful in Chapter 10 for finding good start values for 'more traditional' optimisation methods.

5.4.3 Support Vector Machine

The SVM algorithm was tested in a similar manner to the two algorithms above. Training sets of varying size, and a test set of 1000 points was set up to investigate the behaviour of MATLAB's QUADPROG routine.

The ad hoc distances between points were recorded, as was the optimum width parameter and the number of centres. Besides this the test was the same as above.

There are a number of things that can be noted from the results displayed in Tables 5.3(a) and 5.3(b). The first is that the larger bound (C in Equation (3.28))results in some over generalisation. This is because the network will tend to use less, but larger, weights. This results in certain data points being picked as being 'overly typical' which

Data set size	Accuracy	No. Centres	Triangular Distance	Square Distance	Best σ	SD
20	63%	20	2.36	2.24	4.6	1.67
50	64%	50	2.32	2.28	3.7	1.64
100	65%	100	2.28	2.26	4.9	1.63
200	56%	190	2.32	2.31	2.6	1.64
300	59%	285	2.30	2.29	1.5	1.66
400	58%	356	2.30	2.29	1.4	1.64

(a) Support Vector Machine results. The bound on the weights has been set to C=10

Data set size	Accuracy	No. Centres	Triangular Distance	Square Distance	Best σ	SD
20	57%	20	2.36	2.24	9.4	1.67
50	56%	48	2.32	2.28	4.2	1.64
100	62%	99	2.28	2.26	3.3	1.63
200	68%	183	2.32	2.31	3.4	1.64
300	55%	283	2.30	2.29	2.6	1.66
400	62%	356	2.30	2.29	3.0	1.64

(b) Support Vector Machine results. The bound on the weights has been set to C=100

Table 5.3: Support Vector Machine Results.

λ	Accuracy	Number of centres	Optimal σ
0.05	61%	188	3.5
0.10	74%	174	3.7
0.15	76.3%	127	4.2
0.20	92.3%	78	5.1
0.25	96.3%	67	5.2
0.30	96%	65	4.4
0.35	96.2%	58	5.4
0.40	98.5%	55	6.3
0.45	99%	46	6.4

Table 5.4: Results of a complexity study on the SVM algorithm.

may be observed in the size of the optimal width parameter. The second is that the routine seems to struggle with data sets which have more than 200 points in them. This is a numerical problem, as was observed in the above methods. It is hypothesised that, as the training set grows, in the limit σ will tend to 0. This is because a large data set allows a more complex model.

A second test was to observe how the SVM algorithm adjusts the complexity of the resulting model when the 'overlap' of the data set parent distribution is varied. The results are shown in Table 5.4. The complexity of the network model is proportional to the number of nodes and inversely proportional to σ [69].

It can be seen that the complexity of the model reduces as the parent distribution overlap reduces. Most importantly, σ reduces which means the resulting discriminant function implemented by the model becomes smoother.

Hence the behaviour predicted in the previous chapter has been confirmed:

- As the size of the training data set increases so does the model builder's success. This is because more complex models may be robustly identified.
- As the parent distribution becomes more simple the model becomes smoother, as manifested in the increase in σ.

5.4.4 Discussion

It can be observed that complex models, though giving greater scope to fitting to experimental input/output data, can be extremely costly in terms of computer time and the amount of data required to ensure robustly estimated parameters.

This can be problematic for two reasons: the input to the pattern recognition algorithm may be of high dimension; the amount of data available may be limited. These are characteristics, as shall be seen, of electronic nose experiments, especially for biomedical or microbiological experiments.

It is becoming apparent that fixing the centres *a priori* to training is a good technique. This is for two reasons. Firstly the reduction in computational complexity, and secondly changing the position of a centre, but not the associated width or weight parameters, makes very little sense as far as probability distribution estimation is concerned. This pragmatic approach will be seen to be beneficial in the next two chapters. Certainly it has been suggested in [77] that the Kuhn-Tucker conditions for optimisation imply that the centres should be taken from the training set.

It is apparent from the preliminary results that the error function is not very smooth, and may have a large gradient, even in a small neighbourhood of an optimal point.

The Stochastic Gradient algorithm is not very sophisticated in its current implementation. The algorithm appears to try to interpolate all the data points and so yields very large width parameters. This increases the generalisation error when the resulting estimated function is tested against a novel set of data. It might be preferable to add some smoothness constraints that can be enforced by penalising large weight and width parameters. An improvement may be brought about by some analysis of the complexity of the model by using the Minimum Description Length principle [113] [114].

The genetic algorithm shows some promise, in that it runs much quicker than the stochastic gradient algorithm. However it obviously requires some modifications for it to be sophisticated enough to tackle complex parameter estimation problems such as those detailed here.

As a final thought on analysing the SVM algorithm, it is noted in [77] that sample

complexity is influenced heavily by the radius of the image of the training set under the kernel map. It is conjectured that this is why the SVM algorithm can produce 'worse than guessing results' on sets with large overlaps between classes. In these cases a small value of width parameter is preferable, but this reduces the generalisation ability of the network; that is the complexity of the model increases as was observed above.

The results above in Section 5.4 demonstrate that Radial Basis Function Neural Networks can still be sensitive to noise in the data. It has also been demonstrated that large amounts of data are required to gain reasonable classification results. It appears that, with correct *a priori* parameter setting, SVM trained RBF ANNs are the most robust, and seems the favourite for parameter estimation. In the next two chapters the insight gained from these computer experiments will be used to find optimal discriminant models from laboratory produced data.

Chapter 6

Bacterial Strain and Life Phase Identification.

In this chapter, volatile analysis using a quadrapole mass spectrometer, specifically an Agilent 4440*, is examined. The data processing techniques discussed and developed in Chapters 4 and 5 are utilised to good effect. The Agilent determines the abundance of molecules in the headspace of a biological sample over the mass range 46 to 550 Daltons. It is investigated whether it is possible to distinguish between different samples by detecting only the volatiles that evaporate from each sample. To this end, three different laboratory experiments are carried out. The data are analysed using the black box models examined in Chapter 3. This is in order to measure the discriminant power of the laboratory data to identify the nature of the samples. Different aspects of the samples were recorded, these being the specific strain and life phase of the bacteria in the sample.

As discussed in the first chapter, headspace analysis has the potential to realise rapid medical diagnosis, cutting the time for diagnosis from days to minutes. This type of analysis does present certain challenges. The mixtures analysed are extremely complex and the accuracy and repeatability required for medical applications are extremely strict.

Here, three different types of data set are considered. This is done to demonstrate

*Kindly donated by Agilent Technologies Inc., Delaware, USA.

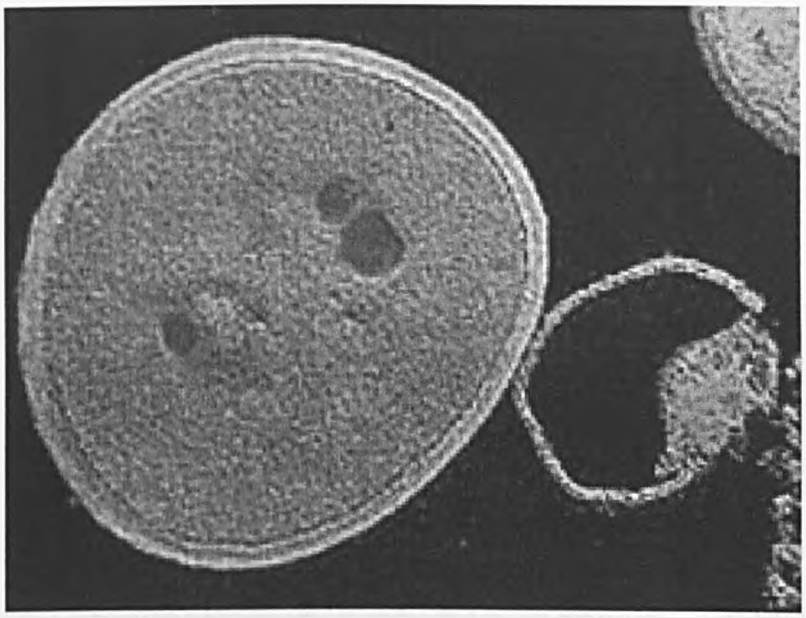


Figure 6.1: *S. Aureus.*

that the techniques outlined can be used to detect both structural and metabolic changes in a colony of bacteria without knowledge of the chemical constituents of the sample. These would be difficult to identify using a mass spectrometer given the complex mixture present in such biological samples. Instead of the mass spectrometer being viewed as producing a mass spectrograph signature of molecules it is viewed as producing a static response, like that of the relative change (2.2)of an array of gas sensors. From this view point it does not matter which output corresponds to which mass because the molecules which produced these responses cannot be deduced. Therefore black box models are suitable for this application.

All data were collected in the Department of Biological Sciences. The first data set consists of samples of two strains of *Staphyloccocus aureus* (Figure 6.1). One strain is antibiotic resistant, the other is not. These are referred to as MRSA (Methicillin Resistant Staphylococcus Aureus) and MSSA (Methicillin Susceptible Staphylococcus Aureus.), respectively. Cultures of the above two strains were grown in laboratory nutrient media. The second data type consists of *E.coli* (Figure 6.2) cultures measured at different growth stages. This is used to show how binary classifiers can be easily

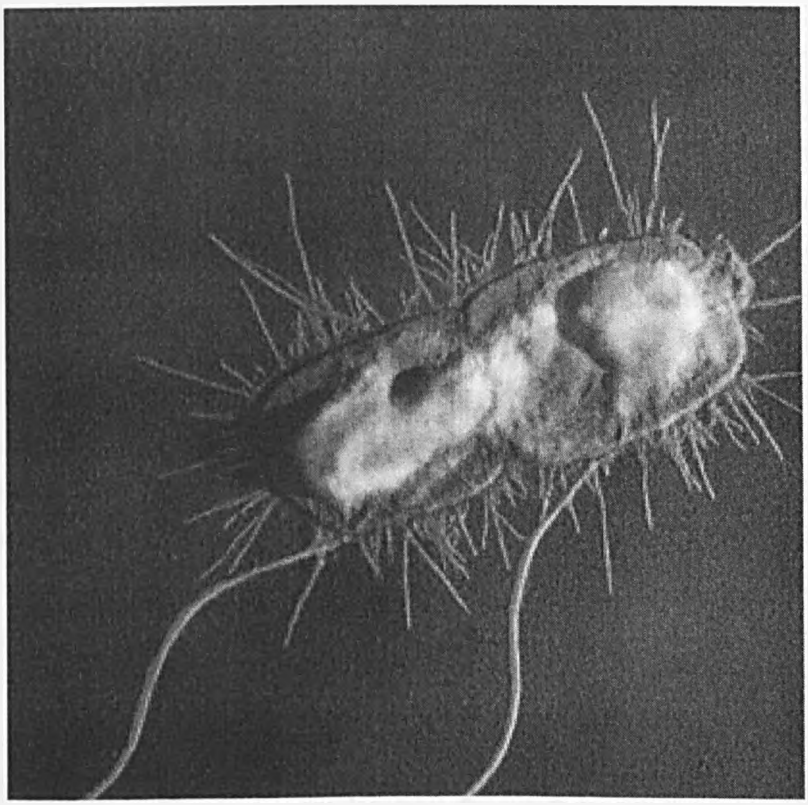


Figure 6.2: *E. coli* under the microscope.

extended to multi-class cases. The final set is related to bacteria cultured in blood. This is to demonstrate that headspace analysis is successful in detecting pathogenic organisms within a biological sample, in this case, blood. It will also be noted that there is a great deal of information in the headspace of a sample, some of which is useful for discriminant analysis, and some of which is not.

6.1 The Problems of Antibiotic Resistance

Due to the emergence of antibiotic resistant micro-organisms, the rapid screening of microbial pathogens has become a crucial issue in clinical care. *Staphyloccocus Aureus* is a serious human pathogen responsible for many cases of septicemia and toxic shock syndrome [115]. Methicillin Resistant *Staphylococcus aureus* (referred to as MRSA) was first reported in 1961, soon after the antibiotic methicillin entered clinical use, and the pathogen is now becoming a major problem in hospitals. This pathogen is also responsible for mastitis in cows and sheep, with severe economic repercussions [116].

Not only has it become apparent that identification of the strain of the organism is important, but it is also necessary to be able to identify its metabolic state. This relates to the viability of a microbe and hence its response when challenged with antibiotics. Moreover, micro-organisms, including certain strains of *Escherichia coli*, display temporal expression of particular genes [117]. In the case of *E. coli O157*, responsible for many cases of food poisoning, verocytotoxin gene expression is enhanced in dormant cells, i.e. a low growth rate stage.

Standard microbiological laboratory techniques require long incubation periods for bacteria that may be present within a sample. Presence and scale of infection are measured by the number of colonies in the growth medium. Species identification is normally through the use of staining agents, though this does not necessarily pick out a particular species. Identification is also possible by using modern PCR DNA amplification techniques. However, these techniques can be slow and expensive. Qualified and trained personnel are required to run the hospital laboratories. As a result, a full diagnosis may take up to three days to perform.

Blood screening is also very important. It is one of the major diagnostic tech-

niques [118] [119] for infection and so, as above, rapid processing of samples would have major benefits. Quite often the practitioner is seeking evidence or not of bacterial infection. Hence, one of the studies detailed here attempts to simulate infection in the blood by culturing known species isolates in blood.

6.2 Experimental Method

It is therefore essential to be able to quickly identify metabolic and structural characteristics of such biological material. A series of investigations were set up to enquire fully into the biomedical application of headspace analysis.

Three experiments were set up to test the hypothesis that there is a relationship between the headspace of a sample and the presence of bacterial organisms; the relationship with the metabolic state of cultures was also investigated. Integration of the Agilent 4440 into such an experiment is very simple. Samples for analysis are prepared in the laboratory and, in the case of the experiments detailed below, are in a liquid state. To introduce the samples to the Agilent 4440, 10 ml of each sample is placed in a 25 ml crimp sealed vial. The analysis process is automated by means of a motorised hopper and a robotic arm which moves the sample vials between processes.

The aim of the experiments was to mimic closely hospital clinical conditions. Thus samples either were clinical (that is from real hospital patients), or were produced in such a way to be like those samples processed by a hospital laboratory. This procedure enabled the techniques to be validated in a reasonable way.

6.2.1 The Agilent 4440 Chemical Sensor

The 4440 chemical sensor (See Chapter 7, Figure 7.1) is produced by Agilent technologies but is marketed by Gerstel (Berlin) in Europe. It consists of two standard manufactured subunits, these being separately available. The first is an Automatic Headspace Sampler (AHS, parts number 7694). This device processes the samples in 10 ml crimped vials. An oven elevates the temperature for 4 minutes, this being set by the operator, to allow equilibrium between the gas and liquid phases to be reached.

The atmosphere within the vial is then drawn off, by puncturing a septa seal in the vial's lid with a needle, and it is then passed on to a mass spectrometer.

The Agilent 5973N mass spectrometer is a quadrapole type unit. Molecular mass is detected using the selected ion monitoring technique. This means that only a particular mass range is recorded. The output of the device is a mass spectrum consisting of a total count for each molecular weight in the specified range. It should be noted that this count is only proportional to the total number of molecules harvested, it is not an absolute measure.

The advantage of using a mass spectrometer is that it will not be affected by environmental conditions such as temperature and humidity. The mass range monitored may also be set, rather than the entire spectrum being recorded. This particular unit also maximises the number of volatiles driven out of the liquid phase. The range of 46 to 550 Daltons was selected as it had been shown previously [117] that this range provided relevant information on bacterial activity.

6.2.2 Data Collection

The data were collected in the Department of Biological Sciences at the University of Warwick and were used for a previous study [117]. Headspace samples were formed by the injection of pure helium gas into a 25 ml vial containing 10 ml of culture, followed by robotic transfer of the sample vial into a heated stage and its stabilisation at a temperature of 80($\pm$0.1)°C. The headspace was then injected into a quadrapole mass spectrometer. The headspace autosampler had a repeatability of about 0.25% by volume. The mass spectrometer analysed the mass content of the headspace with the range set to 46 to 550 Daltons and a resolution of 0.1 Daltons. The spectrometer can record individual masses but the abundance was typically 1,000s of mass units. The unit also contained a series of internal diagnostics to check ion gauge currents and vacuum levels. Figure 6.3 illustrates the experimental set up. The AHS settings are shown in Table 6.1.

The samples were introduced, at each time point, to the mass spectrometer. The Agilent 4440 heats the sample vial up to 80° for 4 minutes, and then draws off the

Parameter	Value
Zone Temperatures	
Oven	80°C
Loop	90°C
Transfer Line	100°C
Event Times	
HS Cycle Time	4 minutes
Vial Eq. Time	12 minutes
Pressurisation time	0.30 minutes
Loop Fill Time	0.15 minutes
Loop Equalisation Time	0.02 minutes
Injection Time	0.3 minutes
Pressure	
Carrier Pressure	4.5PSI
Vial Pressure	14PSI

Table 6.1: Agilent Automatic Headspace Sampler 7694 settings.

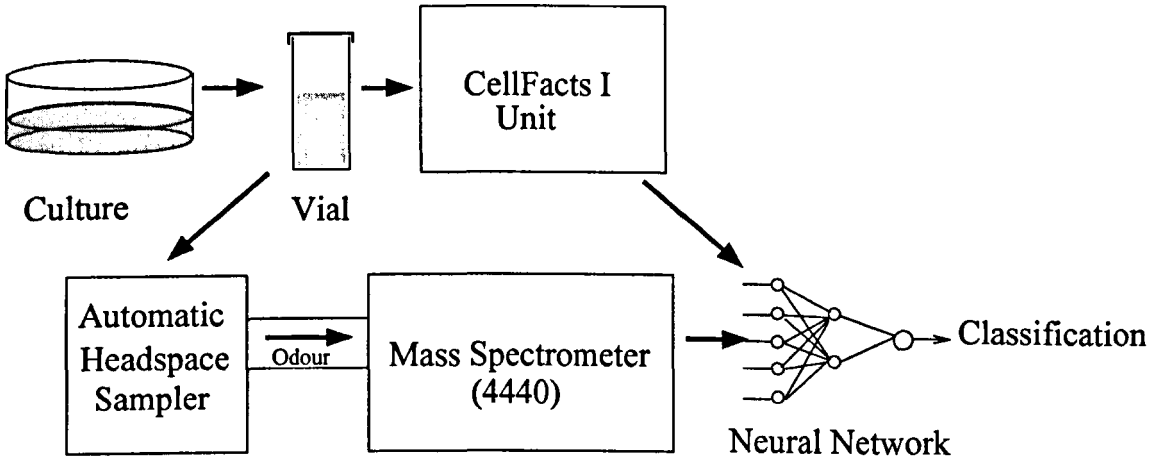


Figure 6.3: Schematic of sample and data flow.

resulting saturated static headspace. This vapour is then passed to the quadrapole mass spectrometer for analysis. The whole process takes approximately 5 minutes. For more information see [117].

The data output is in the form of a mass distribution (called abundance) in the range here of 46 to 550 Daltons. The abundance of each mass is the count of particles of that mass in the sample. The output data are logged on a PC running software provided by the manufacturers of the Agilent 4440 and analysed to see whether it is possible to discriminate between the different classes of samples. Each data point is labelled using the classification achieved following the methods detailed below.

The samples were classified in the laboratory using established methods. The methods were based upon samples being cultured in agar type dishes, classification being based upon number, size and colour of colonies if they are present.

6.2.3 Plated *S. Aureus*

For the *S. aureus* data this was a case of checking whether a sample consisted of the required strain. This was easily checked as separate cultures of the two strains were taken from the Department of Biological Sciences isolates collection. The samples were taken at hourly intervals, so the data also contained information about the growth

state of the culture. Two sets of experiments were carried out yielding two data sets with 68 samples in total.

6.2.4 *S. Aureus* Inoculated Blood

The experiments were carried out in the Department of Biological Sciences at the University of Warwick. The procedure breaks down as follows:

Preparation of Frozen Samples

In order to obtain a culture of a known value of colony forming units per ml, frozen aliquots of the various species considered were created. For each strain three universals were prepared using 10 ml-brain heart infusion (BHI) and one colony inoculated into each. These inoculated bottles were incubated overnight at 37°C and shaken at 75rpm. Following incubation one bottle containing each culture was removed, vortexed and pippeted in 1 ml aliquots into 1.5 ml Eppendorfs. The aliqouts were then centrifuged into a pellet (13000 rpm, 2 minutes) and the supernatant removed. The second culture bottle was then aliquoted in a similar manner into the Eppendorfs containing the pellets and the above steps repeated. The third bottle was processed similarly. Finally 1 ml BHI with 15% Glycerol was added to the Eppendorf containing the pellet and vortexed to create a suspension which was quick frozen to -80°C.

Inoculation of Blood Culture Bottles

All experiments were conducted in BacT/Alert SA (aerobic)(Biomerieux UK Ltd) sterile culture bottles. The bottles contain 40 ml media plus an internal sensor that detects carbon dioxide dissolved in the culture medium. The media formulation consists of pancreatic digest of casein (1.7% w/v), papaic digest of soybean meal (0.3% w/v), sodium polyaneatholesulfonate (0.035% w/v), pyidoxine HCl (0.001% w/v) and other complex amino acids and carbohydrate substrates in purified water.

Microorganism presence and the consequential production of carbon dioxide results in a colour change in the gas permeable sensor at the bottom of the tube.

The culture bottles were injected with 10 ml blood immediately prior to use to recreate a clinical situation of adding a patient blood sample. 10 ml of blood is recommended, although lower blood volumes can be used, recovery may not be as great. During the investigation experiments were conducted using defibrinated horse blood (no preservative)(Oxoid Ltd).

The bottles were inoculated using the above prepared microorganism pellets and incubated at 37°C and shook at 75 rpm to recreate the incubator used in a Hospital.

Preparation of Samples for the Automatic Headspace Sampler

The samples to be processed in the chemical sensor were transferred as 1.5ml aliquots in 10ml sterile flat bottomed headspace vials (Agilent Technologies, Inc). Blanks of BacT/Alert SA and blood were run alongside inoculated samples in addition to BHI broth as standards to confirm the accuracy of the equipment.

A raw mass spectrometer output is shown in Figure 6.4. These data are from the screening of blood samples, though the same form of data was recorded for the other experiments considered in this chapter. Notice that there appears to be very little difference between the two classes of data, this can be observed in the normalised cumulative plot. This shows the cumulative abundance of the spectra, normalised against the total molecular count.

6.2.5 Growth Phase Analysis of *E. coli*

For the *E. coli* data, growth phases were identified via cell count and size [117]. The actual data are displayed in Figure 6.5. These were measured using a CellFacts† I instrument, which measures the size distribution of particles in a liquid sample. For further background information on this system see [120].

Escherichia coli NCTC 10538 (a K-12 strain) was used throughout this part of the investigation. Cultures of E. coli were incubated aerobically in Luria Bertoni broth (LB) at 37°C whilst been shaken at 200RPM. E. coli were incubated in LB overnight and a final concentration of 10^6 cells ml^{-1} was obtained. The flasks were then incubated

†A commercial cell counting and sizing device manufactured by Microbial Systems Ltd., UK.

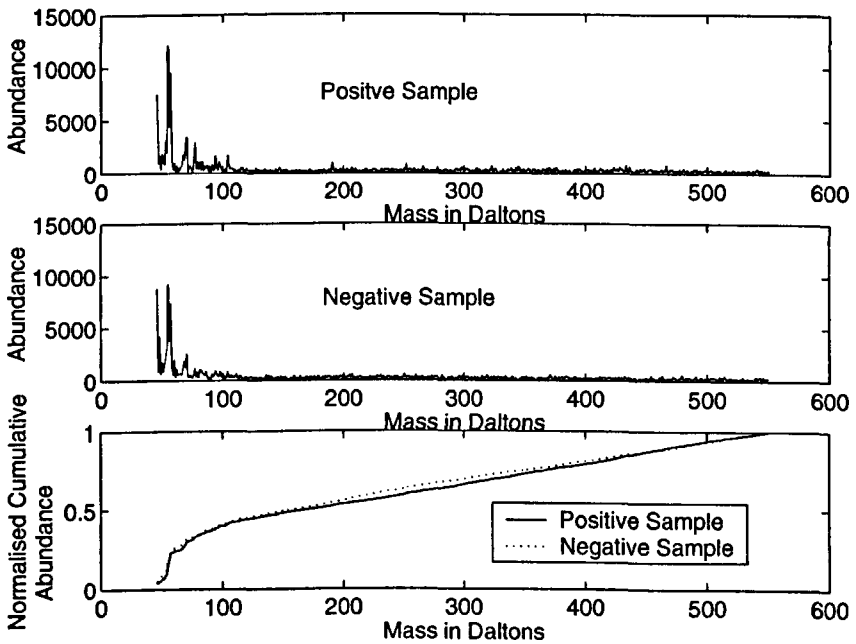


Figure 6.4: Typical mass spectra for positive and negative samples for the blood data. The bottom plot is a normalised cumulative plot which compares the two spectra.

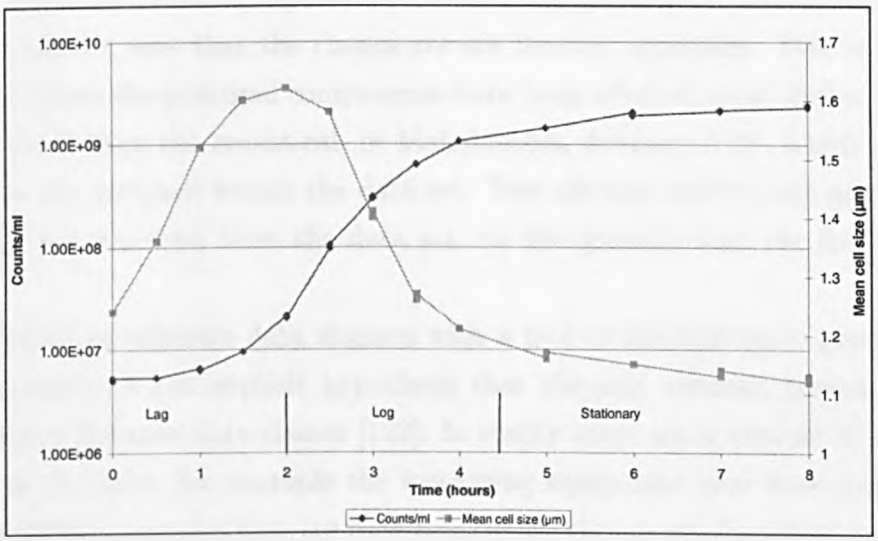


Figure 6.5: CellFacts data showing the change in cell and population size.

and shaken as before and samples drawn off at regular intervals to monitor the growth of the bacteria.

Samples were drawn from the culture at hourly intervals for the first 8 hours of the experiment and then at 24, 48 and 72 hours, respectively.

6.3 Data Processing and Pattern Recognition

As discussed in the previous two chapters, it is necessary to analyse the resulting experimental data with respect to some model in order to interpret the results. The data were logged in the form of tab delimited text files which were read into MATLAB v5.3 for processing.

6.3.1 Principal Components Analysis for Data Visualisation

Principal Components Analysis (PCA) of the data sets collected was performed using MATLAB version 5.3 software. This was in order to visualise the data prior to processing with neural networks. Figure 6.6 shows a PCA plot obtained from one of the

S. aureus data sets, here 'x' represents a MRSA sample and 'o' represents an NCTC sample. It can be seen that the classes are not linearly separable. This is illustrated with a plot where the principal components have been selected using cluster separating criteria derived from the statistical, or Mahalanobis, distance [121], which is a metric based upon the variance within the data set. Two obvious outliers are also observed. These were not removed from the data set, on the grounds that the data sets were small.

Attempting to separate data clusters with a plot of the first three principal components is made on the implicit hypothesis that the only variance present is due to the differences between data classes [122]. In reality there are a number of reasons for variation in the data, for example the measuring equipment may have some built-in error. In addition, samples that are used to estimate the parent distribution will not be perfect, for example in the case of binary classification, as they may have alternative features that are not of interest to this study. For example, when looking for structural differences in bacteria, there may be variation due to metabolic processes. These factors may result in the within-class variation being greater than that between-classes.

It was therefore necessary to find an algorithm to detect which principal components describe the greatest separation between two classes. In [122], it is shown that the Mahalanobis distance between two classes, based upon the kth principal component, is a monotonic increasing function of the expression given in (6.1),

$$[V_k^{\mathsf{T}}(\mu_1 - \mu_2)]^2/\lambda_k. \tag{6.1}$$

Here μ_1 and μ_2 are the within class average vectors, V_k is the kth principal component and λ_k is the variance described in the direction of V_k.

Notice that if principal components are selected which describe a small proportion of the total variance it does not mean that noise has a bigger effect. The components are selected based upon statistically significant separation of classes and this will not be effected unduly by noise.

This suggests a very useful criterion for selecting the 'best' principal components to plot. Equation (6.1) is maximised over all principal components. A three component plot can be achieved by choosing the components which yield the highest values of

Strain	Symbol
MRSA	○
NCTC	+
S. Epidermis	*
S. Warneri	◇
S. Simularis	□
S. Haemalyticus	⋆
S. Lugdenensis	✡
Growth Medium	.

Table 6.2: Key to Data plots 6.8(a) and 6.8(b).

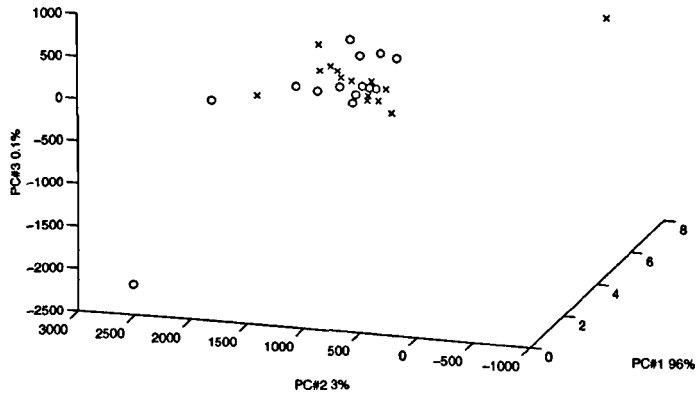
(6.1)(as in Figure 6.6(b)). In this way the distance is maximised between classes observed in the plot.

Looking again at the 3D principal components plots, it is evident that a multilayer perceptron neural network, using a linear hidden layer, would not accurately discriminate between the two strains. In fact there are only sufficient data to train a very simple network. This is because a linear method estimates hyperplanes separating the two classes and it is apparent that here no such hyperplanes exist.

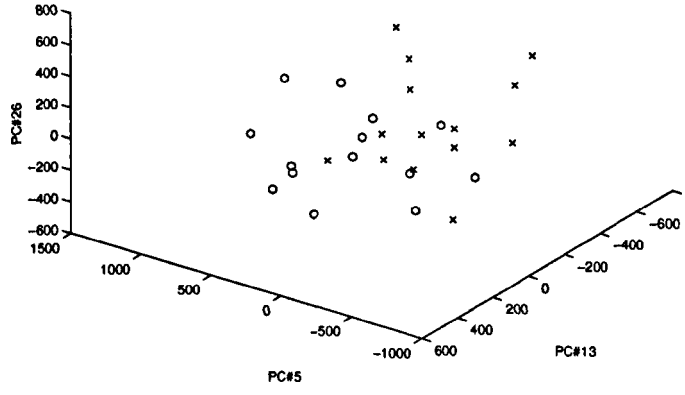
An examination of a PCA plot for the *E. coli* data in Figure 6.7 shows that these data exhibit stronger clustering. However there is still some overlap of classes. The arrows are overlaid to illustrate how the data are changing with time. There is obviously a relationship between the growth phase and headspace of a culture.

An examination of the PCA plots for the *S. aureus* inoculated blood shows a good degree of separation between the different species.

The result of the correlation technique of section 4.5.1 may be observed by comparing Figures 6.8(a) and 6.8(b): note that separation has not been affected but the data may now be analysed using a less complex model form.



(a) First three principal components (percentage variance expressed is also shown). The data used were the raw output from the Agilent 4440.



(b) Selected principal components.

Figure 6.6: Principal Components plots for *S. aureus* data file. 'x' represents antibiotic resistant cultures, 'o' represent susceptible colonies.

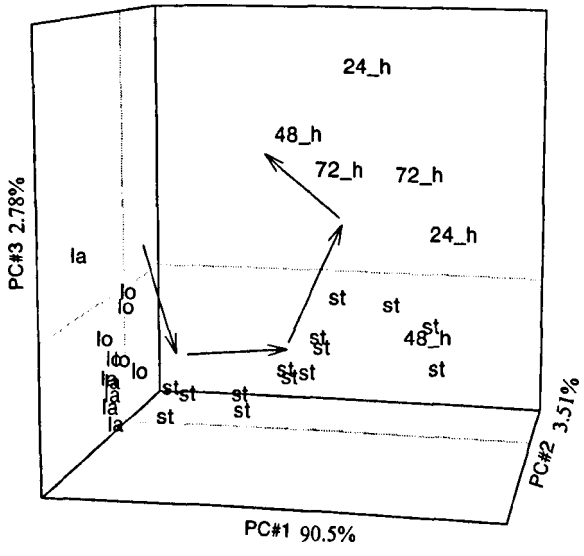
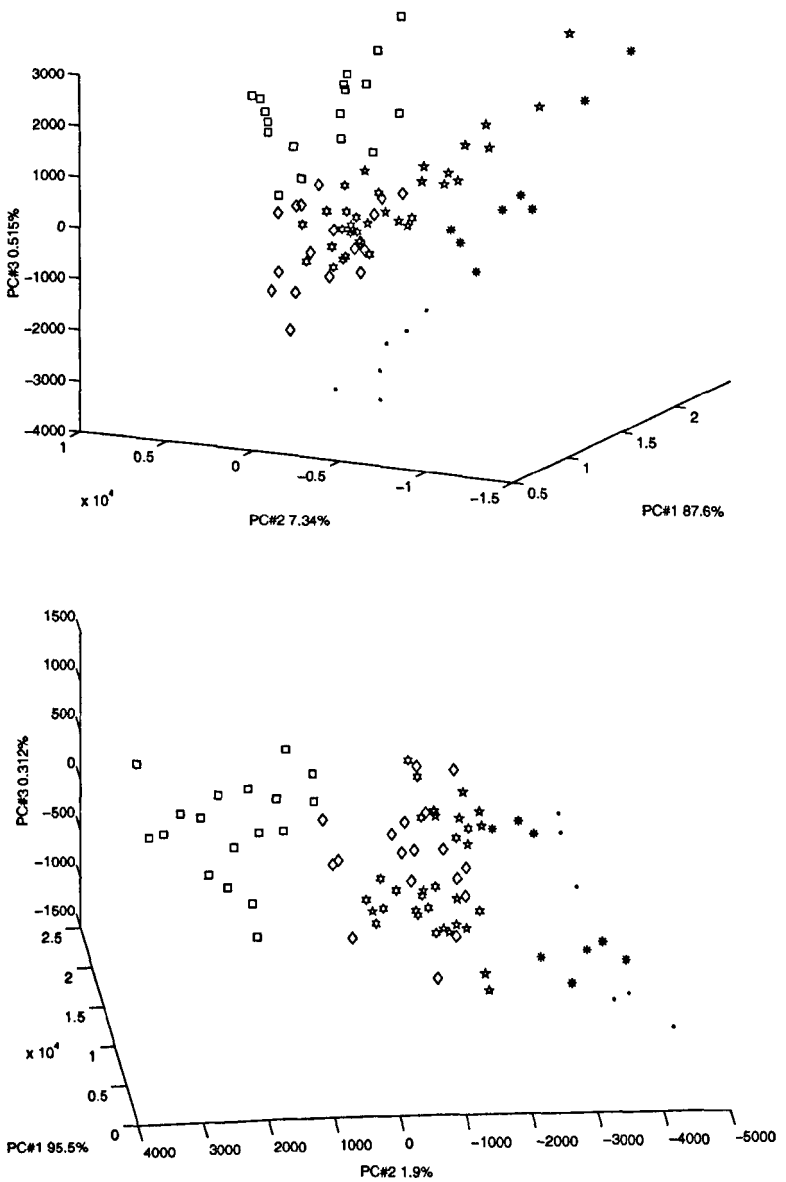


Figure 6.7: PCA plot of *E. coli* data. la=Lag, lo=Log, st=Stationary. Other points are labelled with their time key, so 24h=24 hours, 48h=48 hours and 72h=72 hours. The arrows have been overlaid to illustrate the change with time.



(b) PCA plot of blood data from reduced mass range, using three separate data sets and a correlation technique.

Figure 6.8: PCA plots of blood data. Refer to Table 6.2 for key.

6.3.2 Radial Basis Function Neural Networks

The RBF networks considered here are binary prediction classifiers. One class is assigned the label '1', and the other the label '-1'. As the output of a network is a real number, this was achieved by using a three stage activation function. Thresholds for classification were set such that an output between -0.5 and 0.5 would be classified as 'Unknown'. This activation function is given in Equation (6.2).

$$\theta(t) = \begin{cases} -1 & t \leq -0.5 \\ 0 & -0.5 < t < 0.5 \\ 1 & t \geq 0.5 \end{cases} \tag{6.2}$$

It was felt this is a much better test of the accuracy of the network because it does not force the solution into two states; an unknown bacterial species needs identifying and the identification must be correct. The unknown region allows for the possibility that:

a) The sample is of an unknown species of bacteria.

b) The sample is of a known bacteria, but is sufficiently different for it to be incomparable with the training set.

The networks discussed below employing Gaussian RBFs were implemented in MATLAB. The scripts were custom written as MATLAB has no standard tool box to implement the SVM training method described in 3.2.4. Also the decision to use our own code was based upon the control this gave, and the ease with which the data could be explored.

6.3.3 Model Parameter Estimation and Validation

The two *S. aureus* experiments yielded two sets of data files. These two data sets are denoted A and B to aid exposition. Set A contains 28 samples and set B contains 40. In each set there were equal numbers of MRSA and NCTC samples. Networks used here were trained using the support vector method. This is a batch training regime.

The *E. coli* data set consists of 32 data points. They were organised in duplicate pairs, in the sense that each sample was split into two, and placed in separate vials for sampling. The data were split up into 4 classes; the three main phases discussed above and a fourth phase called 'Later'. The 'Later' class consists of the samples taken at 24, 48 and 72 hours. This fourth stage was introduced because these samples represent bacteria much further along the time scale than the other samples.

The *E. coli* data set consisted of 6 Lag points, 6 Log points, 14 Stationary points and 6 'Later' points. Hence it was necessary to extend the binary classifier. The simplest solution was to train a network for each class, each network recognising one class as being 'positive' and the others as 'negative'. This therefore required four networks in total, but only a single element output. It also permitted independent training and so gave better performance of unscaled abundance data.

The four network outputs were combined into a four dimensional vector. The target output for each class was a positive value in the corresponding component of the output and a negative result in the other components.

Here a classification is 'forced', using the sign of the output, rather than setting thresholds. In the event of a zero output, we classify this as 'unknown'. Hence the predicted result is the most positive component of the output vector. This was felt to be acceptable as the four classes are separated temporally and the classifications are based on 'eye' inspection of the CellFacts output.

The blood screening experiments yielded 78 data points. Instead of splitting the data into two sets, the leave-one-out algorithm was used.

6.3.4 Selection of Training and Validation data

The *S. aureus* data were collected over a full growth cycle so the data sets not only contained information on structural differences but also metabolic information [117]. To illustrate this the data sets are split into two halves, one containing early growth stage samples and the other containing later stages of growth. It must be emphasised that training and validating with two separate sets of samples produced on different days, the accuracy and the robustness of the resulting networks can truly be estimated.

The *E. coli* growth phase part of the work only yielded one data set. However, as discussed above, the points are paired in duplicate samples. Hence two data sets were produced by splitting up the duplicate samples. These two sets will be referred to as X and Y.

6.3.5 Setting the Width Parameter

A problem yet to be satisfactorily solved with RBF networks trained using the SVM method is the setting of the width parameter, σ, of the radial basis functions, as was mentioned in Chapter 5. A balance must be met where the parameter is sufficiently large to represent the distribution of the class, but also small enough so that points from another class are not included.

Many techniques have been suggested for choosing σ (see [123] and [124]) such as the average distance between training points: as demonstrated in Chapter 4. In this chapter the accuracy of the networks was optimised by trial and error after using the average distance between points as a rough estimate. The width parameter was increased in increments, using the average distance as a guide, until the accuracy began to drop, then the interval defined by the last two values was explored using finer steps. Though simple, this procedure was surprisingly quick, requiring very few iterations and in doing this the computational time was kept to a minimum.

6.4 Results

The classification results of the models applied to the data from the three experiments are presented below.

6.4.1 Plated *S. Aureus*

For the *S. aureus* classification problem, all of the training vectors were selected as centres, suggesting an inherently nonlinear problem.

First, a network was produced using the first half of data set A, and validated with the second half. The results are shown in Table 6.3(a).

	True Classification	
Predicted	MRSA(7)	NCTC(7)
MRSA(1)	0	1
NCTC(0)	0	0
Unknown(13)	7	6

(a) Results of training with first half of data set A and validating with second half. The accuracy is 0.

	True Classification	
Predicted	MRSA(10)	NCTC(10)
MRSA(7)	7	0
NCTC(8)	0	8
Unknown(5)	3	2

(b) Results of training with first half of data set A and validating with first half of data set B. The accuracy is 0.75.

	True Classification	
Predicted	MRSA(10)	NCTC(10)
MRSA(10)	10	0
NCTC(10)	0	10
Unknown(0)	0	0

(c) Results of training with first half of data set B and validating with first half of the data set A. The accuracy is 1.

	True Classification	
Predicted	MRSA(14)	NCTC(14)
MRSA(14)	14	0
NCTC(14)	0	14
Unknown(0)	0	0

(d) Results of training with data set B (40 examples) and validating with first set A (28 examples). The accuracy is 1.

Table 6.3: RBF Network Confusion Matrices for *S. aureus* Strain Identification. Accuracy is defined as the number of correct classifications divided by the total number of test samples.

The prediction rate is very poor, and similar results were obtained from data set B. However, as discussed above, the data vary with growth phase. This means the network was trained to recognise early phase data, and so could not cope with later phase data.

To explore the growth phase information contained in the data the next regime employed involved training with the first half of data set A and validating with the corresponding portion of set B and vice versa. The results are shown in Tables 6.3(b) and 6.3(c). These results are much improved, especially when it can be seen that good validation was achieved against a different data set.

Finally, the approach was extended to train over an entire set. Data set B was used for this test as it had the most training vectors. The result is shown in Table 6.3(d). Here a 100% accuracy was estimated.

6.4.2 *S. Aureus* Inoculated Blood

The blood data were recorded using the mass spectrometer. Due to instability in the measurement of the first two samples it has been found to be a useful data preparation technique to remove the first two measurements of each data file.

Applying an ARX model to the full data set was not suitable due to this being a very poorly determined problem, so data reduction methods were applied. By using the top 200 positively correlating masses, identified using the method described in Section 4.5.1 a cross validated accuracy of 94% was achieved.

Applying the hybrid model gave results of 80% accuracy. However, the best results were achieved by using the a projection obtained using the Sammon map and training a standard RBF network. This resulted in an accuracy of 100% success, cross-validated.

The results are summarised in Table 6.4 and are compared with a standard Multi-layer Perceptron artificial neural network with 8 hidden nodes and one output.

6.4.3 Growth Phase Analysis of *E. coli*

For the *E. coli* data sets 65-70% of the data were incorporated as centres in an RBF neural network. Tables 6.5(a) and 6.5(b) detail the results using the two data sets. Here the result is not 100% accuracy; in fact accuracies of 68.75% and 81.25% respectively were obtained. It is evident that more data are required for accurate classification of this kind of multiclass problem.

It should be noted that, on the occasions when the prediction was incorrect, it was only shifted by one growth stage with respect to time. As the CellFacts classifications were (to a certain extent) subjective, the boundaries of each stage are undefined, i.e transition between growth phases is not completely determined.

	Blood
ARX	Not feasible
-Corr ARX	50%
+Corr ARX	94%
RBF	71
Sammon RBF	100%
Norm RBF	75
Nonlinear Model	80%
MLP	50%
MLP+Corr	72%
MLP Sammon	65%
MLP Norm	50%

Table 6.4: Summary of Results for the blood screening as percentage accuracy. Note that the ARX model cannot be applied to blood data because there are too few data points to estimate the required number of weights.

6.5 Conclusions

Data have been gathered from three biomedical experiments using the Agilent 4440 chemical sensor. The data were analysed using black box models in an attempt to 'separate' data points from different classes. The models produced were validated using data sets separate to those used for parameter estimation. Classification rates were high- a demonstration of the combination of the above hardware and software.

The modelling process included data set reduction to keep model complexity to a minimum. It was found that this approach was very successful, though there was clear interaction between dimension reduction method and the various families of black box models. These observations have consequences for the computational analysis of electronic nose data.

The potential use of the Agilent 4440 in biomedical applications has been demonstrated. It is apparent that there is information about bacterial presence within the headspace of biomedical samples. It is clear that in certain circumstances the combina-

Predicted	True classification			
	Lag(3)	Log(3)	Stationary(7)	Later(3)
Lag(3)	1	2	0	0
Log(3)	2	1	0	0
Stationary(7)	0	0	7	1
Later(2)	0	0	0	2

(a) Results of training with set X and validating with set Y. Accuracy=68.75%

Predicted	True classification			
	Lag(3)	Log(3)	Stationary(7)	Later(3)
Lag	2	2	0	0
Log	1	1	0	0
Stationary	0	0	7	0
Later	0	0	0	3

(b) Results of training with set Y and validating with set X. Accuracy=0.8125

Table 6.5: RBF Network Confusion Matrices for *E. coli* Growth Phase Identification

tion of techniques such as the Sammon mapping and SVM trained radial basis neural networks are capable of analysing the headspace of samples, so as to produce useful data.

The use of the particular black box models considered in Chapter 3 has been validated. They can cope well with nonlinear separation problems, the type that have been considered here. It has, however, been noted that some preparatory model reduction techniques are necessary; firstly to gain robust parameter estimation, given the amount of data available; secondly, to avoid the model incorporating noise from sensor outputs irrelevant to the problem at hand.

The content of information, relating to different aspects of a sample, within the headspace of that sample, has been demonstrated. It has also been demonstrated that

the SVM model identification method, coupled with Gaussian RBFs, is capable of dealing with parameter estimation using small data sets. However, despite parameter estimation appearing to have been robust, the accuracies of the models are only estimates due to the small size of validation data sets.

It may be possible, in the future, to consider strain and growth phase together. As discussed at the beginning of this chapter, this would be very useful, in particular, for the determination of the correct range of dosage of the appropriate antibiotics. This, of course, would require larger data sets for robust and safe applications.

In the next chapter a study for which a greater number of samples were available will be considered-screening for urinary tract infections. Conclusions drawn from that study will be considered and compared to those above.

Chapter 7

Urine Screening Using an Electronic Nose.

In this chapter the progress made during a large scale project to carry out headspace analysis of urine samples is discussed. It was decided to carry out a much larger exercise after the promising results gained from the relatively small data sets discussed in the previous chapter. A C320 electronic nose produced by Cyranosciences was also incorporated in the experimental set-up to test the use of cross-sensitive carbon black-polymer sensors. This unit is produced commercially by Cyranosciences. It is a self contained unit that incorporates pump, fluidics, valves, sensors and controlling circuitry. The unit is hand held and designed to be used in a variety of different applications.

The aim of the study was to ascertain whether it is possible to identify Urinary Tract Infections (UTIs) from the headspace of patient urine samples. UTI is a generic term for infection through the urethra and bladder, and as far as the kidneys. The terms UTI and bacteriuria, the presence of bacteria in the urine, and thus an important symptom, are intermixed in the literature [125]. In this study the definition of a positive sample is based upon evidence of an infection, not just bacterial presence, but also leukocytes (white blood cells) and erythrocytes (red blood cells). This is because leukocytes are an indication of an immune system response. Erythrocytes are indicative of internal bleeding and so are also an indication of infection to some extent.

The reason for 'in-house' screening of the samples is due to them being delivered undiagnosed, which results in all diagnostic tests being carried out in parallel with headspace analysis.

Urinary Tract Infections are presently a widespread and commonly occurring medical problem accounting for some 8.3 million visits to general practitioners in the UK every year (NHS statistics). This may be judged by the amount of literature published on the subject every year. It is estimated that in the year 2002-2003 some 400* or so papers were published on the subject. The literature covers a wide range of cases, investigations and uses of scanning for UTIs. It has also been suggested [126] that UTI screening may be useful for tracking other infections in intensive care units, such as antibiotic resistant bacteria.

7.1 Urinary Tract Infections

Urinary tract infections are caused by a number of micro-organisms [127]. For example *E. coli*, *Klebsiella pneumoniae* and *Enterobacter cloacae*. There are a number of specific complaints that fall under the UTI umbrella, these being dependent upon where in the urinary tract the infection lies. This form of infection is common and is responsible for a large proportion of visits to doctors' clinics. These may also be caused in hospital by the insertion of catheter tubes.

Due to their nature, UTIs are difficult to identify conclusively. The only manner which has been proven effective is to carry out laboratory cultures, which can take several days. During the interim period [128] empirical treatment with antibiotics is normally prescribed. This is problematic as it can lead to antibiotic resistant pathogens (see last chapter). The problem is exacerbated by the fact that different organisms require different treatment protocols ranging from 3 to 14 days of drug therapy [129]. This again may yield drug resistance if the course of drugs does not last long enough.

The cost in terms of time and money has not been fully estimated, except in the USA [130]. Here it is estimated it costs $1 billion a year for health funding for treatment. It is also estimated that a great deal of time is consequently lost from the

*Estimated by a search on the Web of Science: www.mimas.ac.uk

work place [131]. It can be assumed that the rest of the developed world has similar statistics. Thus the importance of such infections is demonstrable.

It was thus decided to carry out a study to see if headspace analysis could detect generic bacterial presence, this being rather different to the studies of the previous chapter which knew *a priori* of the existence but not the identity of the bacteria. Urine presents a complex variable mix of chemicals that relates to compounds that the patient has metabolised. Thus within this mixture will be superimposed compounds relating to bacterial metabolisation. The background constituents of urine include urea, glucose, protein and common salt. The problem then is trying to find the bacterial information in a sample when the size of chemical noise in relation to this is not known.

It was found early on that the Agilent 4440 was, in this case, not capable of producing useful data in terms of discriminant model analysis, therefore only the C320 data are considered here as a contrast to the mass spectrometry data of the previous chapter.

In what follows, experimental procedures, the nature of the data recorded and the discriminant problem are detailed. Further to this such techniques as Data Preprocessing, Neural Networks and System Identification are considered in an attempt to provide systematic classification of the samples obtained. It is concluded that the analysis of the headspace of urine samples is a complex problem, but can be successfully solved.

7.2 Urine Sample Classification

Urine contains a wealth of information about the health of the body. In particular, a great deal can be inferred about the state of health of the bladder, kidneys and the urinary tract itself. To enable this, three key indicators in the urine are considered.

- **White blood cell count - leukocyte (wb)** White blood cells are a good indicator of an immune system response to infection.
- **Red blood cell count - erythrocyte (rb)** Red blood cells in urine indicate internal bleeding. This may be as a result of cancer.

- **Bacteria count** The presence of bacteria suggests infection in the lower urinary tract.

These were measured using the CellFacts I (Cellfacts Ltd, UK) machine, as in the previous chapter. Further to this traditional plating methods were utilised. These methods identified the presence of bacteria, the proliferation via the number of colonies, and staining placed the species into one of two groups. These were yellow-orange and blue-green type staining. These presented themselves as potential labels for pattern recognition.

7.3 Connection of the Cyrano C320 in line with the Agilent 4440

A technique for connecting a modified C320 to the Agilent 4440 has been developed by Cyrano Sciences. An overview of this method is presented here. The Agilent 4440 is split into two submodules: the AHS and the Mass Spectrometer(MS). The automation of the AHS is used to integrate the C320 into the existing set up. Figure 7.1 outlines the method of connection. The internal pump in the C320 has been bypassed, it is therefore important that the pump is set to low speed (40 cc min^{-1}). Part of the headspace sample is diverted to the C320 by incorporating a 1:30 splitter (top of Figure 7.1) in the transfer line of the AHS, whereby the C320 received 30 parts of the sample for every part the mass spectrometer relieved. The pressure produced by the AHS will drive the sample through the sensor chamber. The baseline is provided by the AHS's helium supply, as the purge inlet is connected to the splitter line as well.

The whole assembly is synchronised by the addition of new signal cables (bottom of Figure 7.1). The Agilent 4440 and C320 are both connected to the remote control port of the AHS unit. Tables 7.1(a) and 7.1(b) display settings suggested as guide-lines. These required adjustment once the C320 had been installed to ensure the units were synchronised.

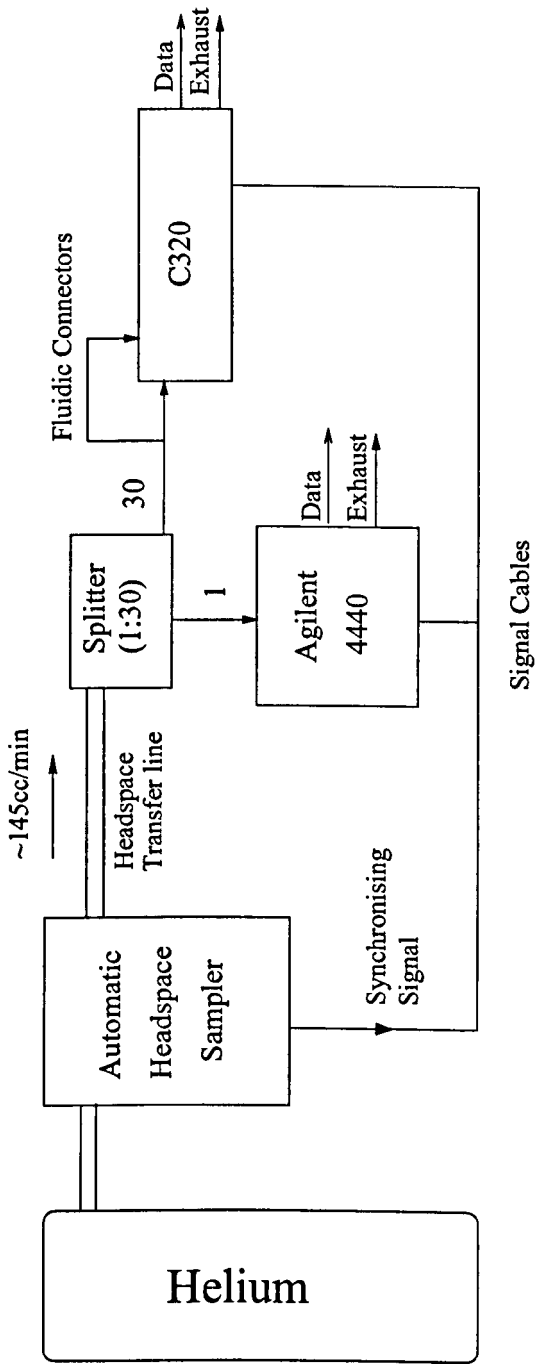


Figure 7.1: Schematic of integration of C320™ with Agilent 4440. Reproduced from Cyrano Sciences: Instructions for Special Use of Cyranose C320 with Agilent 4440 Chemical Sensor.

HS Cycle	60 seconds
Vial EQ	60 seconds
Pressurisation	20 seconds
Loop fill	9 seconds
Loop EQ	1.2 seconds min
Injection	20 seconds

(a) HP4440 settings

Baseline purge	42 seconds
Sample draw	30 seconds
Sample purge	0 seconds

(b) C320 settings

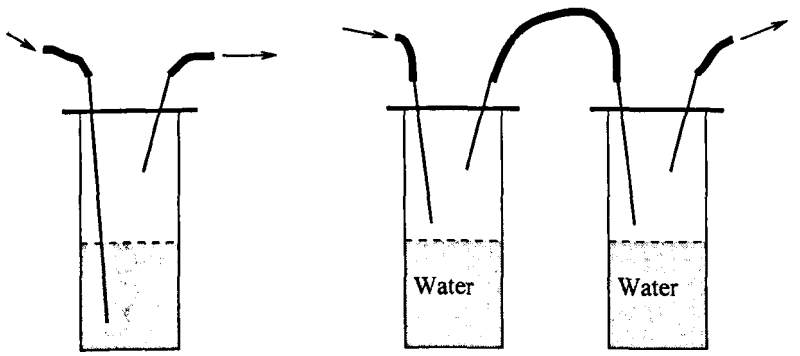
7.4 Experimental Method

The Department of Biological Sciences processed twenty five different patient samples in duplicate, three days a week. Headspace analysis was performed using the Agilent 4440 mass spectrometer. Data logging was also automatic.

For a rigorous experimental design to be implemented, the following issues were considered.

7.4.1 Sample and baseline preparation

The samples as they were provided would not necessarily yield the greatest potential headspace, nor would purging from ambient laboratory air yield the most discriminatory power. It was entirely possible that the sensors would respond strongly to substances that were to be found in all samples. This response may mask the presence of an infection. It is preferable to offset the baseline by treatment of the purge supply in such a way that there will not be a response to 'background' substances. Matching purge and sample temperature is very important, especially as use is made of the AHS's oven. If the sensors are too cool with respect to the oven then there is the risk of the sample condensing, thus possibly damaging the sensor array. Here methods are considered to control and amplify the sensitivity of the sensors via careful preparation of the samples. These are standard aspects of electronic nose experiments.



(a) Sparging. (b) Humidifying without sparging.

Figure 7.2: Methods of humidifying the carrier gas.

7.4.2 Humidifying

A great deal of the content of urine is water and the sensors of the C320 respond to this. Hence it was necessary to investigate if this had a great effect upon the device's ability to react to important features of the sample headspace. In some circumstances it is necessary to humidify the purge supply to match that of the sample headspace. It may be tested whether this is necessary or not by comparing humidified with non-humidified measurements. If humidification is needed then an improvement in sensor response and the separation of data classes, as viewed with PCA, will be observed. Figure 7.2 shows two methods of humidifying the C320 purge air supply using 'bubblers'.

The first method, (a), is very simple, however this will produce water in aerosol form, which can be detrimental to sensor performance. This can be avoided by using a 0.2μm polycarbonate or PTFE filter. This would reduce flow rate by a few percent, but will protect the sensors.

The second method, (b), will not be as effective, and to counteract this two units in series are needed. Also it would be preferable to use vials with a large cross sectional area to aid evaporation.

7.4.3 Additives

It was debated whether to use additives to increase the volatiles driven out of the liquid phase of the sample. One simple method is to raise or lower the pH of the sample. However, as noted above, the chemical constituents of each sample are widely varying. This means that the pH of each sample would have to be 'tuned', to some preset level, by hand. Hence it was decided that this method would be to no advantage at all, as it would slow the experimental procedure down.

7.4.4 Data Logging

Due to the proposed connection of a C320 unit to the HP4440, the data logging should be kept as automated as possible. The *save* routine in the PCNose software provided with the C320 allows adjustment of sample time and maximum duration. The routine also logs the status flags of the unit.

Dynamic data were required in order to assess the advantage of using the dynamic information to discriminate between samples at a later date. It was therefore decided to log continuously over the entire day's run. The individual samples were then separated in the MATLAB environment using the cycle flags also saved by the data logging software. The only consideration then is how frequently to sample and how long the total run will take. The sensor responses were observed to be on a scale of ten seconds, and therefore a one second sample regime was chosen. This sample rate resulted in 9000 entries for a 40 sample run, for which the available computer resources were sufficient.

7.4.5 Experimental Procedure and Sub Experiments

The following procedure was followed.

Checks before Commencing Each Experimental Run.

Before each run the level of water in the humidifiers, if used, should be checked. Periodic replacement of the filter was required, and this was carried out as per the manufacturer's

instructions. It would also be sensible at this point to check that outlets are above liquid level so as to avoid introducing water into the C320 device.

As the unit will have been sat dormant for up twenty four hours, the sensors will have equalised with the environmental humidity. During the first purge cycle this moisture will desorb from the sensor surface, and so affect the signal output. This is what is described as the 'first sniff' issue [132]. It is recommended in [132] to use the manual test option to run the pump continuously for five minutes. This is the 'wake up' cycle. As the motor was disabled it was simply a case of running the C320 for a few minutes before commencing the sampling run. The AHS device undergoes its own warm up cycle, so the 'wake up' cycle was easily implemented.

Data Logging

As in section 7.4.4 it was proposed to log continuously for the HP4440's run. Therefore, all that was required of the operator was to set the sample rate and length, and to specify a filename for the data to be stored under. A desktop PC was set up with the logging software

7.5 The Effects of Humidity

Besides the standards used by the Agilent 4440 (β-ionone and n-tetradecane) it was decided to use pure water and a boric acid solution as standards for the C320. The purpose of this was two-fold; firstly to monitor for sensor poisoning and secondly to investigate the response of the C320 device to humidity with a bubbler humidifying the carrier gas. Water and boric acid were chosen as they are both present in the urine sample but represented no use as far as discriminant power was concerned.

To assess the efficiency of the humidifier, and whether it was needed for the urine screening process, a series of tests was carried out. A single run of standards consisted of six water samples and six boric acid solutions. Two runs were performed: in the first the bubbler was active whereas in the second it was removed from the experimental set up. For the dry helium run a fresh micro-pore filter was used to ensure zero humidity

and an air flow comparable to that with an active bubbler.

Data logging was carried out as standard with the sample rate set at once per second. Responses for all thirty-two sensors were recorded so that sensor selection could be performed at the data processing stage.

The dynamic data were then processed using the relative change measure (2.2). It was hoped to assess whether discrimination was aided by humidifying the purge supply.

Standard Principal Components Analysis plots were produced for the two sets of data and are shown in Figure 7.3. The top plot (Figure 7.3(a)) shows that, without the bubbler, discrimination is poor, there is no clear separation between classes. However the results gained by using the humidifier are no better with respect to discriminatory power. Figure 7.3(b) actually appears to possess less cohesion between points of the same class than in the previous plot.

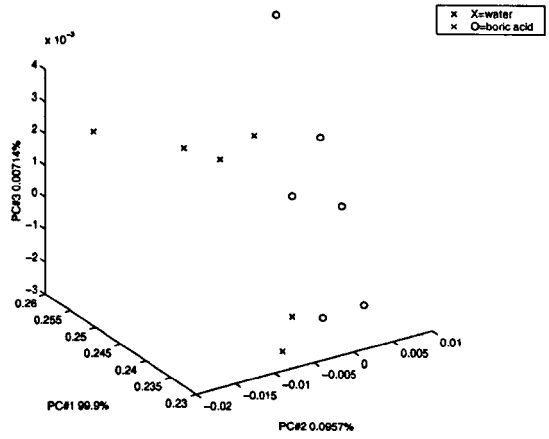
Examining the C320 results for water and boric acid, Figure 7.4, demonstrates that the humidifier is not working to any advantage. There is substantial variation in both the water and the boric acid signals. It was hoped that the water data points would be tightly clustered, because humidifying the purge should reduce the sensors' response to water. However, using the Mahalanobis distance, the derived selection of components does demonstrate that:

- Other components of urine other than water are detectable. In particular the C320's sensors respond to boric acid.
- These two 'standards' are linearly separable.

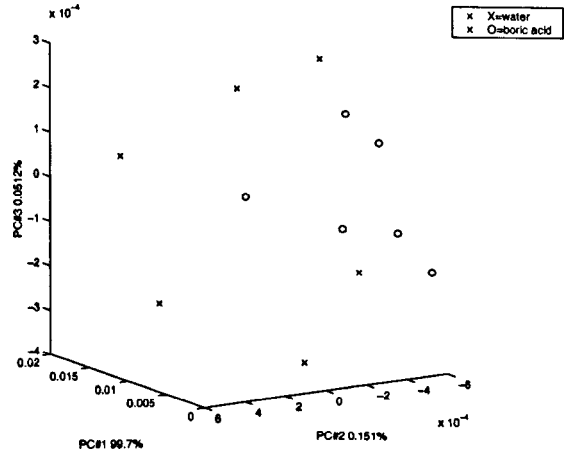
The results of the urine screening experiment, which are shown in Figure 7.5 are now considered. Here screening with and without the humidifier are compared.

If these results are compared with the mass spectrometer data for the same sample set, it seems that the C320 was providing information on the nature of the samples. These plots are shown in Figure 7.6, and it can be observed that the two classes are quite heavily overlapped. Hence the HP4440 was not considered in this study.

Examining the actual time series plots the result of adding humidification can be seen. The response from sensor five was used because an examination of the Principal Components coefficients yielded this sensor to have the most varying response. This

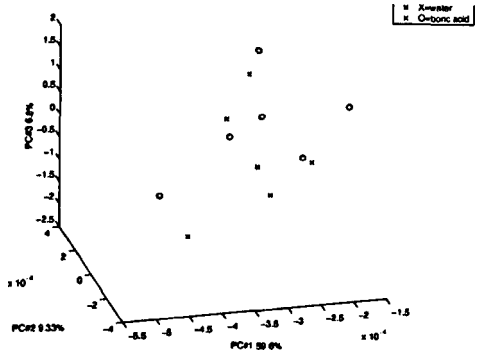


(a) PCA plot of data without Bubbler

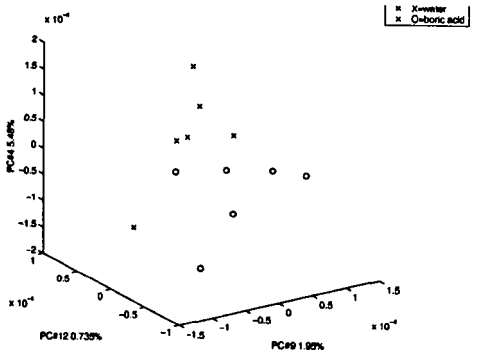


(b) PCA plot of data with Bubbler

Figure 7.3: PCA plots of water and boric acid standards

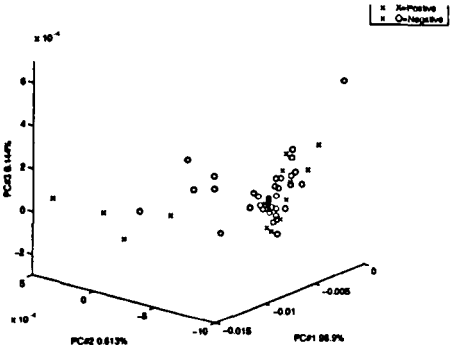


(a) First three PCs (percentage variance variance along the components is also shown).

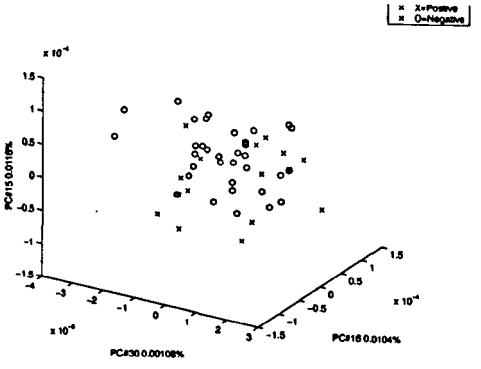


(b) First 3 PCs from a different angle.

Figure 7.4: PCA plot of water and boric acid samples.

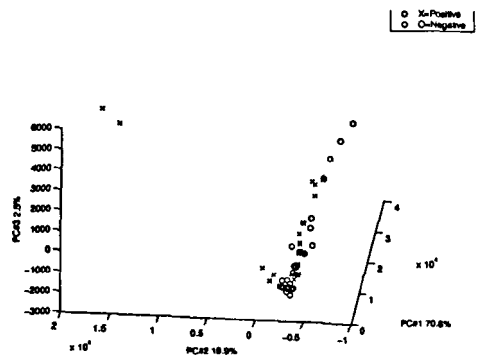


(a) First three principal components (percentage variance along the components is also shown).

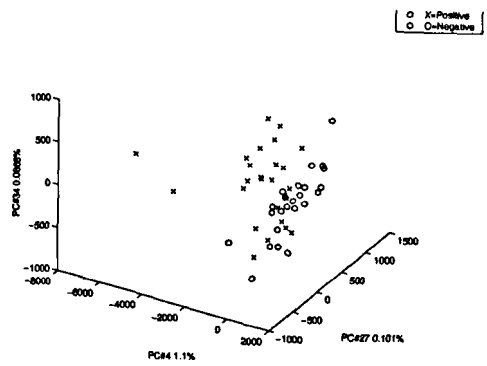


(b) Plot using principal components selected using Mahalanobis distance.

Figure 7.5: PCA results using the bubbler humidifier.

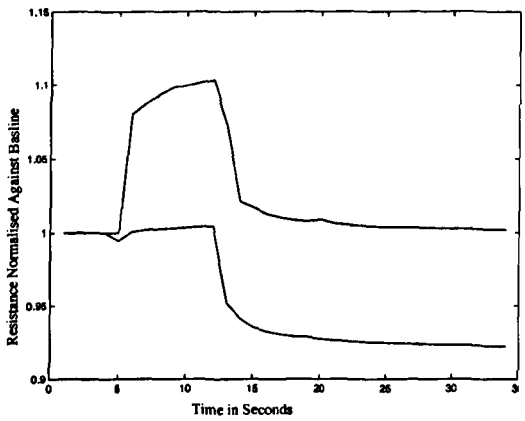


(a) First three principal components (percentage variance expressed is also shown).

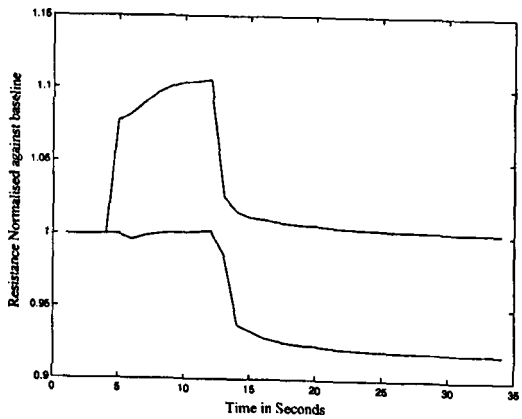


(b) Plot using principal components selected using Mahalanobis distance.

Figure 7.6: Plot of mass spectrometer data from humidifier test.

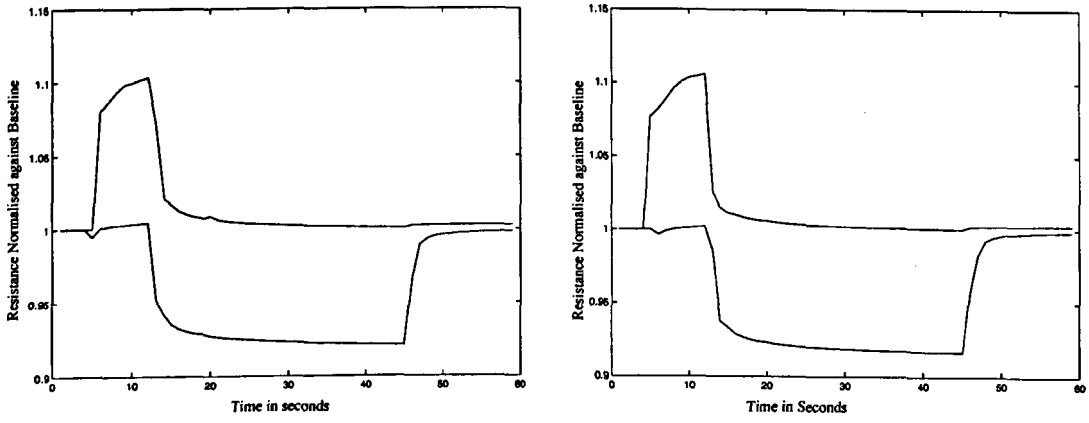


(a) Time series plot of response to water.



(b) Time series plot of response to boric acid.

Figure 7.7: Time series plot of C320 sensor 5 response. On each plot the upper curve is for the case when the bubbler was not used.



(a) Response for water.

(b) Response for boric acid.

Figure 7.8: Long term time series demonstrating return to baseline. On each plot the upper curve is for when the bubbler was not used.

was true for both data sets. Each plot shows the response of this sensor with and without humidification, where the response has been normalised by dividing through by the baseline resistance. The responses in Figure 7.7 show baseline on the left and the response when the sample inlet is opened. The first thing that must be noted is that in both cases of humidification the response actually drops below the baseline after twelve seconds of sample introduction. A corresponding drop is shown in the non-humidified responses. This is due to the headspace being exhausted, resulting in dry helium being swept across the sensors. Secondly the response with humidification is very small in comparison to the dry purge response. Hence in 7.7(a) the 'dry to wet' and 'wet to dry' response of the sensors may be observed.

Figure 7.8 is included to demonstrate that baseline is re-achieved after sample introduction and that humidification of the purge supply is not having an adverse effect on the sensors. However, it was apparent that it was not required.

7.6 Data Preparation

The raw C320 dynamic data were read into MATLAB v5.3. The C320 data were preprocessed using the standard fractional change measurement (2.2). The first 4 samples of each data file were removed as it appeared from PCA plots that these were commonly outliers (see also discussion of instability of the measurement of the first few samples in Section 6.4.2). CellFacts and plating information were provided by the Department of Biological Sciences in the form of a Microsoft Excel spreadsheet. This file detailed the following:

1. **Sample Number.** This is coded by run number and the position of the sample within that run.

2. **Cellfacts screen classification.** This detailed whether, based on medical standards, the CellFacts hardware and software decided it had detected an infection. This column contained the values 'Positive' or 'empty '.

3. **Bacteria.** This contained the CellFacts assessment of the bacterial content of the sample. The allowed values were '-','+' and '++'. These denote the absence of bacteria, the presence of bacteria and a very strong presence respectively.

4. **Red Blood Cell content.** This detailed the number of red blood cells per μl.

5. **White Blood Cell count.** This detailed the number of white blood cells per μl.

6. **Comments.** Comments on the CellFacts run.

7. **Plating +/-.** Stated whether plating had indicated the existence of bacteria in the sample.

8. **Colour.** Colour of the colonies, if any grew.

9. **Comments.** Comments on the plating results.

Bacteria		WBC+RBC		Species		Cellfacts+Plating	
From	To	From	To	From	To	From	To
-	-1	>n	n	Orange	-1	Negative	-1
+	1	<n	n	Blue	1	Positive	1
++	2	n	n	None	0		

Table 7.1: Rules for converting text labels into numeric labels

Columns 1,6 and 9 were removed as they were not pertinent to automatic processing of the data. Columns 2,3,5 and 7 were filtered so that they contained classification in a numeral form by the rules shown in Table 7.1. This was necessary for the implementation of the neural networks for data analysis.

Further to this, a second labelling was applied to the white blood count information. It was decided to split the samples into three classes of low, medium and high white blood cell count. The boundaries were set at 50 and 300 counts per μl as these are national medical thresholds for urinary tract infection.

7.7 Preprocessing Of Data

The data contained a large number of different types of information. The aim of preprocessing should be to remove or minimise unwanted effects such as interference and noise. This noise could take many forms, for example:

- Equipment error.
- What the patients have consumed in the previous 12-24 hours.
- Illness elsewhere in the body.

A great deal of variability could also be due to the water content of the urine samples which dilutes the detectable masses and could alter the overall abundance of detected molecules.

Taking these effects into account, filters were designed to preprocess the data. The static data from the C320 were viewed as being points in a vector space. It was then

assumed that this vector space could be decomposed into a subspace in which all the unwanted variation and noise lies, and an orthogonal subspace where our desired 'information' resides. This is not entirely justifiable because the effects of noise sources 1 to 3 above are not fully understood. However, it was necessary to investigate noise reduction techniques to maximise the quality of the processed data.

Henceforth in this chapter the term 'noise' is used to include all unwanted variance, from systematic, random and/or other errors.

7.7.1 Principal Components Analysis

The reason for performing PCA analysis here was to demonstrate that data classes were separable. PCA is considered here as a data visualisation tool.

7.7.2 Removing variation in Negatives

By the argument at the start of Section 7.7, negative samples provide reasonable estimators of the background noise. In theory, if there was no information in the urine besides that of the presence of a UTI, then all the negative samples should coincide; that is they would be 'close' chemically, and collectively represent a pseudo-standard urine mixture.

A mapping is sought that will remove most, if not all, of the variation in the negative samples. Two candidates are suggested:

1. **A Linear Translation:** The mapping removes the component that is most typical of negative variation from the abundance vector.

2. **A Linear Contraction:** A contraction in the directions most typical of negative variation.

In both cases PCA is used to identify the directions in which the negatives vary.

The former mapping is of the form $T(v) = v - P(v)$, where $P(v)$ is a projection of the sample measurement v into a subspace that is spanned by the negative sample vectors. This is the subspace representing the noise. $P(v)$ is calculated by transforming

into PCA space and then back to the sample space, reconstructing using the PC's as a basis of the 'noise' subspace. Hence the estimation of noise is influenced by the number of 'negative' samples available.

This mapping is expressed as

$$T(v) = v - LL^{\top}v \tag{7.1}$$

where L is the load matrix with the principal components for the negative samples in its columns.

The second mapping is a contraction, and works in very much the same way

$$S(v) = LDL^{\top}v \tag{7.2}$$

where

$$D = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \dots & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & 0 & \dots & 0 \\ 0 & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \frac{1}{\lambda_n} \end{pmatrix} \tag{7.3}$$

and the λ_i's are the eigenvalues associated with the principal components. This means that all the original principal components calculated for the negatively labelled samples are normalised to have unit variance. When principal components analysis is performed again on the full data set then this will not be true, but the variation in the negative samples will have been reduced relative to the positive samples.

7.8 Black Box Model Identification and Validation

The objective of this investigation was to see if it is possible to detect urinary tract infections using the relatively fast technique of headspace analysis. The model reduction techniques previously discussed are used in order to obtain robust black box model parameter estimates.

Output	Classification
< -0.5	No Infection
$-0.5 \leq x \leq 0.5$	Not sure
> 0.5	Infection

Table 7.2: Classification thresholds used for urine data

For the RBF neural networks the SVM technique was used to select the number of nodes in the nonlinear layer and the weights in the single node output layer.

The training responses were set to negative for 'no infection' and a positive output for 'infection'. Classification thresholds for test data are shown in Table 7.2. These are the same as those used in the previous chapter.

A common value of σ, the width parameter, for the Gaussian kernels to use on all data sets was required. Thus code was written that tried a range of values and outputted the cross-validated accuracy for each value. These results are not presented here as in a typical execution of the algorithm 100 different values of σ were tested. For an illustration of typical results see Figure 4.5.

ARX type models as well as Multilayer perceptrons were also implemented so that again linear and nonlinear models may be compared.

A data set of 183 points was generated by the experiments and the leave-one-out algorithm was used for identification and validation. Some experimental data were missed due the time set for the length of the experiment in the software for the C320 device. This is not crucial however as it has no effect upon the data that were logged.

7.9 Implementation of the models

The data analysis models considered below in this chapter were implemented using the MATLAB System Identification toolbox. An M-Script was written that automatically considered a range of different order models.

A number of pre-processing algorithms were considered. The first was an attempt to reduce the dimension of the input. Correlation analysis was performed to seek groups

of sensors that were closely correlated. The ARX model is linear in the inputs and outputs, hence the sensors were ordered with respect to the absolute correlation with the output, that is the absolute value of the correlation coefficient. Starting with the highest scoring sensor, the correlation matrix is examined to see which other sensors correlate highly with it, these were then removed.

Secondly linear projections derived from the Sammon map are used. The Euclidean normalisation of data is also considered here.

7.10 Results of Data Analysis

The results of analysis of the C320 data output were as follows. Both fractional and absolute change are considered as static measures of sensor response.

7.10.1 Principal Components Analysis

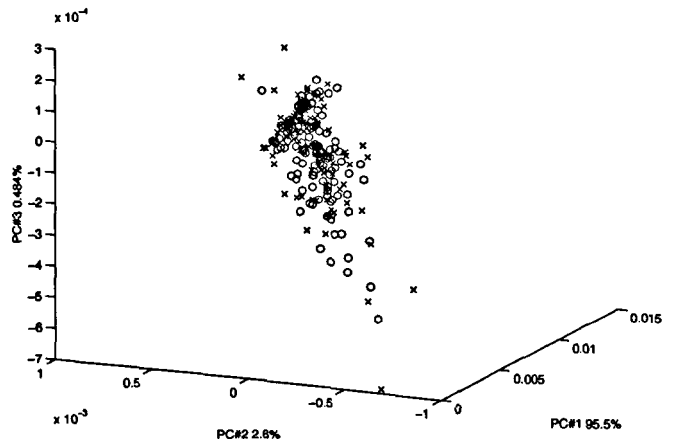
A principal components analysis plot is shown in Figure 7.9(a) using fractional change. Figure 7.9(b) shows the results of using the absolute change in sensor response. The nutrient medium plating techniques discussed above were used as the label. There is a great deal of overlap. Therefore pre-processing techniques are considered.

Figures 7.10(a) and 7.10(b) show the results of applying normalisation to the same preprocessed data set.

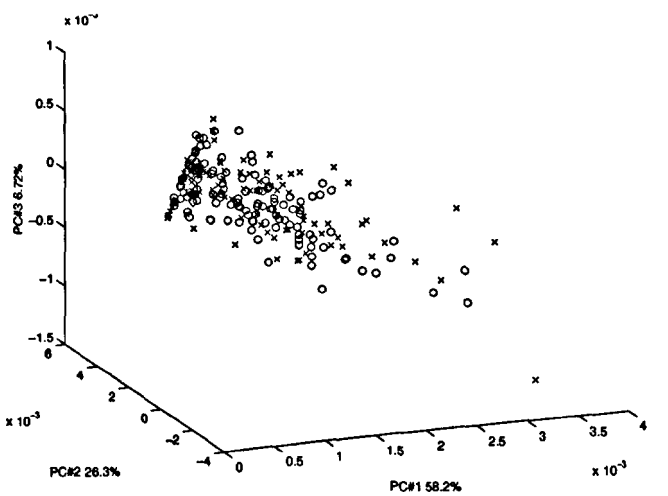
The translational and contraction type maps were applied to the C320 data. Figures 7.11(a) and 7.11(b) show the results of applying the translational mapping. It is clear that this has not aided the separation of the classes. The same is true for the contraction mapping, the results of which are shown in Figures 7.12(a) and 7.12(b).

7.10.2 Black Box Model Results

For the following, all classification success rates are measured by using the leave-one-out algorithm. That is the model parameters are estimated using all but one of the data points, this remaining sample is then used to test the resulting model. The process

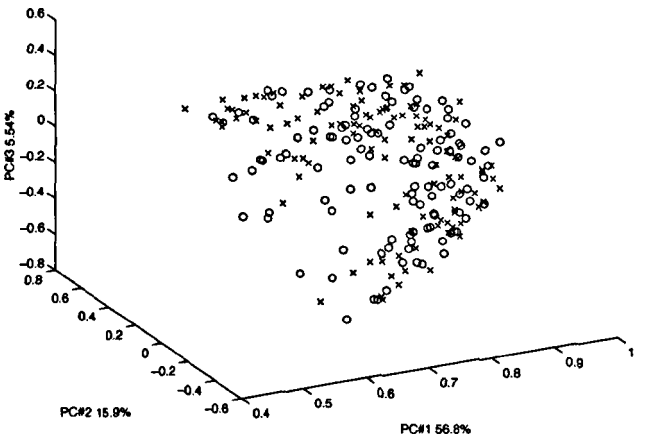


(a) PCA plot of C320 data after using fractional change preprocessing.

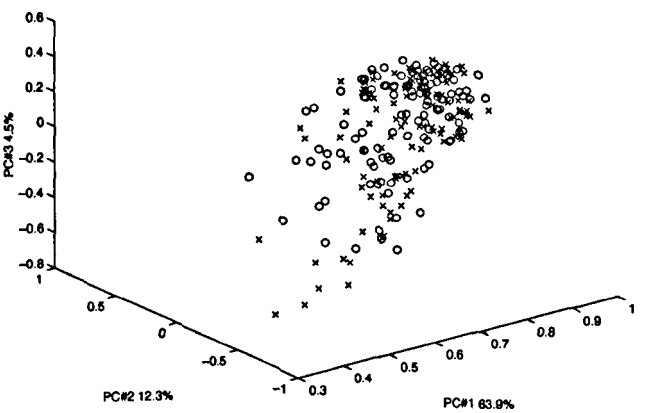


(b) PCA plot of C320 data after using absolute change preprocessing.

Figure 7.9: PCA plot of C320 data where x='infection' and o='no infection'.

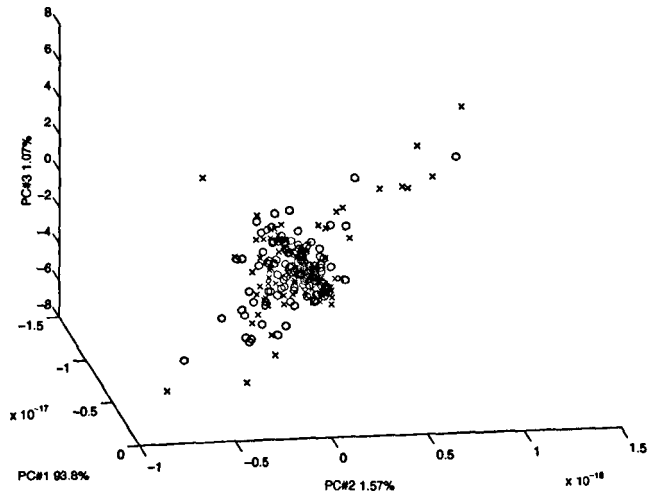


(a) PCA plot of C320 data after applying Euclidean normalisation to the fractional change.

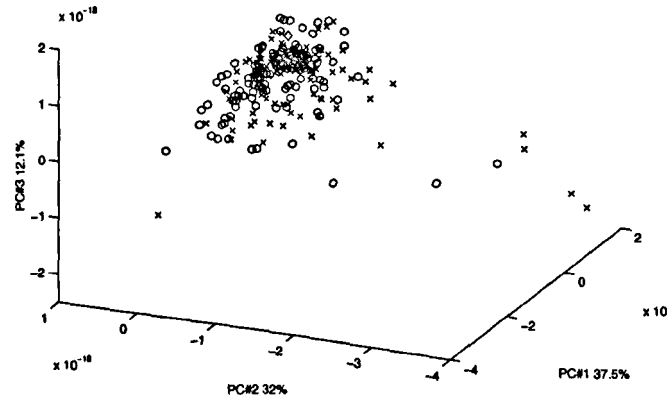


(b) PCA plot of C320 data after applying Euclidean normalisation to the absolute change.

Figure 7.10: PCA plot of C320 data where x=positive and o=negative. Here normalisation has been applied.

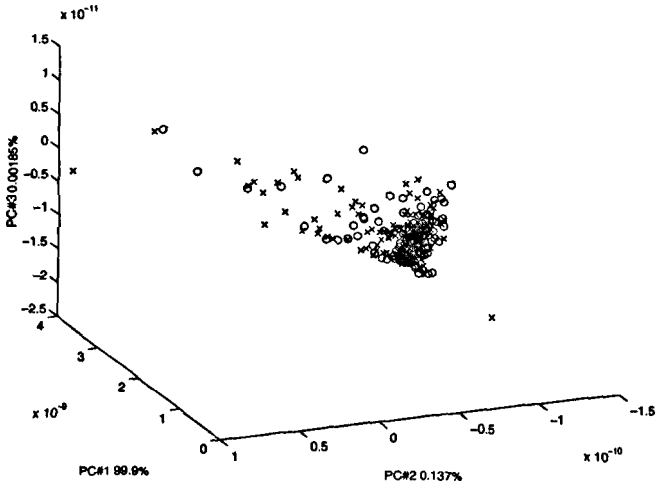


(a) PCA plot of C320 data after applying a translation mapping to the relative change data.

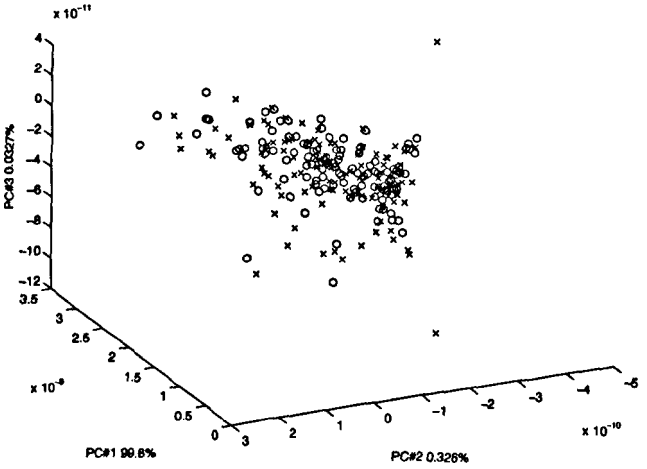


(b) PCA plot of C320 data after applying a translation mapping to the absolute change data.

Figure 7.11: PCA plots of data after translational mapping has been applied.



(a) PCA plot of C320 data after applying a contraction mapping to the fractional change data.



(b) PCA plot of C320 data after applying a contraction mapping to the absolute change data.

Figure 7.12: PCA plots of C320 data after applying a contraction mapping.

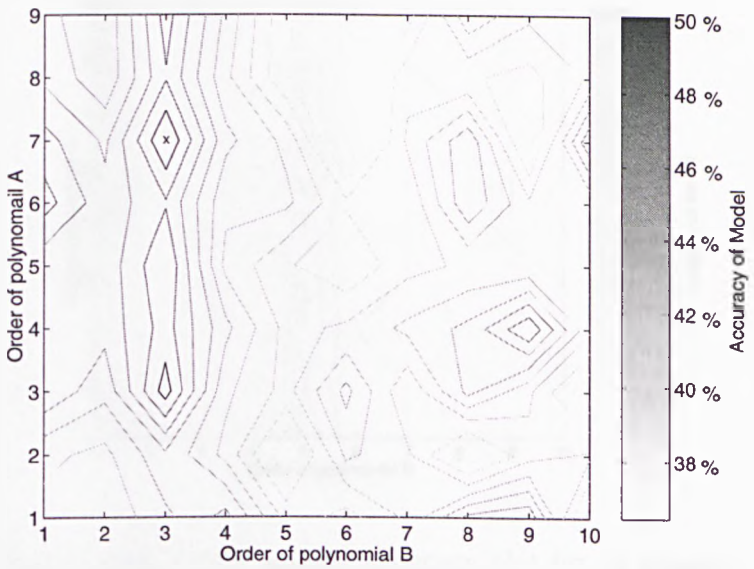


Figure 7.13: Order against accuracy plot for 32 sensors

is carried out with each data point left out in turn. The classification rate is then the average rate over the entire set.

The urine screening experiment resulted in a set of 183 sensor responses from the C320. By using data from all of the sensors an accuracy of 65% successful classifications was attained using ARX as defined in (3.9). The number of sensors was reduced from 32 to 19 by using the most negatively correlating sensors, giving 67% successful classifications. The plots in Figures 7.13 and 7.14 show the relationship between order and accuracy, the crosses represent the positions of peak accuracies. The former is for the full array of 32 sensors and the latter is for the reduced 19 sensor set. Note that once sensor reduction has been performed, lower order models are preferred. This suggests that the sensors that have been discarded contained no information as regards the discriminant task and so only provided some false 'trends' that the model attempted to fit to. Here can be seen the manifestation of an overly complex model. Normalisation gave 60% successful classifications but when coupled with correlation (most negative) technique a rate of 71% was achieved.

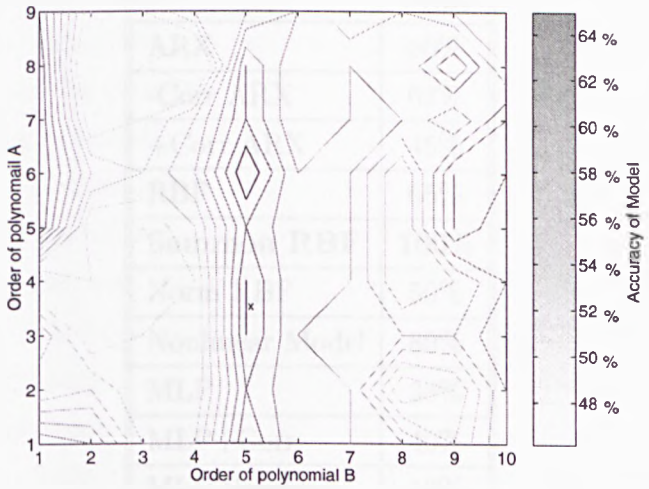


Figure 7.14: Order against accuracy plot for 19 sensors

RBF results on these data were 50% accuracy. Applying these correlation results to RBF networks it was found that a maximum accuracy of 65% was possible.

The hybrid NARX type model was considered and, setting q=3 and r=4 in (3.41), 80% was the maximum successful classification rate achieved. Reducing the number of sensors reduced the successful classification rate to 73%. This was possibly because the hybrid model relies on distance between points, which the extra dimensions increase. Further, it should be noted that Radial Basis Functions have a regularising, smoothing-like property which allows them to avoid over fitting.

The most impressive result was achieved by combining the Sammon map projection with a standard RBF network. A classification rate of 100% successful classifications was achieved. The results are summarised in Table 7.3.

7.11 Conclusion

This chapter is closed with observations on the results obtained above with those of the previous chapter.

Two different types of data sets have been analysed using dimension reduction

	Urine
ARX	60%
-Corr ARX	67%
+Corr ARX	45%
RBF	65%
Sammon RBF	**100%**
Norm RBF	50%
Nonlinear Model	80%
MLP	38%
MLP+Corr	40%
MLP Sammon	10%
MLP Norm	63%

Table 7.3: Summary of Results for analysis of UTI data in terms of successful classifications.

techniques and black box models. A marked change (of the order of 30%) has been observed in the predictive power of the models after judicious pruning of dimensions.

The complexity and variability of the urine samples have been demonstrated. It can be seen that it is difficult to estimate the background chemical noise of the samples. The plots (Figures 7.11 and 7.12) for the negative variation reduction demonstrate this. However the model validation results suggest that it is unnecessary to perform such estimation.

The high order of ARX models necessary for the modelling of the original urine data suggests that the history of experiments needs to be taken into account, this is evidence of a time dependent system. This could mean anything from sensor drift to environmental factors affecting the sampling equipment, and also the samples themselves possibly changing over time in some sense. Why this is so for the urine data and not the blood data considered in the previous chapter could be due to the different sampling equipment used for the two experiments. The urine sample headspace was analysed using a C320 which uses chemical sensors which react with volatile molecules.

The blood data analysed in Chapter 6 were produced using an Agilent 4440, which simply measures the distribution of molecular masses using a quadrapole mass spectrometer. Thus it may be that the C320's response is dependent upon what samples it has interacted with previously.

It appears that correlation techniques work most successfully with linear models, such as ARX, whereas the Sammon map works best with the nonlinear models. This may be due to the nonlinear models using Radial Basis Function neural networks. These functions are spherically symmetric, hence distance information is important. With a linear model, using linear combinations of sensor outputs, correlation between sensors implies redundancy. However, it is unclear why negative correlation is most effective in the case of the blood data and positive correlation is effective in the case of the urine data.

It is possible that the Sammon map works well as it is biased to conserve the distance between mutually remote points rather than neighbouring points, which would be hypothesised to belong to the same data class. Correlation is effective as it 'prunes away' redundant, correlating sensors.

The results for the various models suggest that separation between classes is inherently nonlinear; RBF based techniques faired much better than the linear MLP. In addition the SVM method is designed to avoid over fitting, whereas the MATLAB ARX algorithms are not.

The variation in the success of modelling techniques between this chapter and the last is due to two different systems being considered. The C320 electronic nose and the 4440 mass spectrometer operate on inherently different principles and so it is not surprising that the system behaviour is best represented by different black box model types.

These techniques still beg some questions. The first is that of the optimal dimension of the input into the black box model. As discussed above, lower dimensions result in a much more robust estimate of the parameters (an experimental problem). However, Cover's theorem dictates that a high dimensional feature space is desirable† for maximal

†The basis of the proof of this is the greater degree of freedom afforded models on such spaces-assuming complete knowledge of the system being modelled.

discriminant power. Hence an optimal dimension is a trade off between experimental ability (possibly due to physical constraints upon the observer) and the complexity necessary of a model to explain the input-output behaviour of the physical system 'sufficiently well'.

For the urine data, the least squares error of the Sammon map algorithm can be used to estimate how well the data fits into a given number of dimensions. In this way the 'intrinsic dimension', as discussed in Chapter 4, of the data may be estimated by looking at how much of the topological structure is preserved. The result of this may be compared with how much variance is described by each principal component resulting from PCA. Figure 7.15 shows this comparison for urine data, the same analysis not was performed for the blood data as it was too computationally demanding; the eigenvalues of each principal component are used to denote the variance. It may be seen that both suggest that the urine data may be reduced to just 3 or 4 dimensions.

The Sammon mapping technique focuses on using a family of projections for candidate maps. Thus it was assumed that the relationship between variables was linear. However it is conjectured that a family of nonlinear maps, perhaps RBF type maps, may result in a more efficient reduction in dimension in the case of nonlinear relationships.

It should be noted that the definition of intrinsic dimension given above requires that the form of co-dependence of channels be known. At the very least some assumptions about the form of these relationships have to be made in order to estimate the intrinsic dimension. It seems then that this is another aspect of black box modelling, and requires the same in-depth analysis and investigation that has been afforded input-output models.

The work above suggests that data set reduction can improve the robustness of a model for discriminant analysis and that the specific method used is dependent upon both the data type and the black box model considered. It also demonstrates the success in applying headspace analysis to a biomedical problem.

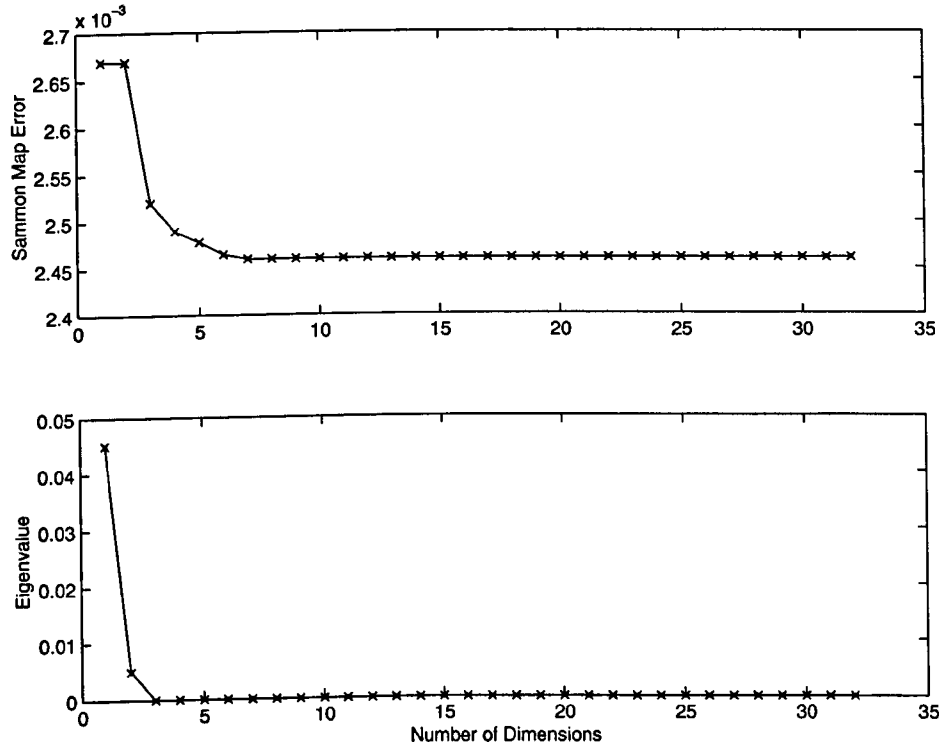


Figure 7.15: The two graphs demonstrate different ways of measuring the intrinsic dimension of the urine data.

Chapter 8

Mechanistic Modelling: The Response of Carbon Black-Polymer Composite Gas Sensors

In the previous chapters it was demonstrated that the response of gas sensors may be analysed in a static mode using black box models. These techniques have proven to be effective in predicting and classifying the form of the input, that is a sample, to the system given the output using inverse model identification. However, it has become evident that the systems under consideration are time dependent and nonlinear with respect to sample concentration. This may be observed in the PCA plots of the last chapter. Much of this dynamical information is lost when processing the response using static measures such as relative or absolute change. It may be possible to retain this important information if the system is analysed in a dynamic manner, the hope being that the discriminant power of the data, in conjunction with some dynamic model, would be greatly increased.

It has been suggested that the dynamic response of a gas sensor contains a great deal more information than the static response. To a large extent this is a vacuous statement as a static response, such as fractional change, may be extracted from dynamics, but the converse is, in general, not true. One way of extracting the dynamic information is to curve fit to experimental data using a mathematical model of the system. Thus in

this chapter, a model is developed that can be experimentally validated. It represents the response of a carbon black-polymer sensor when it is exposed to a constant single species concentration.

First a novel model is derived based upon thermodynamic changes in the system. Having found that it agrees broadly with experimental observations, this is simplified to produce a model that may be used to analyse experimental data. These simplifications are shown to be sensible and be consistent with what is known about the system.

8.1 Physical Knowledge of the Process

It has been noted for many years [133] that when a polymer matrix mixes with a compatible solvent it increases in volume. This is often referred to as a 'swelling effect'. Advantage may be taken of this phenomenon by mixing a polymer with small carbon nanospheres (30nm diameter approximately). If the volume fraction of the carbon is sufficiently large (threshold percentage volume) then the result is that the composite material becomes a good conductor. Swelling, upon mixing with a solvent, will increase the average distance between the nanospheres, and so reduce the conductivity of the material.

Composite materials are interesting substances to model, not least because they exhibit physical properties that are unlike their components. Carbon Black-Polymer materials are polymers which have been mixed with small grains of carbon. If the volume fraction of the carbon is sufficiently large (approximately 25% [134]) then the material is electrically conductive. This watershed in the electrical properties of the material is called the *threshold volume* and is denoted ρ_c.

The polymer component of the material can react and mix with other substances. These can be in a liquid or gas stage and, as this reaction is technically the polymer dissolving into the analyte, shall be referred to henceforth as a *solvent*. Absorption of the solvent causes physical changes to the composite material. In the case below a change in conductivity can be observed. This makes such materials attractive candidates for gas sensor devices, mainly due to the vast range of polymers available with different specifities.

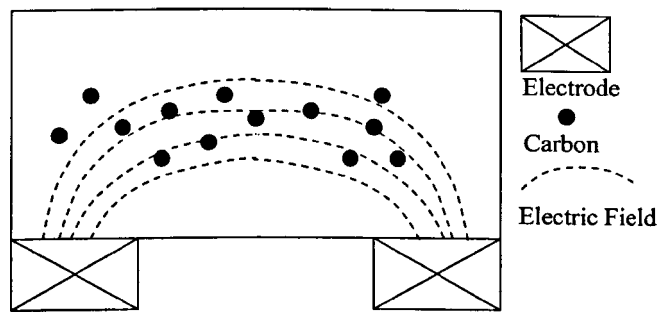


Figure 8.1: Cross sectional view of conduction in Carbon-Polymer material.

Polymers frequently used include Polypyrrole, Polyethylene and Polyvinylchloride [135], [136] and [137]. Blends as well as pure single polymer sensors have been considered experimentally [138] and [139].

Given the wide range of polymers used in such sensors, any dynamic model must encompass the important mechanisms which are common to a broad spectrum of different polymers. In this way the output of electronic noses may analysed and categorised.

The process breaks down into the two stages of solvent uptake and the resulting conductivity change. The conductivity of the polymer sensor is calculable, as demonstrated later, by an integral of the electric flux between the two electrodes and the conductivity as a function of space, see Figure 8.1. The model is inherently one dimensional due to a number of assumed symmetries, these being that the concentration of the solvent is constant across the surface of the sensor and the electric field (see Section 8.6) is constant in the direction parallel to the sensor electrodes. These assumptions greatly simplify this electro-magnetic calculations.

8.1.1 Mass transport in polymers

Early work on solvent solubility in polymers did not consider mass transport, but in the main only considered steady state effects. One of the main, experimentally validated

models is the boiling point model, which will be examined later in this chapter. This type of model predicts the steady state sorption of the composite material using the solubility of a given solvent in the polymer from which the composite is made. In [140] a survey is given of this model as well as an expression relating steady state volume fraction of solvent to boiling point temperature, actual temperature and mixing energy. In that study and also in [141] this model is shown to agree with experimental evidence.

There has been a great deal more experimental work on diffusion in polymers than theoretical studies. With the advent of such devices as Quartz Crystal Micro balances (QCM) there have been detailed studies of transport phenomena. In [142] a study was made comparing the validity of various sorption models, concluding that a Flory-Huggins type model agreed well with experimental evidence. Attempts have been also made to correlate mass transport with the physical geometry of the polymer [143] [144].

Experimental observations have suggested that diffusion is not driven entirely by the concentration gradient. Evidence includes a sharp diffusion front followed by a constant concentration [145], [146]. This rules out the possibility of a Fickian type flow. The major problem with diffusion equations adapted from heat transport is that it is assumed that all concentrations are equally likely, that is, that the medium allows concentrations that can tend to infinity. However, the swelling effect described above, along with associated thermodynamic changes, rules out this possibility. It is clear that these results suggest a variable diffusion rate that is dependent upon solvent concentration.

There have been some previous attempts to model diffusion, such as the steady state models found in [133]. Also there are a number of models considered in [147]. Some work has been carried out to predict the diffusion constant for given solvent-polymer interactions, for example [148] [149]. However these models still have deficiencies [150] and do not explain all experimental data.

For the mass transport of the solvent there are a number of considerations. The first is the cross linking of the matrix, which can introduce nonlinearity into the system via entropy changes. Briefly the solvent mixing necessitates a polymer conformation change, resulting in a volume change. The cross links result in a constraint on this.

Hence it can be demonstrated* that as the cross link density increases, the volume change necessary to reach nonlinear swelling responses lowers. It is therefore assumed, for the moment, that the polymer matrix has no cross links. As a result for small concentrations the response will be linear.

A second problem is the 'non addibility' of volumes in real solutions [151]. By this it is meant that while mass is always conserved in a mixture, volumes are not. The resulting volume is difficult to calculate *a priori* and is due to the polymer's non-ideal nature. Again it is assumed that at small concentrations, and so volumes, the solvent approaches ideality.

In previous models [152] [149], it was assumed that the polymer body did not expand significantly, so that there could be well defined boundary conditions. However, it is apparent that swelling, that has been reported to be in the region of five to six percent, would change the geometry of the polymer film. This should be taken into account as this again will have an effect upon concentration, and so mass transport.

8.1.2 Conduction Models for Carbon Black-Polymer Materials

The electrical properties of carbon black-polymer sensors have been investigated for the last two decades. A number of models have been suggested and these are discussed to justify the particular choice made for the model development.

Initially the interest centred upon the use of carbon black-polymer materials as an anti-static material until their vapour sensitivity was noted [153]. In this paper two conduction effects were considered to coexist: conductivity due the proximity of the carbon granules and electron hopping.

The first effect is achieved by considering the carbon granules forming an RC network. At a low volume ratio of carbon, the conductivity is essentially that of the polymer. By increasing this ratio, a point is reached where the composite material becomes a good conductor. This is referred to as the percolation threshold, v_p, and corresponds to the formation of continuous filaments of carbon granules through the

*In the Flory-Huggins equation the elastic term becomes more dominant over the mixing term

polymer.

Percolation theory [154] is used to describe the distribution of these filaments based on a statistical model of the individual granules' distribution. It is a macroscopic model which is most accurate near the percolation threshold. It also predicts the change in conductivity once the composite material contains a greater volume fraction of carbon black than the threshold value. Hence the conductivity is a function of the volume fraction of the carbon black granules.

Electron hopping, or tunnelling, is a quantum mechanical effect which can take place in the absence of physical contact between the granules. Here an electrical potential difference exists between the granules. Thus this model considers microscopic phenomena and extends it to a macroscopic level by appealing to Effective Medium Theory [155]. Considering a statistical model that covers the whole of the composite material's body it may be seen that the conductivity is again a function of the volume fraction of the carbon granules. The advantage of this model is that it is valid over a much wider range of carbon granule volume fractions than the percolation model.

In [42] it is stated that the primary effect is the percolation effect, however it was pointed out in [153] that this is only true when the volume fraction of the carbon is much higher than the threshold value.

To date much of the work investigating the conductivity of carbon-black polymer sensors has been experimental. The main aim of these experiments has been to characterise the sensitivity, selectivity, and environmental effects of these composite materials as chemoresistors. In [156] the response of CB-PVC to chlorinated hydrocarbons was investigated. The vapour concentration was varied and it was concluded that these materials could be used as gas sensors.

In [42] and [136] this was extended by considering an array of different polymers. Again the vapour concentration was varied, but temperature was kept constant. It was concluded that, up to certain concentrations, the sensors responded linearly.

In [153] [157] it was noted that the temperature effects associated with these materials make composite materials good thermistors. Swelling is a function of temperature also. In [158] it was demonstrated that the material becomes especially sensitive to temperature changes near the polymer melting point. It was also demonstrated that

the volume fraction of carbon dopant was critical in this matter. Near threshold volume fractions, the composite material was sensitive over a larger temperature range.

There has also been work investigating long term effects of the use of this type of gas sensor on the conductivity. Experiments reported in [135] studied the stability of the composite material's conductivity under a temperature cycling. It was concluded that conductivity increased when the material was cycled from ambient temperature, up to near melting point, and down again. However it seems that the temperature/conductivity relationship is stable for lower temperatures.

8.2 Overview of the Model Dynamics

Consider the reaction between a polymer and a single chemical species. It is desirable to ascertain the rate at which a solvent absorbs into a polymer and hence a diffusion model seems sensible. In polymers there are two main effects to consider. The first is the major driving force behind any diffusion effect, that is the flux due to variation in solvent concentration from point to point. The second is a reluctance of the polymer to absorb due to cross linking between polymer strands. Put simply, there is a limit to how much solvent can be absorbed due the strands being 'stretched apart'.

It is expected that the same dynamics come into play when the sensor is 'cleansed' by a purge supply.

The change in conduction is a result of this solvent absorption. The carbon grains will become correspondingly further apart and this will reduce the conductivity. A model will have to be developed that relates this change in conductivity to the 'swell' of the polymer. An electron hopping model is chosen over the percolation model due to its larger domain of validity.

8.2.1 Some assumptions

As with all models, some assumptions have to be made. It is argued that they are both necessary and reasonable assumptions to make; necessary, as they provide a grounding point for our model; reasonable as any variation between the real situation and the

assumptions below would produce small scale local effects. As the more global effects of conductivity are observed only, such small scale phenomena will hopefully be irrelevant. The model described below will therefore assume the following:

Assumption I The composite material is homogeneous. This means that diffusion will progress in the same manner throughout the polymer. It also means that cross-linking is homogeneous in all directions.

Assumption II Diffusion is a reversible process in the sense that the purge stage will leach all solvent molecules from the polymer. This implies there is no bonding process, only an absorption process.

Assumption III The carbon content is above the threshold volume. The content is assumed initially to be sufficiently high so that the absorption effect never takes it below ρ_c. This will avoid discontinuities in the observed electrical conductivity.

Assumption IV Only one chemical species is present in the solvent.

Assumption V The solvent molecules are taken to be spherical. This is because we would like a generic model and it simplifies the modelling process.

The effect observed is a body effect, so that any local inhomogeneities will will difficult to discern from experimental data (see Chapter 9). Assumption II has been observed experimentally many times and is illustrated in the timeseries of Figure 7.8, where the sensor returns to its baseline after 'purging'. Assumption III will be assured during manufacture of the sensors by adjusting the volume fraction of carbon granules. Assumption IV can also be assured during the experimental validation of the model. Assumption V is an implicit assumption as no account is made in the model of interaction at the molecular level. Implicitly, the polymer and the solvent are assumed to be continua.

It is felt that these assumptions simplify the physical system without being too restrictive. It will be shown that they enable an analytical model of conductivity change to be derived.

A schematic of the processes involved within the sensor is shown in Figure 8.2.

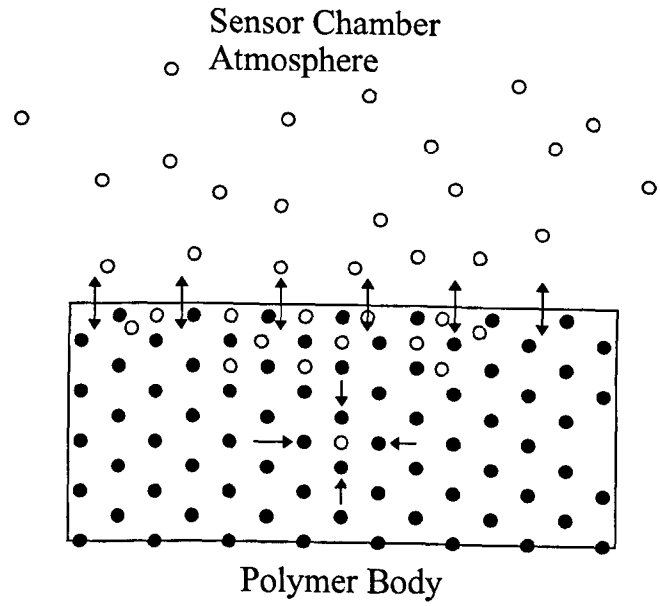


Figure 8.2: Illustration of diffusion in a polymer. Empty circles are solvent molecules and filled circles are carbon granules.

8.3 A Nonlinear Diffusion Model I

Using the assumptions above a nonlinear diffusion model was derived by the author using the thermodynamic concept of chemical potential to drive the solvent flux. A derivation of such a model is given in Appendix A. If $c(x,t)$ is the time and space dependent concentration of the solvent in the body of the sensor and N is the swelling scaling factor then the resulting expression is:

$$\frac{\partial c}{\partial t} = N\left[\frac{v_e}{v_0}\right]\left[\frac{\partial^2 c}{\partial x^2}\left((1-Nc)^{\frac{1}{3}} - \left(\frac{1-Nc}{2}\right)\right) + N\frac{\partial c}{\partial x}\left(1 - \frac{2}{3}(1-Nc)^{-\frac{2}{3}}\right)\right.$$
$$\left. + Nc\left(-\frac{2N}{9}\left(\frac{\partial c}{\partial x}\right)^2(1-Nc)^{-\frac{5}{3}} - \frac{\partial^2 c}{\partial x^2}\left(\frac{1}{2} - \frac{1}{3}(1-Nc)^{-\frac{2}{3}}\right)\right)\right]. \quad (8.1)$$

which is immensely complex. To simplify this highly complex equation we assume an approximation to small concentrations of the solvent and use $(1-x)^{\epsilon} \approx (1-\epsilon x)$ for

small x. This yields

$$\frac{\partial c}{\partial t} = \left[\frac{v_e}{v_0}\right]\left[\frac{\partial^2 c}{\partial x^2}\left(D + \frac{N}{3}c - \frac{2N^2}{9}c^2\right) + \left(\frac{\partial c}{\partial x}\right)^2\left(\frac{1}{3} - \frac{2N}{9}c + \frac{10N^2}{27}c^2\right)\right]. \quad (8.2)$$

for some D. The derivation of this is also provided in Appendix A.

This is an extremely complex model for which as yet no analytical has been found. However (8.2) may be rewritten in the form

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left[D(c)\frac{\partial c}{\partial x}\right]. \quad (8.3)$$

to give

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left[\left[\frac{v_e}{v_0}\right]\left(D' + N\left[(1-Nc)^{\frac{1}{3}} - \left(\frac{1-Nc}{2}\right) + \frac{N}{3}(1-Nc)^{-\frac{2}{3}} + \frac{N}{2}\right]\right)\frac{\partial c}{\partial x}\right] \quad (8.4)$$

for some D'.

The value of this $D(c)$ is shown in Figure 8.3 for small concentrations. The concentration is nondimensionalised as is shown for a random selection of parameter values. Notice that this model predicts the experimentally observed increased diffusion rate in regions of higher solvent concentration.

Boundary value problems with $D(c) = a + bc$ are extremely difficult to solve in general analytically. This would mean that any experimental validation would require some form of computer simulation, possibly using a finite elements numeric solution. Such simulation would be very time consuming, even on a 'fast' computer. It is therefore proposed to use as an approximation a stepwise constant function for $D(c)$, which provides an analytical solution.

8.4 A Nonlinear Diffusion Model II

Experimental evidence confirms the predictions of the above model, that diffusion in a polymer increases in solvent saturated regions. Further, such sharp diffusion fronts have been reported that a discontinuous relationship is implied. The diffusion rate at

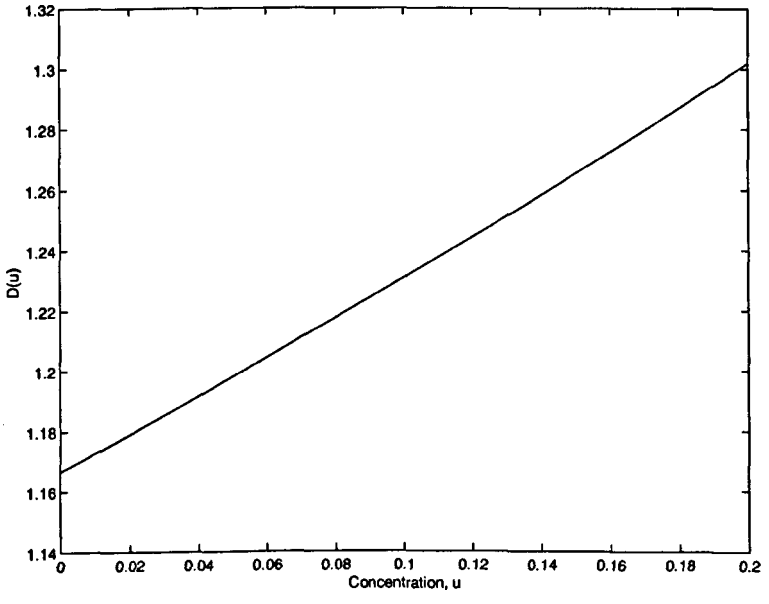


Figure 8.3: Concentration dependent diffusion rate derived from thermodynamic model.

very low concentrations must be very close to zero to produce such a phenomenon. This assumed system behaviour may be analysed analytically, to produce a model that may be validated using experimental data.

8.4.1 Concentration dependent diffusion with one discontinuity

A first approximation to the nonlinear diffusion process would be to set up a linear diffusion model with two region of high and low concentrations of solvent molecules. These two regions, delineated by concentration C_X, could then have two different diffusion rates, D_1 and D_2 associated with them. That is, the diffusion is modelled as

$$\frac{\partial c}{\partial t} = D(c)\frac{\partial^2 c}{\partial x^2} \tag{8.5}$$

where

$$D(c) = \begin{cases} D_1 & \text{if } c(x,t) \leq C_X, \\ D_2 & \text{if } c(x,t) > C_X. \end{cases} \tag{8.6}$$

Further, at $x = 0$ there is a constant concentration of C_1. In the sensor case, this is related to the solvent concentration in the sensor chamber. This form of diffusion is well studied and a good review may be found in [159]. Here the point of discontinuity of the diffusion rate in the polymer body, $c(x,t) = C_X$, is denoted $x = X_p(t)$ (this will be shown to be so below in this section). Figure 8.4 shows a schematic of this; the diffusion rate alters as solvent molecules diffuse through the surface of the polymer. The thickness of the polymer is marked as being 'infinity' as initially solutions to a semi-infinite domain problem are discussed.

Analysis of this problem is reasonably simple via a solution originally proposed by Neumann and first used in [160]. The proposed solution is

$$c_1(x,t) = C_1 + A\text{erf}\frac{x}{2\sqrt{(D_1 t)}}, \tag{8.7}$$

$$c_2(x,t) = C_2 + B\text{erf}\frac{x}{2\sqrt{(D_2 t)}}. \tag{8.8}$$

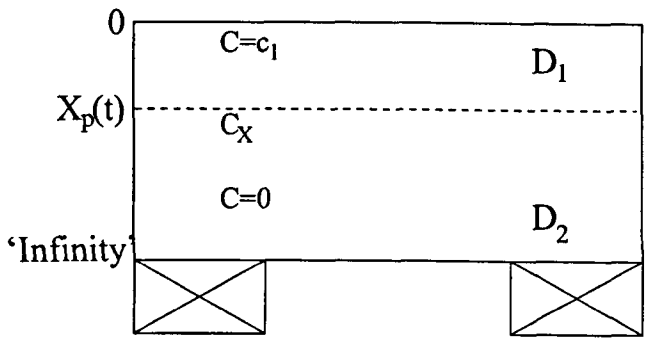


Figure 8.4: Discontinuous diffusion coefficient $X(t)$ represents the position of the diffusion front.

For constants A, B, C_1, C_2. Here $c_1(x,t)$ is the concentration profile in the high concentration region (between air-polymer interface and boundary $X_p(t)$) and $c_2(x,t)$ is the concentration profile in the corresponding low concentration region.

From this we have

$$C_X = c_1(X_p(t),t) = C_1 + A\text{erf}\frac{X_p(t)}{2\sqrt{(D_1 t)}}. \tag{8.9}$$

For this to hold we must have

$$X_p(t) = k_p t^{\frac{1}{2}} \tag{8.10}$$

and the constants A, B, C_1, C_2 and k are determined by the boundary conditions and the two diffusion constants. This will be shown formally later in this section.

The assumption here is that the sorption and swelling of the polymer is much quicker than the diffusion process. Further it is assumed that the actual domain, the polymer film, does not change shape. However it is apparent from experimental evidence that this is not so. A model is therefore also considered where the domain changes.

8.4.2 The General Moving Boundary Problem in a Semi-infinite Medium

A more general moving boundary model may be formulated to take into account the moving interface due to swelling. The case found in [159] is presented, where the

movement of this interface is proportional to the amount of substance transferred across the interface. It is necessary to introduce two coordinate systems that are stationary with respect to the two mediums.

Consider the polymer and the atmosphere, initially separated by the planes $x_1 = x_2 = 0$, to be separated at time t by the planes $x_1 = X_1(t), x_2 = X_2(t)$ (illustrated in Figure 8.5). There are three further conditions for this model:

- Equilibrium at the interface;

$$c_1(X_1) = Qc_2(X_2) + R \tag{8.11}$$

  for parameters Q and R.

- Conservation of the solvent the interface;

$$D_1\left(\frac{\partial c_1}{\partial x_1}\right)_{x_1=X_1} - D_2\left(\frac{\partial c_2}{\partial x_2}\right)_{x_2=X_2} + c_1(X_1)\frac{dX_1}{dt} - c_2(X_2)\frac{dX_2}{dt} = 0 \tag{8.12}$$

- Constant of proportionality of movement,

$$X_2 = PX_1. \tag{8.13}$$

In the full model developed below it will be seen that by setting $R = 0$ in Equation (8.11), Henry's law is obtained, where Q is the solubility of the gas. Further it will be seen that the assumption embodied in Equation (8.13) is superfluous as $P = 0$; the atmosphere coordinate system remains stationary.

Using the particular solutions (8.7) and (8.8) and considering initially that the interfaces X_1 and X_2 are fixed it is deduced that:

$$\frac{c_1(\infty) - c_1}{c_1(\infty) - c_1(0)} = 1 - \text{erf}\left[\frac{x_1}{(D_1 t)^{\frac{1}{2}}}\right], \tag{8.14}$$

$$\frac{c_2(-\infty) - c_2}{c_1(-\infty) - c_2(0)} = 1 + \text{erf}\left[\frac{x_2}{2(D_2 t)^{\frac{1}{2}}}\right]. \tag{8.15}$$

These are solutions for semi-infinite media, though it will become apparent that they also will satisfy (8.11), (8.12) and (8.13).

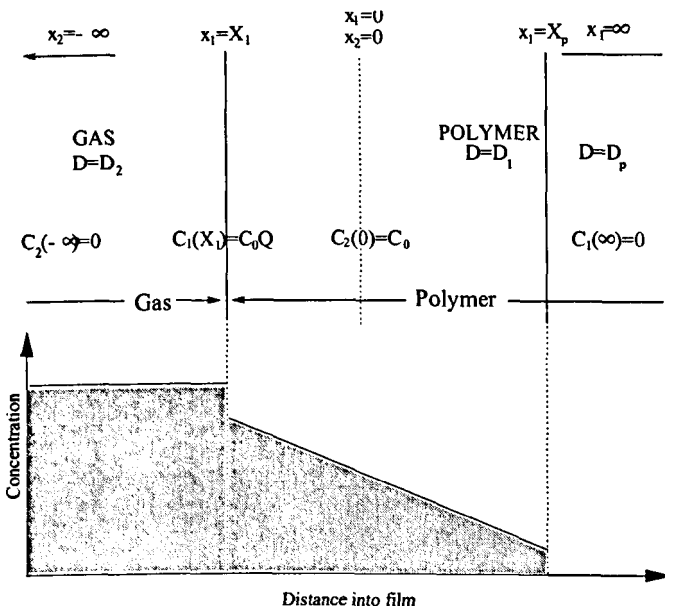


Figure 8.5: The general moving boundary problem. The two bold vertical lines represent the air-polymer interface X_2 and the diffusion rate discontinuity X_1.

Now the case is considered where the movement of the media relative to the boundary is caused by the transfer of diffusing substance across the interface. This is referred to as the Case A situation in [159]. It is only necessary to specify two out of five boundary conditions: $c_1(\infty), c_1(0), c_1(X_1), c_2(-\infty), c_2(0)$ [159]. In accordance with the assumption that the interface moves due to the influx of substance, this movement is expressed as

$$\frac{dX_1}{dt} = S\left\{D_1\left(\frac{\partial c_1}{\partial x_1}\right)_{x_1=X_1} + c_1(X_1)\frac{dX_1}{dt}\right\}, \tag{8.16}$$

$$\frac{dX_2}{dt} = SP\left\{D_2\left(\frac{\partial c_2}{\partial x_2}\right)_{x_2=X_2} + c_2(X_2)\frac{dX_2}{dt}\right\}. \tag{8.17}$$

Here in both equations the first term in the bracket represents the flux across the interface due to the concentration gradient and the second represents flux due to the interface moving relative to the diffusing substance. S is a constant that relates concentration to volume.

Combining these with (8.12) and (8.13) gives,

$$\frac{dX_1}{dt}\left\{\frac{1}{S}-c_1(X_1)\right\} = \{c_1(\infty)-c_1(0)\}\left(\frac{D_1}{\pi t}\right)^{\frac{1}{2}}\exp\left[-\frac{X_1^2}{4D_1t}\right] \quad (8.18)$$

$$\frac{dX_1}{dt}\left\{\frac{1}{S}-PR-Qc_1(X_1)\right\} = \{c_2(0)-c_2(-\infty)\}\left(\frac{D_2}{\pi t}\right)^{\frac{1}{2}}\exp\left[-\frac{P^2X_1^2}{4D_2t}\right] \quad (8.19)$$

where the rightmost two terms of (8.18) represent the derivative of the error function with respect to t. Also by setting $x_1 = X_1$, $c_1 = c_1(X_1)$ in (8.14) the following is obtained

$$c_1(X_1) = c_1(0) + \{c_1(\infty) - c_1(0)\}\text{erf}\left\{\frac{X_1}{2(D_1t)^{\frac{1}{2}}}\right\}. \quad (8.20)$$

It is apparent that Equations (8.18), (8.19) and (8.20) are satisfied, if and only if X_1 is of the form

$$X_1 = 2\alpha(D_1t)^{\frac{1}{2}}. \quad (8.21)$$

Of course, here diffusion in the polymer is constant at a constant rate D_1. It is therefore necessary to combine the two models above in a novel way to simulate the process of diffusion in polymers.

8.4.3 Model with Discontinuous Diffusion Constant and Expanding Domain

In the following medium 1 (all with subscripts 1) is taken to be to be the swollen polymer and medium 2 (all with subscripts 2) denotes the atmosphere. Subcripts p represent unswollen polymer regions in medium p. Note X_2 does not change as the atmosphere does not deform. The time dependent interface between the atmosphere and the polymer is denoted $X_1(t)$ and the concentration front in the polymer by $X_p(t)$.

Note that the concentration profile in the atmosphere is of little interest, but it is present to provide boundary conditions and the moving interface.

The boundary problem is specified thus:

$$\frac{\partial c_1}{\partial t} = D_1 \frac{\partial^2 c_1}{dx^2}, \tag{8.22}$$

$$\frac{\partial c_2}{\partial t} = D_2 \frac{\partial^2 c_2}{dx^2}, \tag{8.23}$$

$$\frac{\partial c_p}{\partial t} = D_p \frac{\partial^2 c_p}{dx^2}. \tag{8.24}$$

Here D_1, D_p are the diffusion constants in the polymer for high and low concentrations respectively. That is, in the polymer

$$D = \begin{cases} D_p & \text{if } c_1(x_1) < C_X, \\ D_1 & \text{if } c_1(x_1) \geq C_X. \end{cases} \tag{8.25}$$

The boundary conditions are set as

$$D_1 \left(\frac{\partial c_1}{\partial t}\right)_{x_1=X_1} - D_2 \left(\frac{\partial c_2}{\partial t}\right)_{x_2=X_2} + c_0 Q \frac{dX_1}{dt} = 0 \tag{8.26}$$

$$D_1 \left(\frac{\partial c_1}{\partial t}\right)_{x_1=X_p} - D_p \left(\frac{\partial c_p}{\partial t}\right)_{x_1=X_p} = 0 \tag{8.27}$$

and initial conditions are specified to be:

$$\begin{aligned} c_2(0,t) &= c_0 \\ c_2(X_2,t) &= c_0 \\ c_2(-\infty,t) &= 0 \\ c_1(X_1,t) &= c_0 Q \\ c_p(\infty,t) &= 0 \\ c_1(X_p,t) &= C_X. \end{aligned} \tag{8.28}$$

where X_p represents the position of the diffusion front within the polymer.

Note that these are point initial conditions at the interface. It would be preferable to set more realistic conditions such as a constant concentration in the atmosphere

(assuming good mixing). The following form for the particular solution is considered

$$c_1(x,t) = C_1 + A\left[\text{erf}\left[\frac{x_1}{2(D_1t)^{\frac{1}{2}}}\right] - \text{erf}\left[\frac{X_p}{2(D_1t)^{\frac{1}{2}}}\right]\right] \tag{8.29}$$

$$c_2(x,t) = C_2 + B\text{erf}\left[\frac{x_2}{2(D_2t)^{\frac{1}{2}}}\right] \tag{8.30}$$

$$c_p(x,t) = C_p + C\text{erfc}\left[\frac{x_1}{2(D_pt)^{\frac{1}{2}}}\right] \tag{8.31}$$

where A, B, C, C_1, C_2, C_3 are parameters whose values have to be deduced.

8.5 Analysis of the Model

A series of analyses for different boundary conditions and assumptions of semi-infinite or finite media are presented below.

8.5.1 Semi infinite media

Given the candidate solutions (8.29), (8.30), and (8.31) the conditions under which they satisfy the boundary value problem above are analysed. Note first that Equations (8.22), (8.23) and (8.24) are satisfied as the error function is a standard solution of the diffusion equation for point initial conditions. Applying the boundary conditions (8.28) yields:

$$c_1(X_1,t) = c_0Q = C_1 + A\left[\text{erf}\left[\frac{X_1}{2(D_1t)^{\frac{1}{2}}}\right] - \text{erf}\left[\frac{X_p}{2(D_1t)^{\frac{1}{2}}}\right]\right] \tag{8.32}$$

$$c_1(X_p,t) = C_X = C_1 + A[0] \tag{8.33}$$

$$\text{Giving } C_1 = C_X \tag{8.34}$$

$$c_2(-\infty,t) = 0 = C_2 - B \tag{8.35}$$

$$c_2(0,t) = c_0 = C_2 \tag{8.36}$$

$$\text{Giving } C_2 = c_0 \tag{8.37}$$

$$\text{and } B = c_0 \tag{8.38}$$

$$c_p(-\infty, t) = 0 = C_p + 2C \tag{8.39}$$

$$\text{Giving } C_p = -2C \tag{8.40}$$

$$c_p(X_p, t) = C_X = \frac{1}{2}C_p\left(1 + \text{erf}\left[\frac{X_p}{2(D_p t)^{\frac{1}{2}}}\right]\right), \forall t \geq 0. \tag{8.41}$$

For Equations (8.33) and (8.41) to hold is necessary that

$$X_1 = 2\alpha(D_1 t)^{\frac{1}{2}} \tag{8.42}$$

$$X_p = 2\beta(D_1 t)^{\frac{1}{2}} \implies \frac{1}{C_p} = \frac{1}{C_X}\text{erf}\left[\frac{D_1^{\frac{1}{2}}\beta}{D_2^{\frac{1}{2}}}\right]. \tag{8.43}$$

Applying the flux condition (8.26) and comparing with Equation (8.16) gives

$$A\left(\frac{D_1}{\pi t}\right)^{\frac{1}{2}} \exp\left(-\frac{X_1^2}{4D_1 t}\right) - B\left(\frac{D_2}{\pi t}\right)^{\frac{1}{2}} + c_0 Q\frac{dX_1}{dt} = \frac{1}{S}\frac{dX_1}{dt} - B\left(\frac{D_2}{\pi t}\right)^{\frac{1}{2}} = 0. \tag{8.44}$$

Comparing (8.45) with (8.42) gives

$$\alpha = \left(\frac{D_2}{D_1}\right)^{\frac{1}{2}} \frac{Sc_0}{\pi^{\frac{1}{2}}}. \tag{8.45}$$

Applying the flux condition (8.27) and substituting to equation (8.43) gives

$$A\left(\frac{D_1}{\pi t}\right)^{\frac{1}{2}} \exp\left(-\frac{X_p^2}{4D_1 t}\right) + C_p\left(\frac{D_p}{\pi t}\right)^{\frac{1}{2}} \exp\left(-\frac{X_p^2}{4D_p t}\right) = AD_1^{\frac{1}{2}} \exp(-\beta^2) + C_p D_p^{\frac{1}{2}} \exp\left(-\beta^2\frac{D_1}{D_p}\right) = 0. \tag{8.46}$$

Finally substituting (8.42), (8.43) and (8.45) into Equation (8.32)

$$c_0 Q - C_X = A\left[\text{erf}\left(\left(\frac{D_2}{D_1}\right)^{\frac{1}{2}} \frac{Sc_0}{\pi^{\frac{1}{2}}}\right) - \text{erf}(\beta)\right] \tag{8.47}$$

Thus from (8.41) and (8.47) one can obtain

$$A = \frac{c_0 Q - C_X}{\left[\operatorname{erf}\left(\left(\frac{D_2}{D_1}\right)^{\frac{1}{2}} \frac{Sc_0}{\pi^{\frac{1}{2}}}\right) - \operatorname{erf}(\beta)\right]}. \tag{8.48}$$

and

$$c_p = \frac{C_X}{\operatorname{erf}\beta \left(\frac{D_1}{D_2}\right)^{\frac{1}{2}}}. \tag{8.49}$$

Therefore all parameters, except β, have been expressed in terms of the initial conditions and the diffusion constants. However, by substituting (8.48) and (8.49) into the flux condition (8.27) this yields

$$\frac{c_0 Q - C_X}{\left[\operatorname{erf}\left(\left(\frac{D_2}{D_1}\right)^{\frac{1}{2}} \frac{Sc_0}{\pi^{\frac{1}{2}}}\right) - \operatorname{erf}(\beta)\right]} D_1^{\frac{1}{2}} \exp\left(-\beta^2\right) + \frac{C_X}{\operatorname{erf}\beta \left(\frac{D_1}{D_2}\right)^{\frac{1}{2}}} D_p^{\frac{1}{2}} \exp\left(-\beta^2 \left(\frac{D_1}{D_p}\right)\right) = 0 \tag{8.50}$$

which is an equation in β alone. It is apparent however that this equation will have to be solved numerically for β due to the presence of error functions making this equation transcendental.

From here the case of a polymer film of finite thickness is considered and it is shown how β may be ascertained in another manner.

8.5.2 Finite medium with sharp diffusion front

The polymer film is assumed to be of finite thickness, y_0, so that initially the domain of the polymer is the interval $[-y_0, 0]$. Hence to move to a consideration of a finite medium the following condition in (8.28)is changed to

$$c_p(-y_0, t) = 0. \tag{8.51}$$

The following is also assumed so as to allow an analytical solution to the finite medium problem

$$D_p = 0. \tag{8.52}$$

This stops rapid transport of the solvent past the base of the polymer film. Hence the flux condition at the X_p boundary becomes [159]

$$-A\left(\frac{D_1}{\pi t}\right)^{\frac{1}{2}}\exp\left(-\frac{X_p^2}{4D_1 t}\right) = C_X\frac{dX_p}{dt}. \tag{8.53}$$

Assuming the same form for the boundary, X_p, as in 8.43 then the following is deduced,

$$\frac{-2A}{C_X\pi^{\frac{1}{2}}} = \beta\exp\left[\beta^2\right] \tag{8.54}$$

and in the notation above (Equation (8.31))

$$C_p = -C\text{erf}\left[\frac{-y_0}{2(D_p t)^{\frac{1}{2}}}\right]. \tag{8.55}$$

That is that $c_p(x,t) = 0$ is the only possible solution of the form (8.31), which is to be expected given the diffusion rate is zero in this region.

Figure 8.6 shows the time evolution of solvent concentration throughout the polymer film as it expands into the atmosphere. The film/ air boundary is set at zero in the space coordinates, the film is represented by the interval $[-10, 0]$ and the air by $[0, 2]$. It can be seen how the film expands into the air domain as the diffusion front moves through the film.

8.5.3 Steady state solutions in the swollen polymer

One simplification that will prove beneficial when calculating the current between the electrodes of the sensor (discussed in Section 8.6) is that of quasi-steady state in the swollen region of the polymer. Taking the initial and boundary conditions (8.28) as before, an approximate solution is proposed to be

$$c_1(x,t) = [c_0 Q - C_X]\left[\frac{X_p - x}{X_p - X_1}\right] + C_X. \tag{8.56}$$

in the swollen region of the polymer.

It is immediately obvious that this expression satisfies equation (8.22) and the initial conditions (8.28). The flux conditions at X_1 and X_p need to be satisfied and so (8.26)

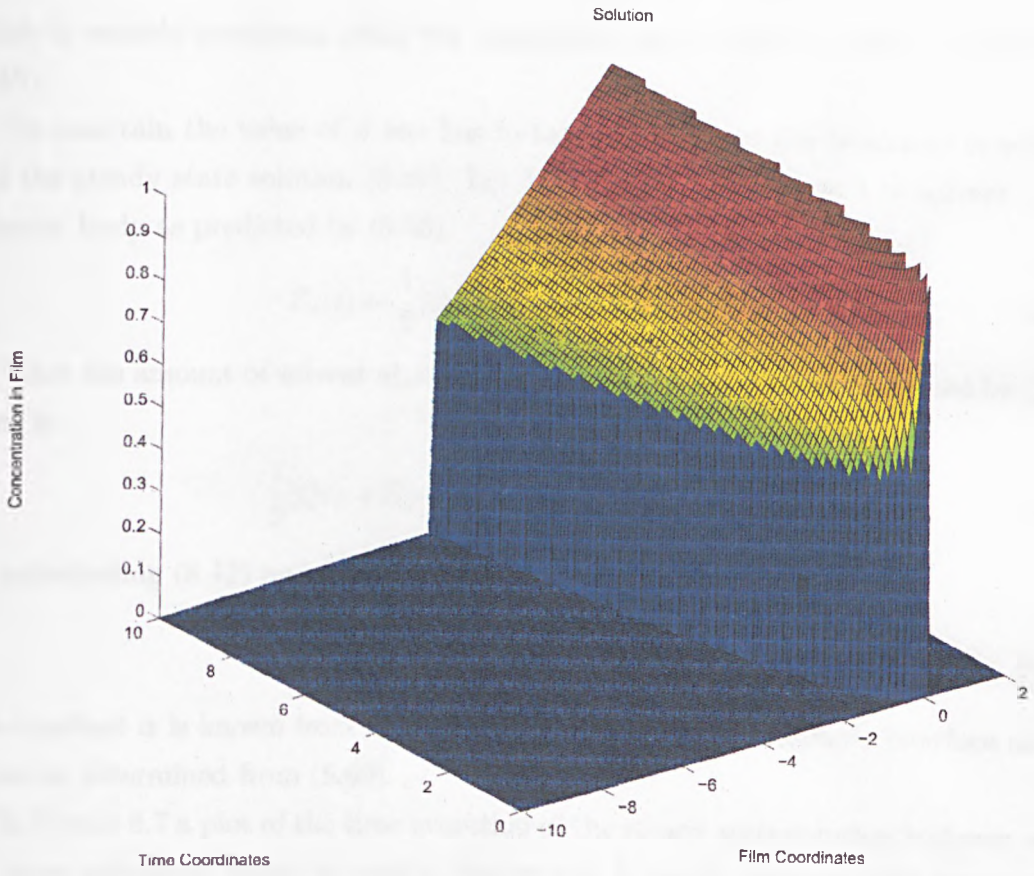


Figure 8.6: Plot of solution to finite media problem. Here $c_0 Q = 1$, $C_X = 0.5$ and $\alpha = 0.1$, $\beta = -1$. The axes are nondimensional as the simulation is for random parameter values.

implies that

$$-D_1\left[c_0Q - C_X\right]\frac{1}{X_p - X_1} + c_p\left(\frac{D_p}{\pi t}\right)^{\frac{1}{2}}\exp\left(-\frac{X_p^2}{4D_pt}\right) = 0 \tag{8.57}$$

which is entirely consistent using our expressions for X_1 and X_p given in (8.42) and (8.43).

To ascertain the value of β one has to take into account the boundary conditions and *the steady state solution* (8.56). Let $\Sigma_c(t)$ be the total amount of solvent in the polymer body as predicted by (8.56)

$$\Sigma_c(t) = \frac{1}{2}[Qc_0 + C_X][X_1(t) - X_p(t)]. \tag{8.58}$$

Then the amount of solvent absorbed at the boundary X_1 must be gained be $\Sigma_c(t)$. That is:

$$\frac{1}{2}[Qc_0 + C_X]\left[\frac{dX_1}{dt} - \frac{dX_p}{dt}\right] = Qc_0\frac{dX_1}{dt}. \tag{8.59}$$

By substituting (8.42) and (8.43) we get

$$\beta = -\frac{1}{2}\frac{Qc_0}{C_X}\alpha. \tag{8.60}$$

The constant α is known from conditions at the polymer/atmosphere interface and so β can be determined from (8.60).

In Figure 8.7 a plot of the time evolution of the steady state solution is shown using the same parameter values as used in Figure 8.6. It can be observed that the solution is qualitatively and quantitatively similar to the analytical solution.

8.5.4 Notes on the quasi-steady state solution

The linear expression (in x) in Equation (8.56) is only an approximation: it does not satisfy the diffusion equation itself. However the quasi-steady state has been used in a number of studies to great effect [161][162]. In [163] the error involved was investigated. Let F be the concentration of free sites for the diffusing substance within the polymer, and let C_b be the concentration at the boundary (within the polymer). The errors which were calculated using numerical methods are summarised in Table 8.1.

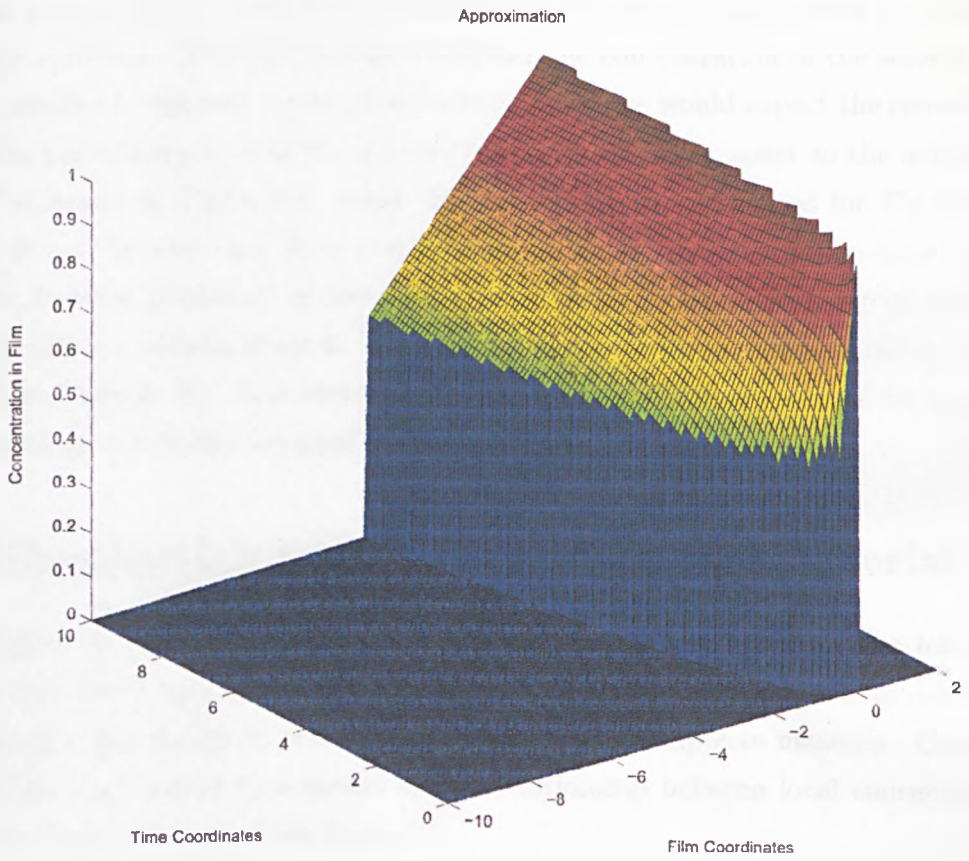


Figure 8.7: Plot of steady state solution to finite media problem. Here $c_0 Q = 1$, $C_X = 0.5$ and $\alpha = 0.1, \beta = -1$. The axis are nondimensional as the simulation is for random parameter values.

$\frac{F}{C_b}$	Relative error
> 10	$< 1\%$
$= 10$	$\approx 1\%$
$= 5$	$\approx 5\%$

Table 8.1: Error values for the quasi-steady state solution.

It can be seen that, if there is an abundance of free sites, then the error is relatively small. For a solvent diffusing through a polymer the concentration of the solvent will be much smaller in relation to the polymer body, thus one would expect the error to be small. The percentage error of the steady state solution with respect to the analytical solution is shown in Figure 8.8, using the same parameter values as for Figures 8.6 and 8.7. It can be seen that there is a peak error of 2.5% (excess with respect to the analytical solution predicted by the steady state solution) which propagates into the film. This is not a very large error, but its effect will be seen when fitting the model to real data in Chapter 10. This steady state approximation will be used to express the time dependent concentration profile of the solvent in the polymer.

8.6 Conduction within the Composite Material

An expression for the time and space dependent concentration of the solvent has been derived using the steady state approximation (8.56), it can now be used to derive an expression for the change in resistance of the polymer composite material. Consider a single electron hopping type model for the relationship between local concentration and conductivity. This is of the form

$$\sigma = \sigma_0 \exp(-\chi_s) \tag{8.61}$$

where σ is the conductivity of the polymer, σ_0 is the conductivity of the virgin polymer and χ_s is a function of the local swelling factor. By assuming small concentrations ($\chi_s \approx 0$), (8.61) may be linearised to obtain

$$\sigma \approx \sigma_0(1 - \chi_s) \tag{8.62}$$

In this model it is assumed that χ_s is proportional to the volume change, ΔV_p, which is proportional to the local solvent concentration, c.

$$\chi_s \propto \Delta V_p \tag{8.63}$$

$$\propto c \tag{8.64}$$

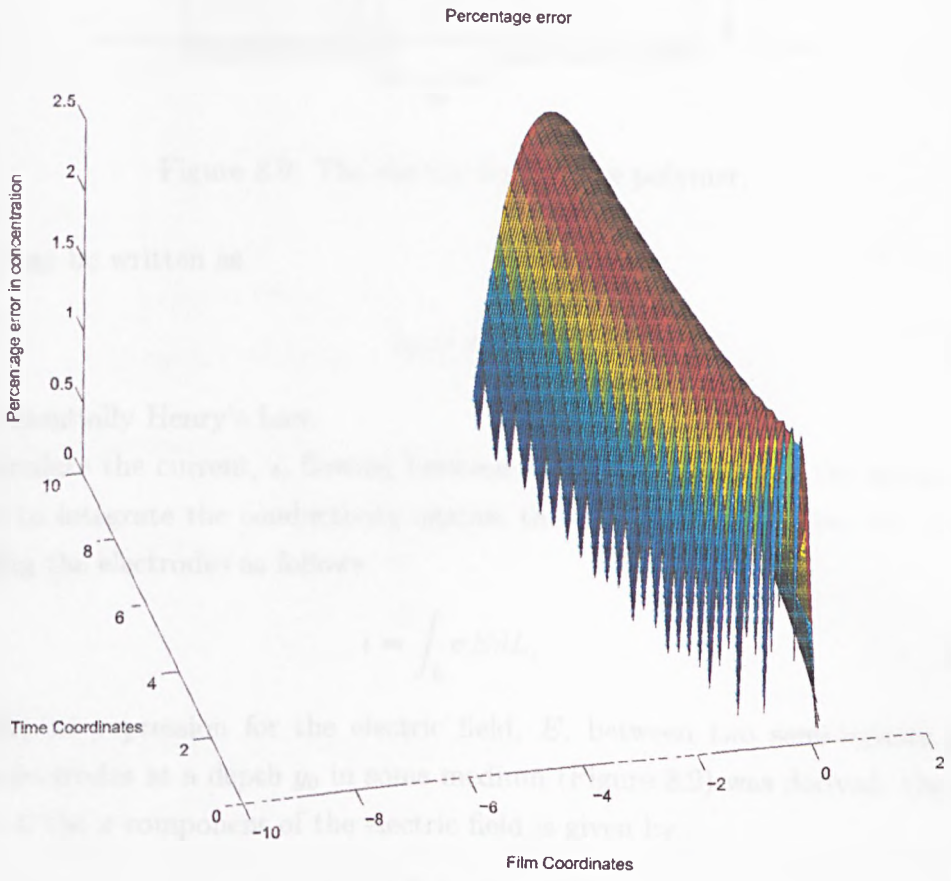


Figure 8.8: Plot of error of steady state solution. Here $c_0Q - 1$, $C_X - 0.5$ and $\alpha - 0.1$, $\beta = -1$

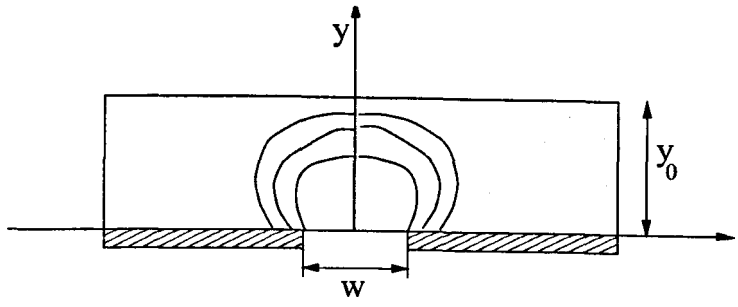


Figure 8.9: The electric field in the polymer.

Thus χ_s may be written as

$$\chi_s = Nc \tag{8.65}$$

which is essentially Henry's Law.

To calculate the current, i, flowing between the two electrodes of the sensor, it is necessary to integrate the conductivity against the electric field, E, over the surface, L, bisecting the electrodes as follows

$$i = \int_L \sigma E dL. \tag{8.66}$$

In [164] an expression for the electric field, E, between two semi-infinite plane coplanar electrodes at a depth y_0 in some medium (Figure 8.9) was derived. Over the plane $x = 0$ the x component of the electric field is given by

$$E(0, y) = \frac{V}{\pi}\left[y^2 - \frac{w^2}{4}\right]^{-\frac{1}{2}}, \tag{8.67}$$

In the geometry of a finite film this must be adapted to

$$E(0, y) = \frac{V}{\pi}\left[(y + y_0)^2 - \frac{w^2}{4}\right]^{-\frac{1}{2}} \tag{8.68}$$

because in the reference frame of the diffusion model, the surface of the film is (initially) at $x = 0$ and the electrodes are at $x = -y_0$.

Parameter Label	Physical Characteristic
D_1	Diffusion rate in virgin polymer
D_2	Diffusion rate in chamber atmosphere
V	Potential across electrode
σ_0	Initial conductivity of sensor
c_0	Initial solvent concentration in the atmosphere
Q	Solubility of solvent in polymer
C_X	Threshold concentration for diffusion rate switch.
N	Constant of proportionality between concentration and volume change
y_0	Initial film thickness

Table 8.2: Summary of Parameters in Model Output

Thus the corresponding current is given by

$$i = \frac{V}{\pi}\left[\int_{-y_0}^{X_p}\sigma_0\left[(y+y_0)^2-\frac{w^2}{4}\right]^{-\frac{1}{2}}dy+\int_{X_p}^{0}(1-Nc_1)\sigma_0\left[(y+y_0)^2-\frac{w^2}{4}\right]^{-\frac{1}{2}}dy\right.$$
$$\left.+\int_{0}^{X_1}(1-Nc_1)\sigma_0\left[(y+y_0)^2-\frac{w^2}{4}\right]^{-\frac{1}{2}}dy\right]. \quad (8.69)$$

Assuming steady state, as in Equation (8.56)

$$i = \frac{V\sigma_0}{\pi}\left[N(c_0Q-C_X)\frac{1}{X_1-X_p}\left[X_p\left[\cosh^{-1}(X_1+y_0)-\cosh^{-1}(X_p+y_0)\right]-\right.\right.$$
$$\left[\exp(\cosh^{-1}(X_1+y_0))-\exp(\cosh^{-1}(X_p+y_0))-y_0\left[\cosh^{-1}(X_1+y_0)\right.\right.$$
$$\left.\left.\left.-\cosh^{-1}(X_p+y_0)\right]\right]\right]-\log 2y_0-NC_X\cosh^{-1}(X_p+y_0)+1\Bigg]. \quad (8.70)$$

The parameters involved in this model are summarised in Table 8.2. A typical plot of the solution defined by (8.70) is shown in Figure 8.10.

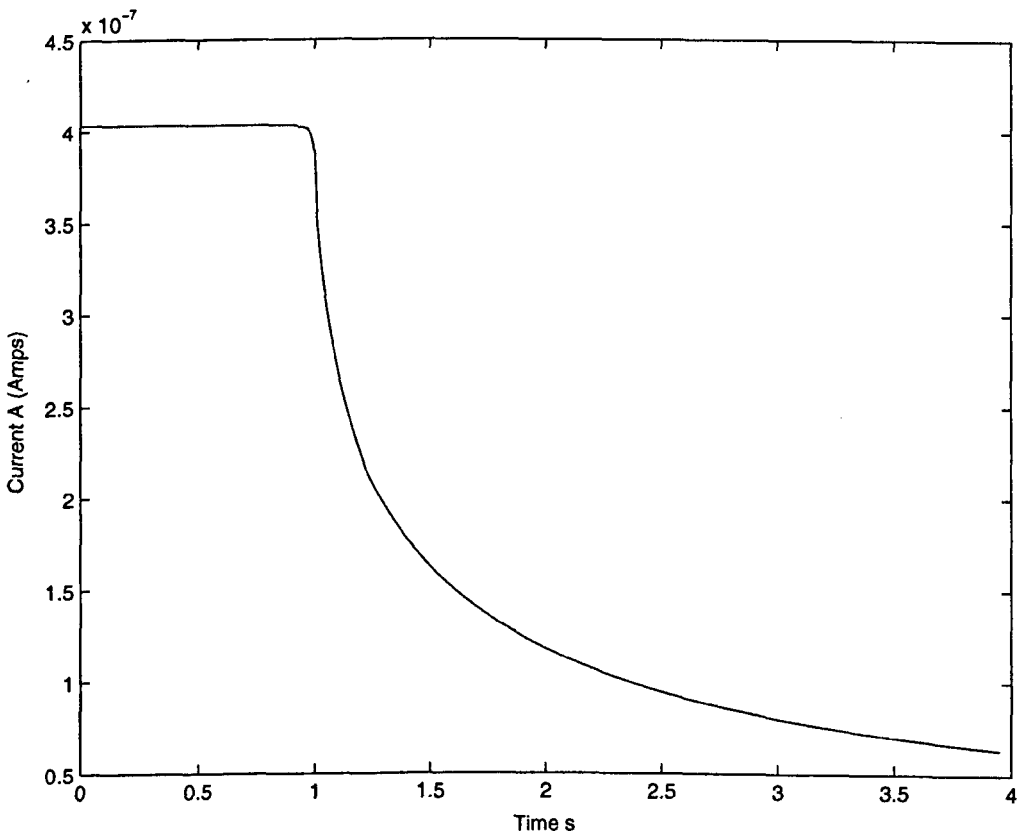


Figure 8.10: Plot of time line defined by Equation (8.70).

8.7 Estimation of Parameters from Experimental Data

It has been shown above that the parameters in the diffusion model are uniquely determined by the initial conditions. The next question is which of these may be experimentally estimated by observing the current flow. Equation (8.70) has parameters representing several physical characteristics, and these are summarised in Table 8.2. However,they appear in such a way that they may not be uniquely determined, for example the product Nc_0 occurs in several places, but the two parameters involved do not appear separately. The analysis is from the perspective of a single polymer-solvent interaction.

First, set

$$X_1 = k_1 t^{\frac{1}{2}} \tag{8.71}$$

$$X_p = k_p t^{\frac{1}{2}} \tag{8.72}$$

as per Equations (8.42) and (8.43)

As a further simplifying assumption, set

$$C_X = \gamma c_0 Q \tag{8.73}$$

because it is assumed that the threshold for the diffusion coefficient discontinuity will be set by the boundary condition.

It is assumed that γ, relating C_X as a fraction of $c_0 Q$, is set prior to the experiment.

For the parameters N, Q, and c_0 it may be seen by inspection that only the product NQc_0 appears, and thus only Q and so only this product may possibly be estimated.

The same argument holds for $V\sigma_0$ in (8.70) and the fact that y_0 may be estimated. Giving a total of six groups of parameters that may be estimated.

The parameters that are most 'important' are those that characterise the interaction of a given solvent with a given polymer: these being the diffusion rates, the solubility and swelling parameters. These parameters may be considered important as they are dependent only upon the material's properties and not the specific physical set up,

such as film thickness. This means they have use as far as the identity of an unknown solvent, use in a modified model, and validation of the model theory.

From the analysis above it seems that if a sensor is considered in isolation only, then the swelling parameter cannot be uniquely identified. In the next Chapter an array of sensors is used to validate the model in order to reduce the number of degrees of freedom in the free parameters.

8.8 Dependency of Parameters on Temperature and Humidity

In the above model it has been implicitly assumed that the system operates at a fixed temperature and humidity. To allow experimental work to investigate the validity of the model over a range of conditions the effects that these environmental conditions will hypothetically have on the parameter values are considered. It is assumed that variation of these parameters is over a much larger time scale than each experimental cycle so that the correlation of model parameters with temperature, etc., may be investigated.

8.8.1 Temperature

The conduction model (8.61) may be expanded to give

$$\sigma = \sigma_0 \exp(-k(S_0 + \Delta s)) \tag{8.74}$$

where S_0 is the 'normal' gap in the virgin polymer. It has been observed in [135] that there is a good linear relationship between temperature and conduction. Therefore a dependence of S_0 with the temperature in the form of a linear expansion model is incorporated:

$$S_0 = S_{00}\left[1 + \theta(T - T_0)\right] \tag{8.75}$$

where S_{00} and θ are constants of proportionality and T_0 is the baseline temperature.

In [144] the correlation between a number of different parameters within polymer solvent-polymer transport was considered. It was found experimentally that the rate of diffusion, and so α and β, increased with temperature.

The prevalent model for the relationship between temperature, T, and the partition coefficient, Q, is the boiling point model and takes into account certain properties of the two interacting substances. The exposition given in [141] is followed, amending where necessary to maintain consistent notation.

Let ρ_1 and M_1 be the density and molecular weight respectively of the the stationary phase. Let p_2 be the saturation vapour pressure of the solute vapour, γ_2 is the vapour activity coefficient (=1 for an ideal solution). The following relationship holds

$$Q = \frac{\rho_1 RT}{\gamma_2 M_1 p_2}. \tag{8.76}$$

Using Trouton's rule and the Clausius-Clapeyron equation, the following expression relating p_2 to T_b the vapour boiling point results [165]

$$\log p_2 \approx 7.7 - T_b \left(\frac{t}{2.303RT} \right) \tag{8.77}$$

where t is the Trouton coefficient for the vapour. Thus, by taking natural logarithms of both sides of Equation (8.76) and combining with Equation (8.77) the following expression is obtained:

$$\log Q = C' + T_b \left(\frac{t}{2.303RT} \right) \tag{8.78}$$

where C' is a constant. This was found to fit well with experimental findings [141]. Thus this would be the relationship one would expect to see between Q for the polymer and a similar relationship would be observed for the atmospheric solvent concentration, as the liquid phase in the sample jar has fixed concentration.

8.8.2 Humidity

The effects of humidity, water vapour etc, represent the interaction of a solvent with the polymer body. It is assumed that the variation of concentration, that is humidity, is such a long term trend that a steady state has been reached within the polymer.

It has been found [166] that there is a linear to sigmoidal relationship between relative humidity at the surface and the steady state concentration of water within the polymer body up 80% of relative humidity. Furthermore *in-plane* mechanical strain was measured against concentration and, more importantly, surface relative humidity. It was found that that the relationship was linear to slightly concave, this concavity perhaps resulting from nonlinearities in expansion near saturation concentrations. Thus humidity is introduced as another linear effect on swelling and modifying (8.75) accordingly gives:

$$S_0 = S_{00}\left[1 + \theta_T(T - T_0) + \theta_H \Phi\right] \tag{8.79}$$

where Φ is the relative humidity.

Thus, if the model is correct, the above relationships between parameters and environmental conditions should be observed.

8.8.3 Processing Techniques

A review of relevant literature shows that the processing techniques, such as curing time and temperature, affect the response of the polymer to environmental factors. The sensors used experimentally were made under the same conditions, so it is hoped that environmental parameters such as S_{00} and θ_T will be constant for each type of sensor. Other processing influences will not be detectable as all sensors will be made under the same conditions.

8.9 Discussion

A more mechanistic approach has been taken to build a theory of how a given sensor should react to a solvent, given the initial conditions; these initial conditions being concentration at the boundary, the constant of swelling, and the mixing parameter. Though the model did give general, experimentally verified, predictions, the resulting PDE was highly nonlinear. It was necessary then to simplify it.

A simplified model for diffusion of a solvent in a polymer (composite) was then derived. A steady state version has been coupled with a conduction model to give a time

dependent expression for the current flowing between the electrodes of a chemoresistor type sensor. It has been demonstrated that the parameters of the diffusion model solution are all functions of the initial conditions and the diffusion coefficients.

There are a number of aspects of the physical system that have not been incorporated into the model. The first that the models take no account for cross links in the polymer structure, this allows for unlimited swelling which does not occur in reality [142]. The second is accounting properly for a finite medium with a nonzero diffusion coefficient in the unswollen state. A solution for this case has not yet been found. However, the analysis has resulted in a number of testable predictions and offers an analytical solution that may be handled by numeric optimisation algorithms.

As demonstrated, it is necessary to estimate the threshold at which the diffusion coefficient should change value, and the values of these different coefficients. There are a number of ways of approaching this. There are experimentally obtained values provided in the literature. For example [144], which contains values for various polymers at different temperatures and in swollen and unswollen states. The values given do confirm the hypothesis of the above models as the diffusion rates differ by up to an order of magnitude between unswollen and swollen states.

In the next chapter, the design of the validating experiments will be considered. This will include the construction of an automated rig for gathering data and parameter estimation methods for data analysis.

Chapter 9

Experimental Validation of the Mechanistic Sensor Model

In this chapter the methods available to validate the mechanistic sensor model developed in Chapter 8 are examined. A model is a mathematical abstraction; that is an attempt to embody any knowledge of the 'actual' system. The aims of such a model are many: to give a greater depth of knowledge of the system; to estimate quantities which cannot be detected directly; to predict the behaviour of the system. For the model of a carbon black-polymer sensor, these criteria are of prime interest. For a model of such a system one needs to be sure that the proposed model is capable of satisfying these criteria.

Carbon black-polymer sensors transduce chemical information into an electrical signal. A model coupling solvent transport with a conduction model for polymer composite materials has been developed in the previous chapter. The result has been an analytical expression relating the current flowing through a sensor to various parameters which characterise the physical and chemical aspects of a specific solvent/polymer interaction.

In such a phenomenological model the parameters relate to certain characteristics of the system. Due to the abstraction of the parameters in such models it is very difficult to elicit the validity of such concepts as 'knowledge' or physical constants, unless, of course, some prior information is known. What is a surer test is the demand that the

system matches future behaviour of the system. In this case, it is considered that the model 'fits' the input-output behaviour of the system.

In order to validate this model, and its underlying assumptions, it is necessary to test it against experimentally obtained data. To this end, the sensor model must be incorporated into a model of a specific electronic nose system model. In this way, experimental data may be produced to validate the model and actual behaviour may be compared with that which is predicted by the model.

Hence, by 'fitting' two issues are covered:

1. For a given output there is a response resulting from the model which approximates 'closely', relative to some error function, the output. That is to say that the model accurately predicts the real system's behaviour with respect to some criteria.

2. The parameters estimated from this response have some dependence on the initial conditions of the experiment. By this it is meant that some dependence may be observed between model parameter values and such factors as temperature of the samples and the humidity of the air.

Good experimental design is required in order to validate the model. If, at all possible, the experiment must work within the constraints under which the model was developed. The model should be validated on its 'own terms'. In that way results are meaningful within the context of the theory, that is to say that certain measurements and parameters obtained from the experiment have some 'physical meaning' within the theory and can be thought to characterise the system's mathematical abstraction. Otherwise, the perceived failure of the theory is nonsense as this failure was outside the context of the model; it only failed to be general enough.

9.1 Overview of the Theoretical System

The full electronic nose system can be broken down into four basic processes [167]: odour production, odour transport, sensor response, and the electronic interface. The

latter part has been chiefly dealt with in the sensor model. The first part of this chapter is concerned with how the odour production and transport result in initial conditions at the surface of the sensor.

Odour production is a thermodynamic process whereby the solvent exists in two different phases: liquid and as a vapour in the carrier gas. The solvent will move between these two phases in order to achieve equilibrium. When the carrier gas is pumped through the odour production vessel, taking the solvent vapour away, the system will be constantly adjusting itself in an attempt to obtain equilibrium.

Odour transport is a combination of fluid flow and diffusion of the odour in a carrier gas. In this respect it is analogous to the process of gas chromatography in a packed column. If in this case the odour is a mixture of several chemical species then they will all diffuse at different rates. For this reason a single species mixing with air will be considered.

9.2 Odour Production

It is assumed that the odour concentration in the sample vessel will tend towards equilibrium over the time scale of the experiment. This equilibrium will be dictated by a number of conditions: the flow rate through the fluidics, the temperature of both the solvent sample and the ambient environment; the size of the sample vessel. With a continuous flow rate the equilibrium will be reached in some finite time. However, with the type of experimental rig that is considered in this chapter, this is not the case. *Flow through the sample vessel only exists when introducing the sample to the sensors.* Hence in this situation there is a constant cycle of concentration lying between static equilibrium and that for carrier gas flow.

It is necessary, therefore, to calculate the concentration of solvent vapour that finds its way into the atmosphere of the production vessel. This is in order to provide boundary conditions for the sensors. The concentration cannot be measured directly in the sample vessel, but can only be observed indirectly in the sensor response.

Two aspects of the evaporation process need to be understood. Firstly, one needs to determine the equilibrium concentration in the atmosphere of the sample vessel when

there is no flow in that part of the fluidics. Secondly, one needs to determine the rate of evaporation when the concentration is perturbed from equilibrium by carrier gas flow. As the main aim of this thesis is the modelling of electronic nose systems, and primarily the sensors, a radical new model of evaporation will not be presented. Instead, in order to understand the effects of environmental parameters on the system's input-output behaviour, a first order kinetic model will be considered.

Figure 9.1 displays a linear one compartment model of solvent vapour generation and elimination. Generation is through evaporation and elimination through condensation and fluidics transport. This system could be written with a second compartment for the liquid phase of the solvent to replace elimination via condensation, but, assuming limited volume changes, the evaporation rate will be constant. The system equation is thus given by

$$\dot{x}_A = k_{LA}x_L^0 - (k_{AL} + k_{AF})x_A \tag{9.1}$$

where

$$k_{AF} = \frac{\dot{M}}{V} \tag{9.2}$$

is the rate of elimination due to the fluidics flow and V is the volume of the gas phase,$\dot{M}$ is the flow rate in the fluidics, in Ls^{-1}, and k_{LA}, k_{AL} are, respectively, the rates of evaporation and condensation. Here x_A and x_L^0 are the concentrations of the solvent in the atmosphere and liquid phase respectively. The initial conditions for this system are derived by considering the system equilibrium with no flow. Thus the initial condition is that

$$x_A(0) = \frac{k_{LA}}{k_{AL}}x_L^0 = Q_s x_L^0 \tag{9.3}$$

where Q_s is the partition coefficient and as such is subject to the boiling point law which predicts a variation with temperature. The solution to this model is

$$x_A(t) = \frac{k_{LA}x_L^0}{k_{AL} + k_{AF}}\left(1 - e^{-(k_{AL}+k_{AF})t}\right) + \frac{k_{LA}}{k_{AL}}x_L^0 e^{-(k_{AL}+k_{AF})t}. \tag{9.4}$$

Note that the asymptotic equilibrium will be

$$x_A^{eq} = \frac{k_{LA}}{k_{AL} + k_{AF}}x_L^0 \tag{9.5}$$

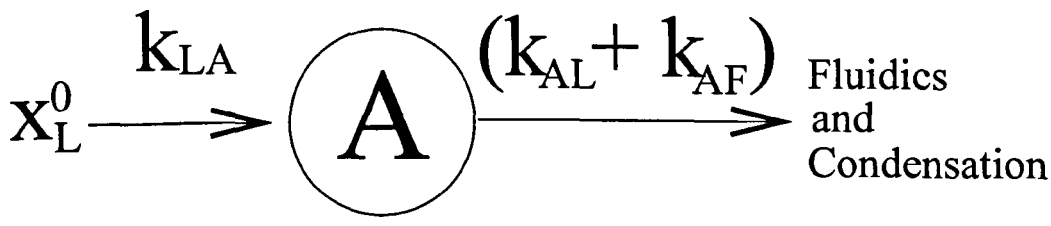


Figure 9.1: One compartment model representing odour production

and so the the concentration that the sensors will 'see' varies with time, as predicted by Equation (9.4). Thus, as the mechanistic sensor model assumes a constant solvent concentration in one of its boundary conditions, the model parameter corresponding to concentration will be some form of time average. Thus it can be seen that the estimated concentration parameter will be a function of temperature, flow rate and sampling time.

9.3 Fluidics

The transport of the solvent vapour to the sensor surface involves an array of pipe-work and valves, Figure 9.4. It is necessary to find a relationship between the concentration of vapour in the sample vessel and that in the sensor chamber. This is so any influence that the flow rate, fluidic design and environment have on the system response can be understood.

9.3.1 Odour Transport

The transport of the vapour involves the flow of fluid within a pipe of a certain diameter and length. This will incorporate a time lag between the impulsive input at the start of the sampling and the solvent reaching the sensor surface. The question is how this is related to the instantaneous concentration in the production vessel.

Fluid flow within a pipe is still very much an experimental area of research. Usually theoretical work deals with the flow at a 'large' distance down the pipe. This is the concept of *fully developed flow* [168]. Here there is a continuous variation in flow

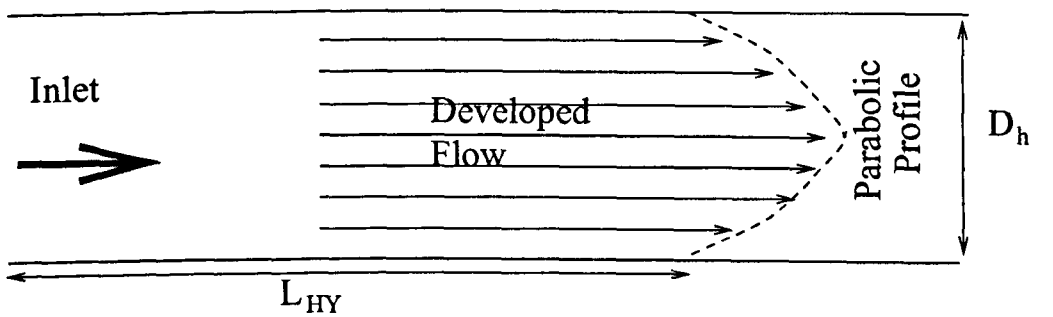


Figure 9.2: Developed laminar flow. An initial inlet profile, to the left, develops into a parabolic velocity profile to the right.

from the tube wall to the centre, where the flow rate is at a maximum. However, the behaviour at the entry to the tube is very different where it is more of a 'plug like' flow, that is a constant velocity profile. In [168] and [169] the length required for a fully developed velocity profile is given by

$$\frac{L_{HY}}{D_h} = 0.59 + 0.0056 R_e \tag{9.6}$$

where D_h is the hydrodynamic diameter in metres, L_{HY} is the length before developed flow occurs and R_e is Reynolds number. Figure 9.2 illustrates this process.

Under fully developed flow the radial velocity profile is parabolic in nature and it is derived in [168] that the average velocity of the fluid is half that at the centre of the pipe. All of this is under the assumption that $R_e \leq 1000$. The tubing used in the rig discussed in this thesis has a diameter of 3 mm and the flow rate is of the order of 100 mm per minute. This gives $Re \approx 100$ so there will be developed laminar flow after 2 or 3 millimetres of the entry into the tubing.

There is a problem if diffusion is significant in comparison with the flow rate in the fluidics, if so then plug flow will not be present. The initial plug profile will spread out due to diffusion. The extent of this effect is reliant upon the flow rate as this dictates the time it takes for the sample to travel along the pipe. Thus again the observed concentration at the surface of the sensors will be a function of flow rate.

Secondly, different solvents will diffuse at different rates and so there will be different time delays present in the initial conditions.

A final consideration is the delay produced by the length of piping and the pressure necessary to drive the fluid along it. The delay will be a linear function of the pipe length and will also have an inverse relation to the flow rate. The tubing will affect the flow rate via the pressure drop along its length. The pressure drop is expressed as

$$\Delta P \propto \frac{l}{d^4} \tag{9.7}$$

where P is pressure, l is the length and d is the diameter of the pipe in metres. Hence pressure drop is linearly proportional to length and thus delay can be expressed as the square of the length of the pipe.

9.3.2 Mixing in the sensor chamber

The final consideration is how the solvent vapour mixes with the 'clean' atmosphere in the sensor chamber. The sensor chamber may be considered to be part of the fluidic system and contribute towards the time delay discussed above, assuming good mixing in the sensor chamber. The rate of mixing will depend upon the structure of the flow. An ideal chamber would be one that preserves the fluid flow, allowing the structured parabolic flow to continue. This will help to produce a step change in solvent concentration at the sensor surface.

9.4 Sensor response

The response of the sensors, given certain initial conditions at the surface of the polymer, has already been modelled in Chapter 8. The expression derived needs to be integrated into the above lumped system model. The response is monitored by measuring the change in resistance using interface electronics.

It is thought best to design a rig which reproduces the necessary constant boundary conditions used in the derivation of the model. In this way the model may be validated and used to analyse the data. In practice the best that can be attained is a system with pure time delay, however it is more likely to achieve a system where the sample concentration variation with time has a ramp profile with a steep slope, see Figure 9.3. This is possible with careful sample production and transport criteria.

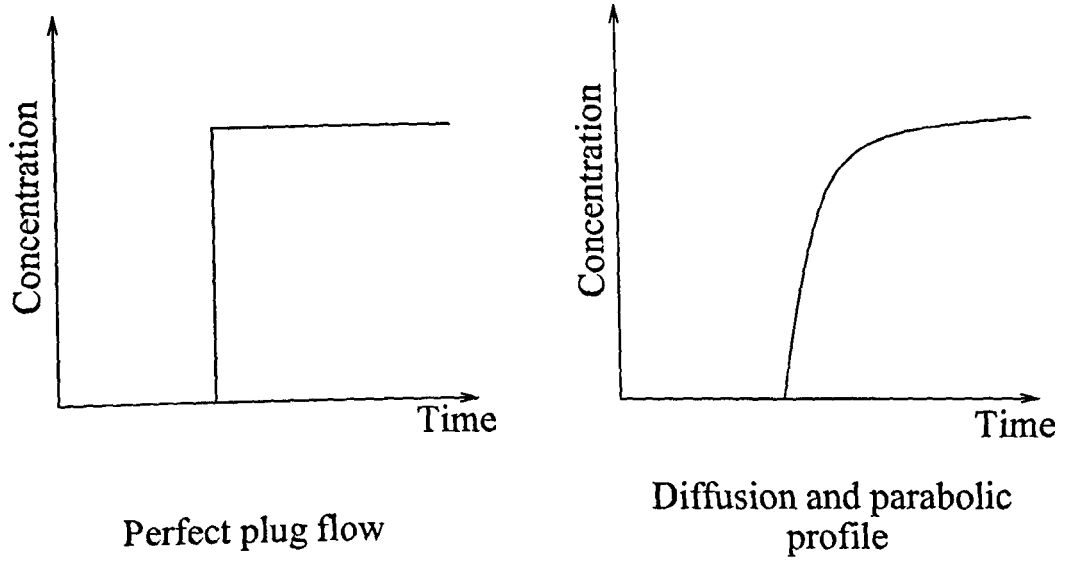


Figure 9.3: Time dependence of solvent vapour concentration in the sensor chamber. The graphs are for illustration only and consequently have no units. The time frame is of the order of seconds.

9.5 The Experimental Test bed

The test bed is a simple electronic nose rig and its basic design has been tested a number of times [170], [171] with success. As discussed in the introduction to this chapter, the main aim of the test bed is to characterise the response of gas sensors and in order to do this it imitates the basic function of a mammalian nose. Conceptually this process includes the production of some odour external to the nose and transport (nasal passage) to the sensor where the odour is detected (olfactory sensors). The response is observed via the nervous system and processed by the brain.

In an experimental set up these processes are mimicked by an electro-mechanical system under the control of a microcomputer. The external odour production is that of solvent vapours, transport is by use of an electric pump and the fluidic system and odour detection is performed by the sensors. The response of the sensor is then compared with simulated responses from the mathematical model developed in the previous chapter.

The modular design may be summarised as follows:

- Sample jars containing solvents
- Fluidics and pump
- Selector Valves
- Sensor chamber and sensors
- Environmental sensors
- Interface cards
- Data logging.

9.5.1 Odour Production

The odours are intended to be single species vapours from certain solvents. Each liquid sample and its headspace is contained within a glass jar. The lid of each jar has two holes drilled into it and then are fitted an inlet with an inline check valve and an outlet connecting to the fluidics of the electronic nose system, leading to the gas sensors. Leakage of solvent into the environment is avoided by using one way slip valves on the intake. To increase odour production the inlet is fitted with a sparge (see Chapter 7).

A number of choices presented themselves: how many different odour samples, what substances to sample, what volume of headspace is required, and at what temperature should these samples be kept at? A constant headspace concentration was required as a constant boundary condition is a an assumption of the model (Equations 8.28).

The vapour pressure of a substance is the pressure at which the liquid and vapour phases exist at equilibrium and an approximation is readily calculable using Trouton's law. Vapour increases with temperature. Boiling occurs when the vapour pressure reaches the atmospheric pressure so it can be seen that the concentration of the solvent in the atmosphere above the liquid phase should increase as well.

It was decided to compare two solvent vapours. The two sample vessels (Figure 9.4) were heated by the environment. To monitor this a temperature sensor was set to

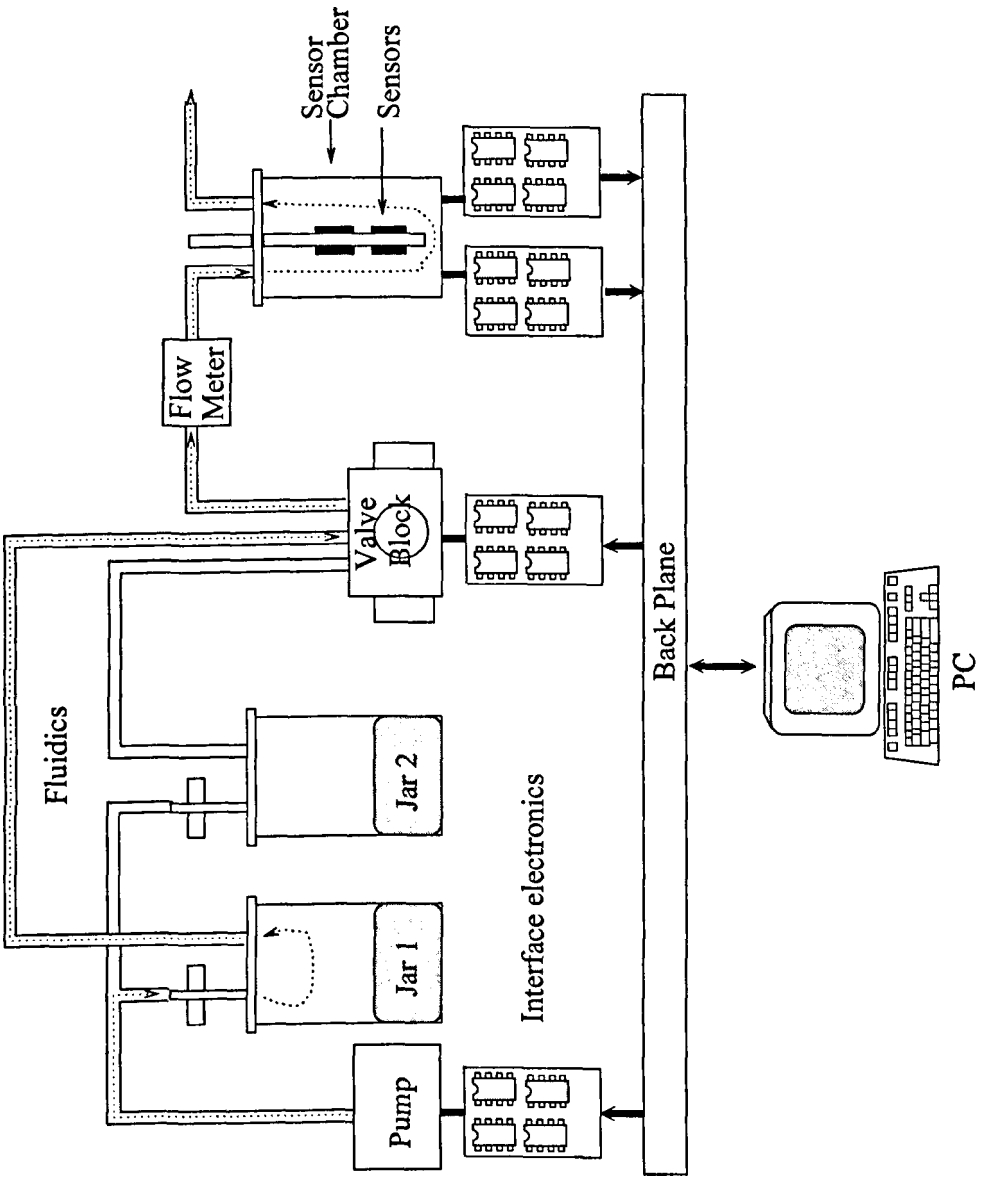


Figure 9.4: Schematic of the relationship between modules of the experimental system.

measure air temperature near the jars. Two solvents were proposed: Ethanol and Acetone. These are known to interact very differently with the polymers discussed below in Section 9.5.5 and so some variation should be observed in the estimated parameter values from sensor responses to these two substances.

The requirement of a constant vapour concentration in the neighbourhood of each sensor dictates that a good headspace reserve is required. Thus 120ml jars were used as sample vessels. These produce around 100ml of headspace.

9.5.2 Fluidics: Odour Transport

The sample vessels were connected to the sensor chamber via 3mm diameter piping. Sample selection is facilitated by solenoid valves mounted in a PTFE block. A small scale diaphragm pump (NP 904 from KNF UK) was used to create fluid flow. This pump was oiless in order to avoid contamination of the sample of the headspace.

There are two things to consider for the transport of the sample vapour:

- Flow rate
- Sampling Cycle

Taken in the context of a requirement of an instantaneous interface concentration the following was suggested. The sensor chamber is of a much smaller volume than the sample vessels. This will ensure a constant sample concentration at the sensor surface.

The following sampling cycle was chosen:

1. Sample vapour in jar 1 (10 seconds).
2. Flush sensors with air (600 seconds).
3. Sample vapour in jar 2 (10 seconds).
4. Flush sensors with air (600 seconds).

9.5.3 Solenoid Valves

The solenoid block consists of a PTFE block, drilled to accept three valves. This block has three inlets, two were connected to the outlets of the sample jars and one took in air, as well as one outlet, which passed the selected sample on to the sensor chamber. The valves were controlled by the computer via an interface card.

9.5.4 Sensor Chamber

For the mounting of the sensors it was decided to use a new 'smart nose' design (Figure 9.5). This sensor chamber uses a narrow channel with the sensors set to be flush with the one side. The resulting flow, for velocities in the range used in the experiments described in this chapter, is laminar with a parabolic profile. The exact dimensions (see Figure 9.5) of the channel were arrived at after a number of prototypes were tested. The aim of this device is to produce a plug type flow in order to eliminate mixing dynamics.

9.5.5 Sensors

The rig, as designed, is capable of housing and performing measurements on six resistive sensors at a time. It is thus proposed to use three polymers in sets of two sensors. These polymers are: Poly(caprolactone); Poly(styrene-co-butadine); Poly(vinyl acetate). The reasons for using two of each polymer are the following:

- Built in redundancy means that the experiment does not have to stop if one of the polymer devices fails.
- The differences in parameter values within the group of the same type of sensors can be compared to see which are characteristic of the solvent-polymer reaction, and those which are dependent upon the precise geometry of the sensor. It is expected that due to fabrication techniques there will be some variation in the thickness of the polymer films.

The carbon black-polymer sensors used in this experiment are produced in-house by the Sensors Research Laboratory. Temperature and humidity sensors were mounted

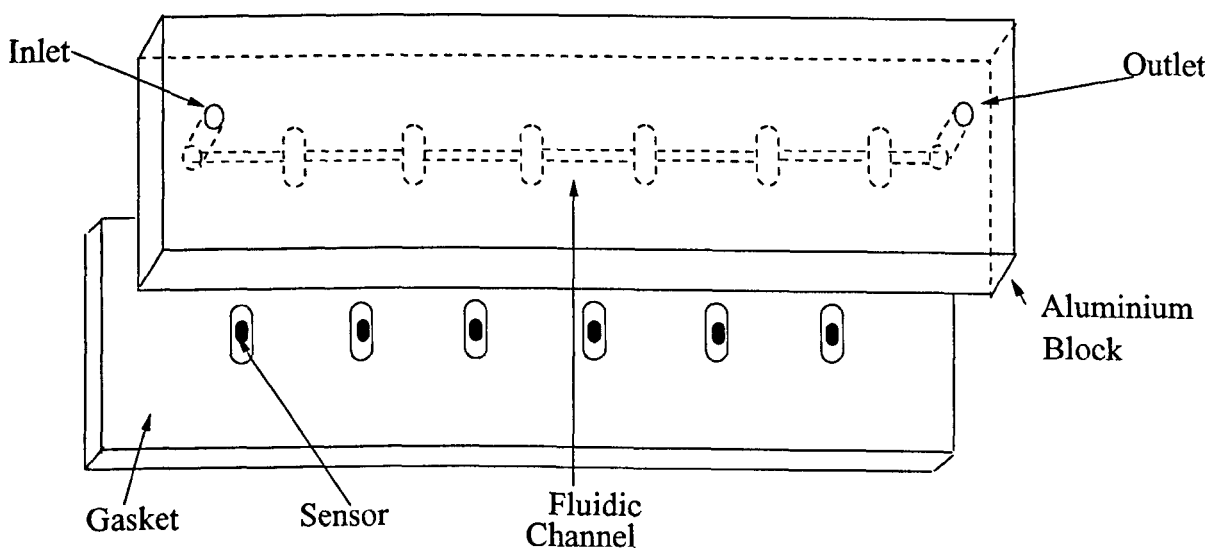


Figure 9.5: Schematic of the smart nose sensor chamber. The Aluminium block was 150mm long, 60mm wide and 15mm deep. The fluidic channel was 5mm in width and depth .

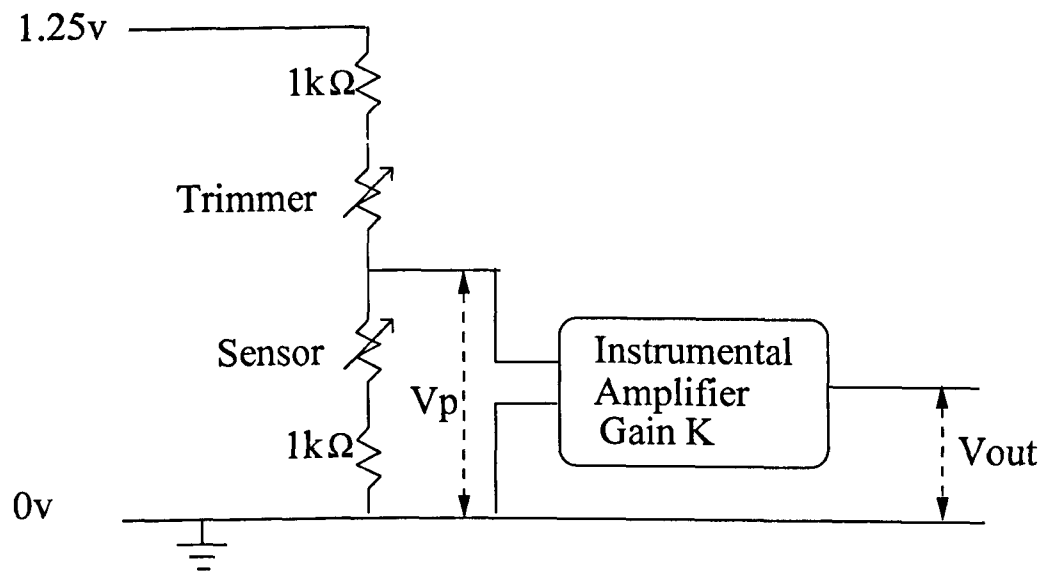


Figure 9.6: Schematic of sensor response measurement. V_{out} is measured by the test bed.

near the sensor chamber in order to measure environmental conditions. This enabled an investigation into the effect of these conditions on the parameters within the sensor response model.

The polymer sensors were monitored using a previously constructed interface card which uses potential divider and instrumentation amplifier to output a voltage proportional to the resistance of the sensor that it is monitoring (See Figure 9.6).

9.5.6 Interface Electronics

The interface electronics for the sensors comprise one card, capable of interfacing with six sensors. It can also interface with a temperature sensor to monitor the sensor chamber temperature.

These interface cards have to be balanced to monitor the sensors produced. During manufacture the sensors were produced to have a baseline resistance of approximately 1 kΩ. Trimming variable resistors are integrated within the cards to balance the potential divider.

Given V_{out}, the trimmer resistance, R_{trim} and the gain, K, the current through the sensor may be recovered. First:

$$V_p = \frac{V_{out}}{K} \tag{9.8}$$

where K is the gain of the instrumentation amplifier. Considering the potential divider.

$$\begin{aligned} \frac{V_p}{1.25} &= \frac{R_{sens} + 1000}{R_{sens} + R_{trim} + 2000} \\ \text{giving } R_{sens} + 1000 &= \frac{V_p(R_{trim} + 1000)}{1.25 - V_p} \end{aligned} \tag{9.9}$$

where R_{sens} is the unknown sensor resistance. Finally:

$$\begin{aligned} i &= \frac{V_p}{R_{sens} + 1000} \\ &= \frac{1.25 - V_p}{R_{trim} + 1000} \\ &= \frac{1.25 - \frac{V_{out}}{K}}{R_{trim} + 1000} \end{aligned} \tag{9.10}$$

The temperature sensor in the chamber was a back up for an integrated humidity and temperature sensor. This device interfaced on a card that also interfaced with the flow meter. The flow rate was monitored through a meter placed in the fluidic system. Data for this did not flow into the PC data acquisition card, but into the serial port of the computer.

The solenoid valves used to control fluid flow were controlled from a separate card. The motor for the diaphragm pump was controlled from another card which will monitor the current drawn and the voltage across the pump, as a form of diagnostic.

9.5.7 Data Logging

The data were recorded at a sample rate of 20Hz and stored using a standard Windows PC. The responses to one acetone and one ethanol exposure were stored along with the environmental conditions in one file. Each file stored one experiment run each.

Phase	Time	Pump	Valve
Baseline	1800s	On	3
Sample	4s	On	1 or 2

Table 9.1: Valve timing information for experimental data gathering.

9.6 Proposed Experimental Regime

The mathematical model is based upon a linear PDE. The model requires that the boundary concentration is constant and instantaneous. It also implicitly assumes that the ambient humidity, temperature, and the temperature of the polymer body are all constant. This is implicit due to the hypothesised dependence of the parameter values upon these conditions. It is also desirable so that the dependence of the sensor response upon these environmental factors may be investigated via parameters estimated from experimental data.

9.7 Discussion

A test rig was designed to test the sensor response model produced in Chapter 8. It was based upon a standard modular electronic nose design. The constraints of the model were considered and thus an experimental regime was designed to satisfy these constraints.

Such issues as evaporation have been considered and their effects upon the theoretical sensor response have been analysed. Thus there are hypothetical relationships between environmental conditions and the model's free parameter values.

The system was designed to measure environmental variables as well as the sensor responses. In this way, not only may the model be validated by reconciling it with the experimental data, in addition the parameters may be validated by observing how the estimated values vary with temperature and humidity.

A representative plot of experimental data collected is shown in Figure 9.7. The two graphs show the data for an acetone and ethanol exposure, thus it is representative of one experimental cycle. The V_{out} data that were recorded by the test rig have

been converted to the current shown in the plots using Equation 9.10. The data are presented as change of current as the range of baseline currents is of an order of magnitude greater than the typical changes; a plot of actual current would leave the reader unable to appreciate the data properly.

The responses exhibit a time delay as discussed in Section 9.4. It may be observed that the response to ethanol is of an order of magnitude lower than the response to acetone. The time series exhibit some low level noise and fluctuations which make the response non-monotonic. The frequency of these fluctuations is too low to be electrical and could possibly be due to slight variations in the flow of the carrier gas. Notice the correlation of the 1-2Hz oscillations between time series.

Once the data have been gathered then the parameters must be estimated. In the next chapter methods for parameter estimation will be explored and the results analysed.

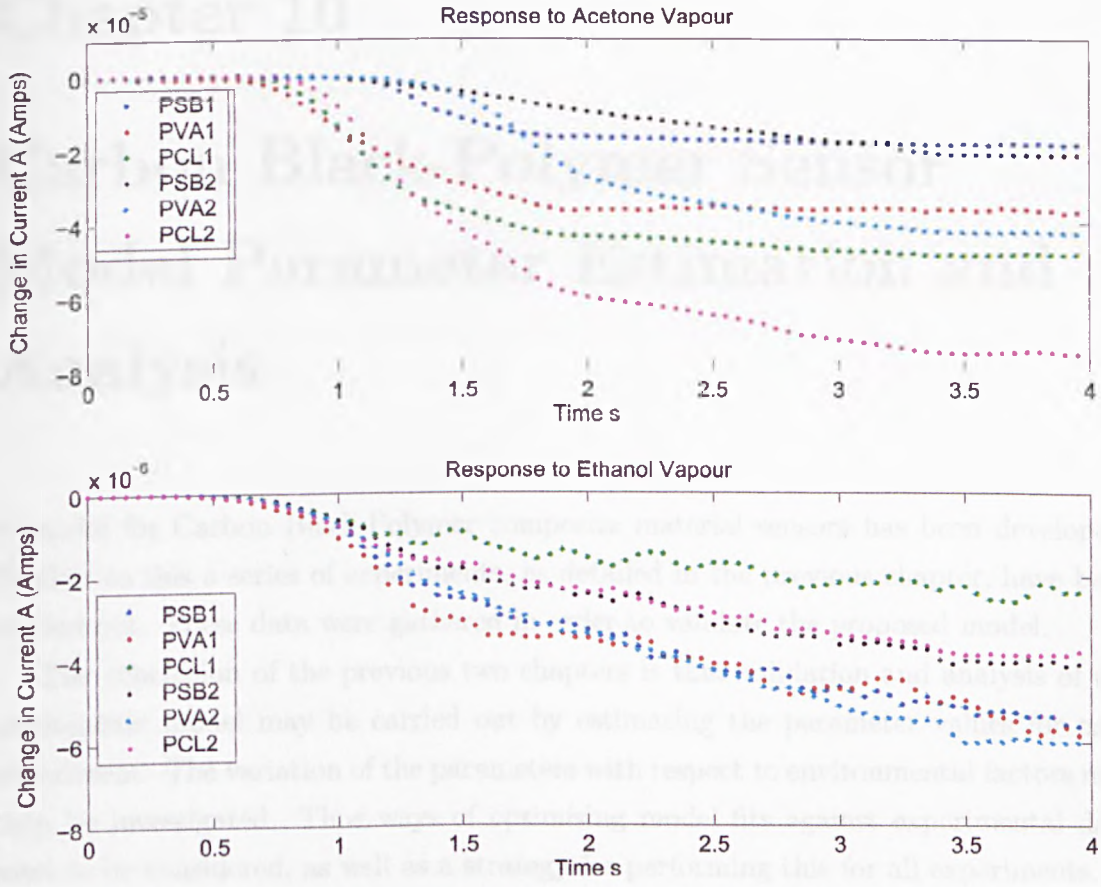


Figure 9.7: Sample time series recorded for an array of sensors for one experimental run.

Chapter 10

Carbon Black-Polymer Sensor Model Parameter Estimation and Analysis

A model for Carbon Black-Polymer composite material sensors has been developed. Further to this a series of experiments, as detailed in the previous chapter, have been carried out. These data were gathered in order to validate the proposed model.

The conclusion of the previous two chapters is that validation and analysis of the mechanistic model may be carried out by estimating the parameter values for each experiment. The variation of the parameters with respect to environmental factors may then be investigated. Thus ways of optimising model fits against experimental data need to be considered, as well as a strategy for performing this for all experiments.

The model produced gives algebraic expressions for one sensor output (8.70). The full experimental system can be simulated by considering six models of the same form, but not necessarily with the same parameter values, in parallel. The expression (8.70) is highly non-linear in time as well as in the parameters. Rough 'ball park' estimates of the parameter values were not initially known and these were needed to start optimisation based upon each experiment. These initial values were sought first and then a batch process was implemented to consider all experimental data collected simultaneously.

The polymers used in the sensor will be abbreviated to PCL for Poly(caprolactone);

PSB for Poly(styrene-co-butadine); PVA Poly(vinyl acetate)

10.1 Parameter Estimation

Given the model structure, as developed in Chapter 8, it is necessary to identify the model most closely characterising the system's behaviour. The experiments involved a series of stimulations (solvent introduction) and from this a candidate model may be chosen to represent the system's response (system identification). It should be noted that input to the system is only partially controlled, as shown in Table 10.1. However, the uncontrolled conditions have been recorded. As discussed in the last chapter, knowledge of environmental factors will help to characterise the system; the model identified will only be valid for the environment present at the time. This will enable a characterisation of the environmental dependencies of the system via the estimated parameter values.

The model chosen gave predicted outputs for the six sensors in the experiment simultaneously. The specific model identified for each experiment was chosen with respect to certain criteria. These may be dichotomised into parameter values and output error. For computer simulation the single sensor response model is parameterised as follows:

In (8.70) it can be seen that a constant voltage V across the sensor is assumed. However (9.8) shows this to be untrue. Hence $V = 1$ is set in the model and so the result of applying (9.10) to the experimental data was divided through by $V_p - 1000i$ to give an idealised response for unit voltage, where i is the current in (8.70). Also set

$$X_1(t) = k_1 t^{\frac{1}{2}} \tag{10.1}$$

$$X_p(t) = k_p t^{\frac{1}{2}} \tag{10.2}$$

Input	Controlled?
Solvent	Yes
Flow rate	Partially
Temperature	No
Humidity	No

Table 10.1: Inputs to the gas sensor system

then from (8.70):

$$i = \frac{\sigma_0}{\pi}\Bigg[N(c_0Q - C_X)\frac{1}{(k_1 - k_p)t^{\frac{1}{2}}}\Big[k_p t^{\frac{1}{2}}\Big[\cosh^{-1}(k_1 t^{\frac{1}{2}} + y_0) - \cosh^{-1}(k_p t^{\frac{1}{2}} + y_0)\Big] - \Big[\exp(\cosh^{-1}(k_1 t^{\frac{1}{2}} + y_0)) - \exp(\cosh^{-1}(k_p t^{\frac{1}{2}} + y_0)) - y_0\Big[\cosh^{-1}(k_1 t^{\frac{1}{2}} + y_0) - \cosh^{-1}(k_p t^{\frac{1}{2}} + y_0)\Big]\Big]\Big] - \log 2y_0 - NC_X \cosh^{-1}(k_p t^{\frac{1}{2}} + y_0) + 1\Bigg] \quad (10.3)$$

It is obvious from the physical interpretation of each of the parameter values that all should be positive except for k_p. Secondly, as pairs of sensors share the same chemical characteristics, they should also share all parameters representing these characteristics. This is easily implemented within the computer simulation code. In order to keep the number of free parameters to a minimum, the assignment $C_X = \frac{1}{2}Qc_0$ was made. The constant $\frac{1}{2}$ was arrived at after trial and error model fits.

A pure time delay is incorporated into the response of each sensor to allow for the delay inherent in the fluidics.

Output error is simply a measure of how well the model prediction corresponds to the system output. For the case in hand it is desirable for the selected model to be a good predictor of the system output. For this reason the sum of the squared error is imposed as an error function. The model selected then will minimise the squared error against the parameter constraints discussed above.

Optimisation techniques are required to tune the parameter values in the model to fit the observed behaviour (Figure 10.1). There are a number of different methods for numerical optimisation. They are all iterative in nature, as discussed in Chapter 4.

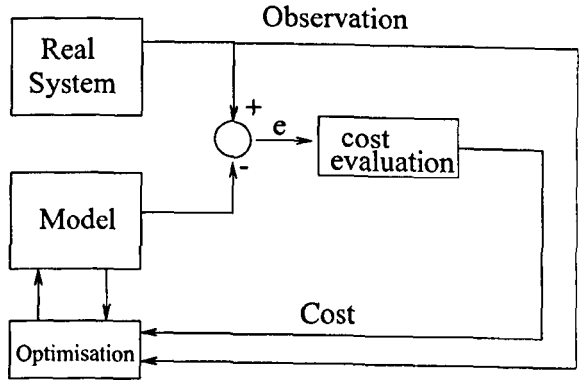


Figure 10.1: Schematic of the optimisation process inherent when estimating parameter values for an observed experimental response.

These methods all require a 'starting' parameter vector on which to begin the iterative process. A good start value was not known so this had to be identified. A good method for this was the genetic algorithm approach, since this algorithm can simultaneously search different areas of the solution space. For final parameter estimation it was proposed to use the Levenberg-Marquardt method. This method was chosen as it is successful at handling nonlinear expressions [105].

10.2 Batch Processing of Experimental Data

The outline of the technique to process the experimental data was as follows:

1. Sort files into the order that the experiments were performed. This was done in order to detect any sensor drift.

2. The earliest file is processed using the best guess found using the genetic algorithm as start values for the Levenberg-Marquardt algorithm.

3. The resulting estimated parameters are recorded in a table, along with the recorded environmental data.

4. The estimated parameter results were then used for the second experiment's start

values; this process of using the final values of the last experiment as the start values for the next was continued until the model had been fitted to all the data.

This strategy was decided upon as it represented an efficient way to process the data. There might be concern about propagation of error because of point 4 above, however the results demonstrate that this is not the case. It was assumed that the environment in the laboratory would vary continuously with time and that the model parameters would vary continuously with the environment. Hence it was assumed that the parameters would vary continuously with time; using the results of the last experiment parameter values supplied a best start parameter vector for the next experiment. In this way it was hoped to minimise the number of iterations of the Levenberg-Marquardt algorithm necessary to complete the task.

10.3 Computer Implementation

Model optimisation was carried out within the MATLAB V5.3 environment. Code was adapted from that used previously in Chapter 5 to create a genetic algorithm to search for candidate optimal models. The Levenberg-Marquardt algorithm has already been implemented in MATLAB as the LSQNONLIN routine. This routine requires the start values of the parameters, the function to minimise and stopping criteria as input arguments. It outputs the fitted parameters, the fitting error and whether successful convergence has occurred.

Common to both of these optimisation stages was the simulation script which produced, based upon specific parameters values, a simulation of two sets of six sensors representing the six experimental sensors reacting to the two solvents. The two sets of six responses for each experiment are interrelated. These relationships are that pairs of sensors share the same chemical characteristics and that between two exposures a sensor retains the same physical characteristics. These relationships were manifested in the simulation by common parameters representing common characteristics.

The script then compared the simulation with the provided single set of experimental data and calculated the squared error. The squared error was weighted using

the range of each time series in order to obtain a uniform fit over the various sensor responses. If w_i represents the range of sensor response i, $y_{i,\text{real}}(t)$, then:

$$w_i = max(y_{i,\text{real}}(t)) - min(y_{i,\text{real}}(t)) \tag{10.4}$$

and so the error for a given parameter vector $\hat{\mathbf{p}}$ is

$$E(\hat{\mathbf{p}}) = \sum_{i \in \text{Sensors}} \sum_{t \in time} \frac{(y(t)_{i,\text{real}} - y_{i,\hat{\mathbf{p}}}(t))^2}{w_i} \tag{10.5}$$

where $y_{i,\hat{\mathbf{p}}}(t)$ is the model simulation for parameter vector $\hat{\mathbf{p}}$.

To identify a good 'start' value for the Levenberg-Marquardt algorithm the genetic algorithm was run with 100 individuals in the population for 1000 generations. Absolute convergence was not sought, rather, variation in the neighbourhood of solutions was preferable in order to fully explore possible start values for the Levenberg-Marquardt optimisation. The algorithm output all 100 individuals from the final population. Each of these was investigated using the Levenberg-Marquardt algorithm. The optimal result was used as the start value for the main optimisation algorithm.

A batch processing algorithm was written to execute as described above in Section 10.2. A limit on the number of iterations of the Levenberg-Marquardt algorithm was imposed for two reasons. The first is that on a few occasions, due to experiments being unsupervised, fluctuations in the equipment produced non-typical responses. These included the pump becoming overheated and the solvent vessels running empty. It was difficult to identify these amongst the hundreds of experiments carried out. Instead the final error of each fit was recorded along with the other data for the experiment and suspect values were identified before final analysis.

10.4 The Identified Model

The model fitted to data with a good degree of accuracy. Figure 10.2 shows an example of the model fit compared to the real data, the change in current is plotted so that all the responses may be viewed on the same scale. The experimental and simulated data are plotted in the same colour for each sensor. This fit has a weighted sum of squared

errors (10.5) of 1.944. The error was measured with respect to magnitude of response and so an average point wise error of 4.5% was estimated. This is considered to be very good given that:

- The constraints on the parameters and the model to fit all 12 responses simultaneously.
- The difference in magnitude of response to acetone and ethanol.
- The data were noisy.
- The experiment was 'uncontrolled'.

Table 10.2 details the estimated parameter values for this fit, as well as the standard deviations (SD). It can be seen that the parameters do give an insight into the speed and magnitude of the sensors' responses to the two solvents. PCL, it appears produces the fastest response to acetone, whereas PVA reacts quickest to ethanol. For magnitude of response it appears that PSB has the greatest (judging by values of parameter N). Notice that the all the parameters have standard deviations of an order of magnitude lower than their respective values, thus giving a high confidence level in the values estimated

Table 10.3 details the estimated correlation matrix for the parameters for the PCL sensor pair for both acetone and ethanol exposures. Notice the correlation is much higher between corresponding terms of the product NQc_0 than for other pairs of parameters. This correlation is to be expected as only the product may be estimated from the model fits.

The parameter standard deviations and correlation coefficients were calculated using the Hessian matrix method detailed in [172]. If $H(\hat{\mathbf{p}})$ is the Hessian matrix of the model error $E(\hat{\mathbf{p}})$ (10.5) with respect to the parameters at $\hat{\mathbf{p}}$, then the parameter covariance matrix, $Cov(\hat{\mathbf{p}})$, may be written as

$$Cov(\hat{\mathbf{p}}) = \frac{2E(\hat{\mathbf{p}})}{M - n_p} H(\hat{\mathbf{p}})^{-1}. \tag{10.6}$$

Here $M = 960$ experimental data points and $n_p = 50$ free parameters.

Parameter	Units	Value	SD	Parameter	Units	Value	SD
kp_{11}	$ms^{-\frac{1}{2}}$	2.5×10^{-3}	3.2×10^{-5}	$k1_{11}$	$ms^{-\frac{1}{2}}$	2.1×10^{-4}	2.7×10^{-13}
kp_{21}	$ms^{-\frac{1}{2}}$	2.1×10^{-3}	2.4×10^{-10}	$k1_{21}$	$ms^{-\frac{1}{2}}$	8.0×10^{-5}	4.5×10^{-23}
kp_{12}	$ms^{-\frac{1}{2}}$	4.3×10^{-4}	1.4×10^{-7}	$k1_{12}$	$ms^{-\frac{1}{2}}$	3.1×10^{-4}	3.8×10^{-12}
kp_{22}	$ms^{-\frac{1}{2}}$	1.2×10^{-4}	2.7×10^{-8}	$k1_{22}$	$ms^{-\frac{1}{2}}$	2.5×10^{-4}	2.3×10^{-15}
kp_{13}	$ms^{-\frac{1}{2}}$	3.0×10^{-3}	3.2×10^{-5}	$k1_{13}$	$ms^{-\frac{1}{2}}$	1.2×10^{-4}	1.0×10^{-28}
kp_{23}	$ms^{-\frac{1}{2}}$	3.1×10^{-2}	3.8×10^{-9}	$k1_{23}$	$ms^{-\frac{1}{2}}$	7.0×10^{-3}	1.2×10^{-19}
N_{11}	M^{-1}	2.7×10^{0}	1.2×10^{-4}	Q_{11}	none	4.7×10^{0}	4.1×10^{-8}
N_{21}	M^{-1}	7.5×10^{-1}	2.0×10^{-5}	Q_{21}	none	2.3×10^{-1}	1.8×10^{-6}
N_{12}	M^{-1}	1.5×10^{0}	3.7×10^{-6}	Q_{12}	none	3.6×10^{0}	3.4×10^{-6}
N_{22}	M^{-1}	1.47×10^{0}	1.1×10^{-4}	Q_{22}	none	5.7×10^{-1}	3.5×10^{-6}
N_{13}	M^{-1}	4.7×10^{-3}	2.1×10^{-5}	Q_{13}	none	6.9×10^{-4}	1.3×10^{-9}
N_{23}	M^{-1}	5.4×10^{-1}	2.5×10^{-6}	Q_{23}	none	1.0×10^{-1}	1.1×10^{-7}
y_1	m	1.4×10^{-4}	1.7×10^{-5}	σ_1	$AV^{-1}m^{-1}$	1.4×10^{0}	2.2×10^{-16}
y_2	m	1.0×10^{-5}	7.1×10^{-5}	σ_2	$AV^{-1}m^{-1}$	3.1×10^{-2}	9.1×10^{-15}
y_3	m	2.1×10^{-4}	5.7×10^{-6}	σ_3	$AV^{-1}m^{-1}$	2.3×10^{0}	5.5×10^{-14}
y_4	m	1.4×10^{-4}	2.5×10^{-8}	σ_4	$AV^{-1}m^{-1}$	1.4×10^{0}	3.6×10^{-16}
y_5	m	1.5×10^{-4}	7.3×10^{-5}	σ_5	$AV^{-1}m^{-1}$	1.5×10^{0}	1.2×10^{-14}
y_6	m	2.5×10^{-4}	7.1×10^{-6}	σ_6	$AV^{-1}m^{-1}$	2.5×10^{0}	1.6×10^{-13}
c_1	M	1.1×10^{-1}	1.6×10^{-7}	c_2	M	1.2×10^{-1}	9.3×10^{-6}

Table 10.2: Estimated parameter values, where SD stands for standard deviation. Parameters are labelled $P_{\text{solvent,sensor}}$ or P_{sensor}. For sensors; 1=PCL, 2=PSB, 3=PVA. For solvents; 1=Acetone, 2=Ethanol. Parameters are as in equation 10.3. For example c_1 represents the concentration of acetone

	N_{11}	N_{21}	Q_{11}	Q_{21}	c_1	c_2
N_{11}	1.0×10^{0}	1.3×10^{-01}	-7.2×10^{-01}	2.7×10^{-01}	9.7×10^{-01}	1.4×10^{-01}
N_{21}	1.3×10^{-1}	1.0×10^{0}	2.2×10^{-1}	-2.5×10^{-1}	4.9×10^{-3}	1.0×10^{0}
Q_{11}	-7.2×10^{-1}	2.2×10^{-1}	1.0×10^{0}	2.7×10^{-1}	-7.4×10^{-1}	2.2×10^{-1}
Q_{21}	2.7×10^{-1}	-2.5×10^{-1}	2.7×10^{-1}	1.0×10^{0}	3.0×10^{-1}	-2.5×10^{-1}
c_1	9.7×10^{-1}	4.9×10^{-3}	-7.4×10^{-1}	3.0×10^{-1}	1.0×10^{0}	-1.3×10^{-3}
c_2	1.4×10^{-1}	1.0×10^{0}	2.2×10^{-1}	-2.5×10^{-1}	-1.3×10^{-3}	1.0×10^{0}

Table 10.3: Correlation matrix for a selection of parameters for one sensor type. Parameters are as in Equation (10.3)

The Hessian matrix of the error was calculated using a finite differences method with perturbations in the parameters being analysed to avoid numerical errors [172].

Figure 10.3 shows a histogram of the errors over all experiments. It can be seen that a good fit was achieved for the majority of the experimental data.

10.5 Analysis of Parameter Results

The resulting estimated parameter values will now be analysed with respect to the changing environmental conditions. The aims of the analysis were to understand how the dynamics of the system change with the environmental conditions.

10.5.1 Flow Rate and Time Delay

The simplest relationship that exists within the system is the inverse relationship between the model pure time delay and the flow rate. Figure 10.4 shows this relationship.

There is reasonable correlation. Error will originate from the sampling rate of the test rig, as well as the fact that the progress of the solvent vapour is observed indirectly via the sensor response. These results demonstrate that the parameter identification process was performed robustly and real variations in the system have been detected.

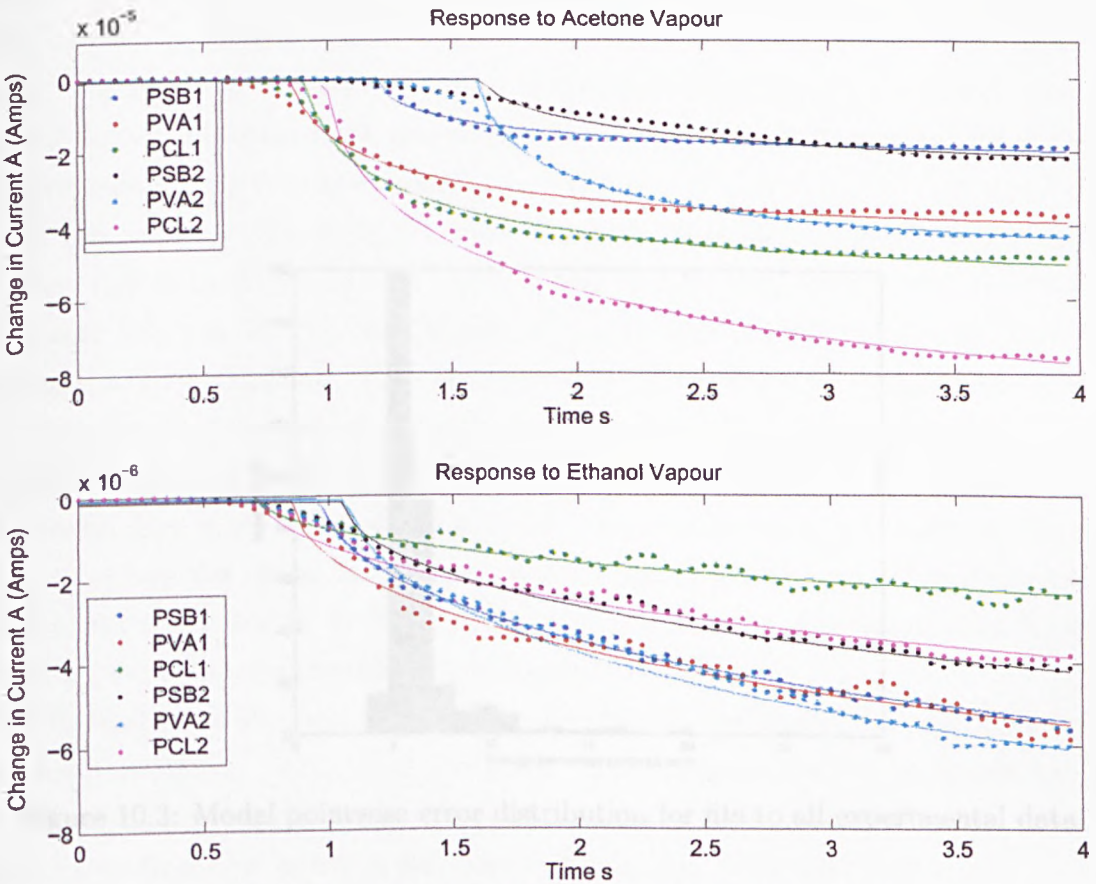


Figure 10.2: Example of the a model fitted to an experimental dataset.

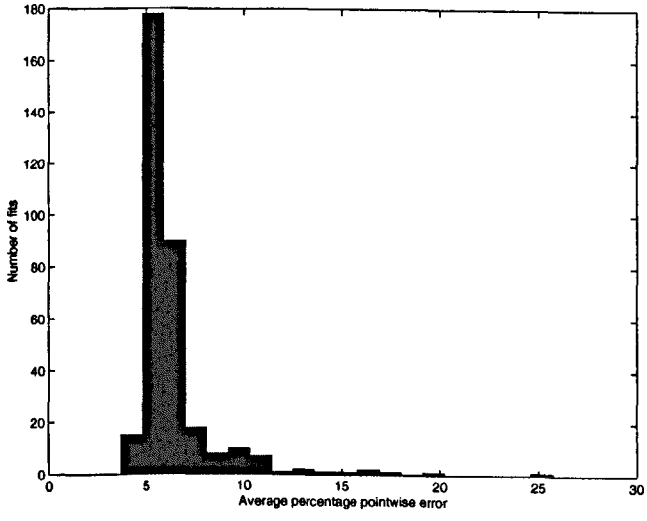


Figure 10.3: Model pointwise error distribution for fits to all experimental data.

10.5.2 Rate Constants

The rate constants in the model relate to the progress of the solvent diffusion front through the polymer films. It would be expected that there would be a strong environmental relationship due to mass action and the fact that temperature dependence of sensor responses has been observed before [57]. It is these environmental dependencies that need to be understood in order to widen the range of electronic nose applications. Figures 10.5, 10.6 and 10.7 show the results. These graphs are scatter plots of k_p, the solvent diffusion front rate constant, as defined in (10.2) against the ambient temperature during the experiment.

It can be seen that there is a much stronger temperature relationship with the acetone responses (Figures 10.5(a), 10.6(a) and 10.7(a)) than there is with the ethanol responses (Figures 10.5(b), 10.6(b) and 10.7(b)). This is probably due to the much stronger acetone responses. The dependences are either linear or slightly sigmoidal.

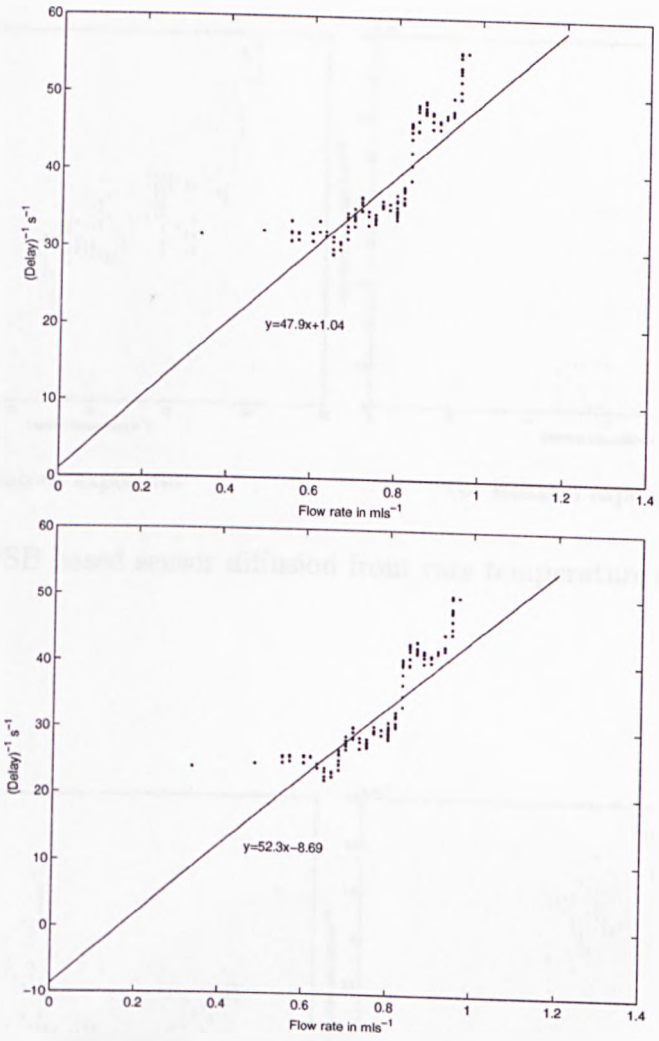
The acetone responses are also affected by the humidity of the atmosphere. This is shown in Figures 10.8, 10.9 and 10.10. The rate of reaction reduces with humidity. Note the discrepancy in the trend of Figure 10.10. This may be due to other factors effecting the rate of response or the sensor characteristics changing with time. The temperature effects were accounted for by fitting a 4th order polynomial to the temperature/k_p data in order to incorporate sigmoidal type dependency.

The final effect observed is that of the flow rate dependence of the rate of diffusion, as shown in Figures 10.11, 10.12 and 10.13. There again appear to be several modes of behaviour, however there are very definite relationships between the rate of diffusion and the environment in which the electronic nose sits. The behaviour is very similar for the three different types of sensors.

10.5.3 Swelling, Solvent Concentration and Partition coefficient

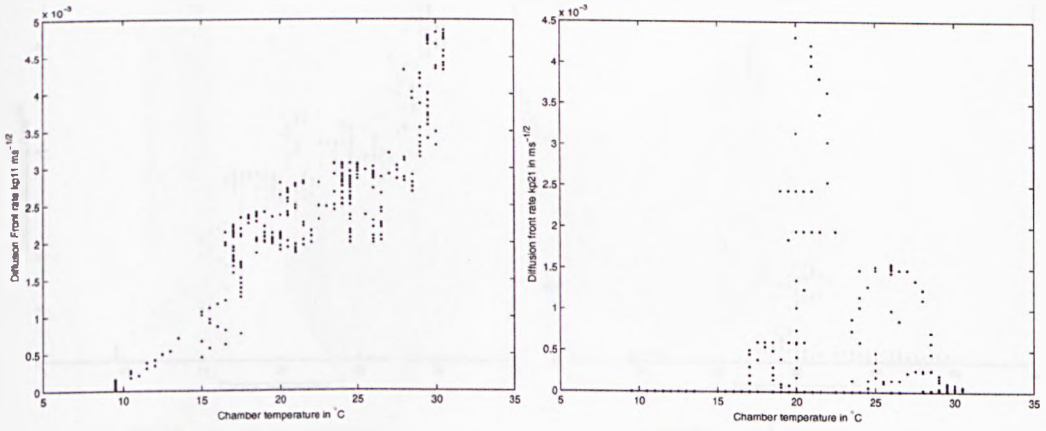
In the proposed model, only the product NQc_0 is estimable. This is proportional to the total change in current, as can be seen by inspecting (10.3).

Figures 10.14, 10.15 and 10.16 show the dependence of the product NQc_0 on temper-



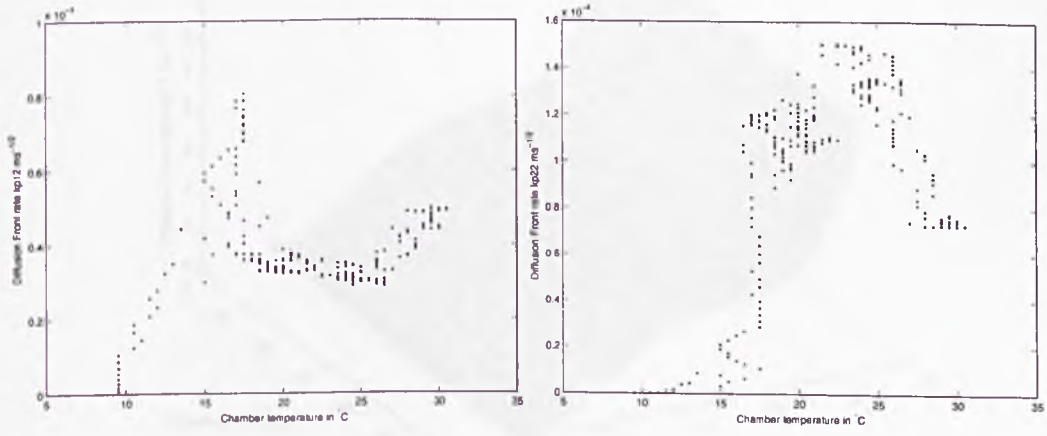
(b) Sensor 6

Figure 10.4: Relationship between time delay and flow rate. The expression relating the flow rate and the inverse of the time delay is written in the form $y = mx + c$ as delay is inversely proportional to the flow rate.



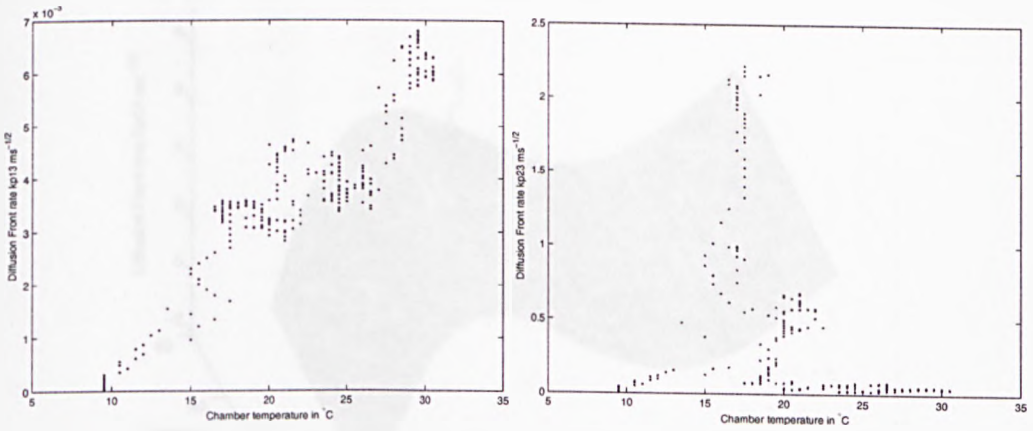
(a) Acetone exposures (b) Ethanol exposures

Figure 10.5: PSB based sensor diffusion front rate temperature dependence.



(a) Acetone exposures (b) Ethanol exposures

Figure 10.6: PVA based sensor diffusion front rate temperature dependence.



(a) Acetone exposures (b) Ethanol exposures

Figure 10.7: PCL based sensor diffusion front rate temperature dependence.

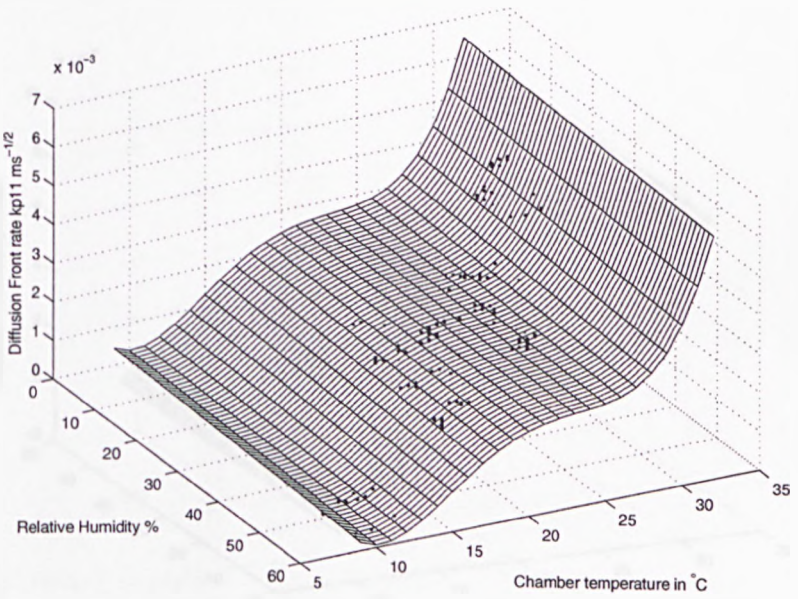


Figure 10.8: PSB based sensor diffusion front rate temperature and humidity dependence.

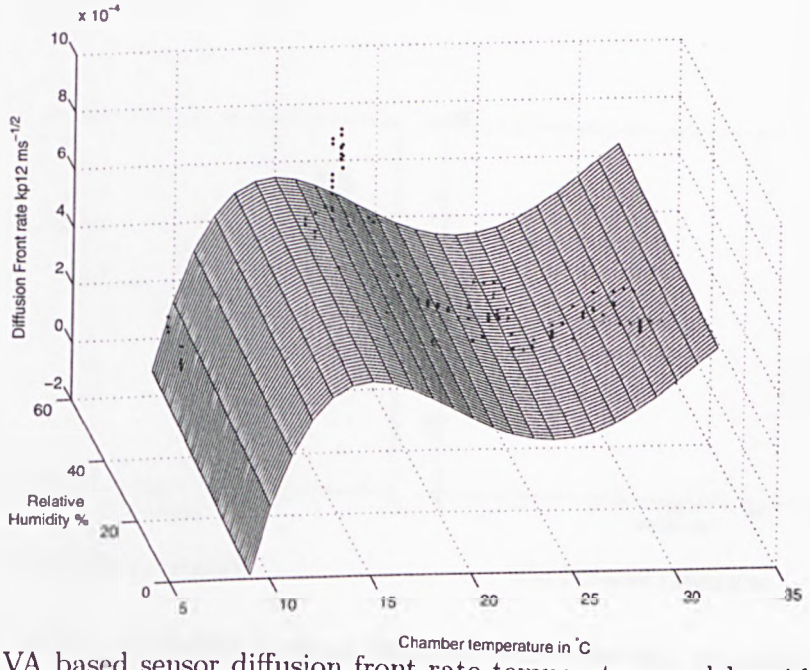


Figure 10.9: PVA based sensor diffusion front rate temperature and humidity dependence.

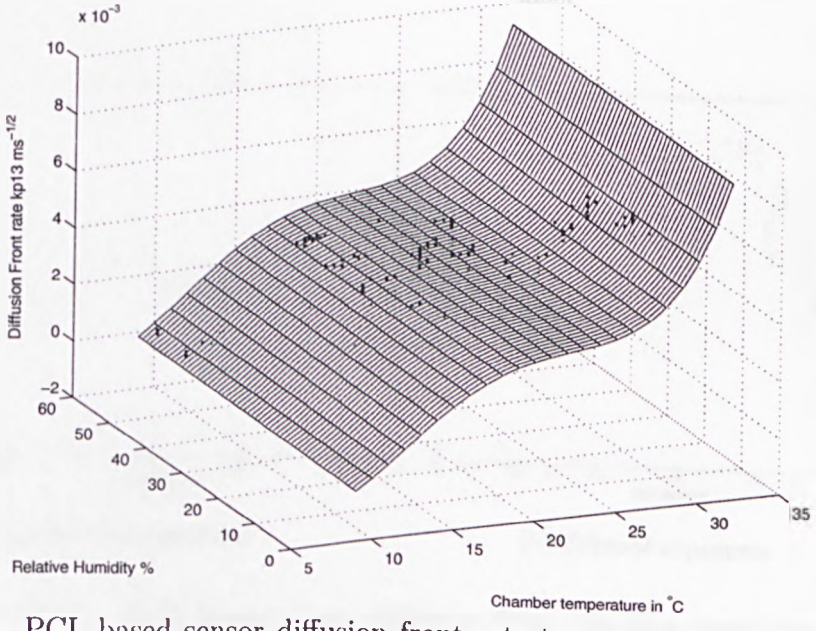
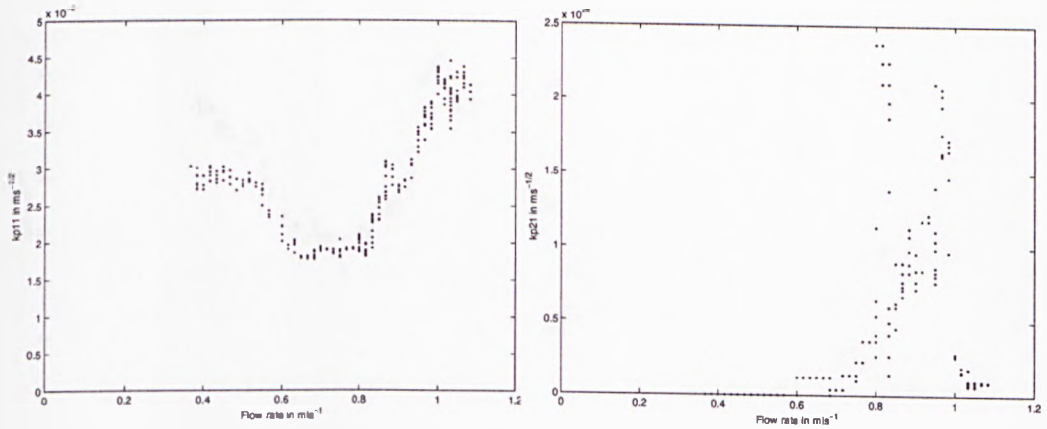
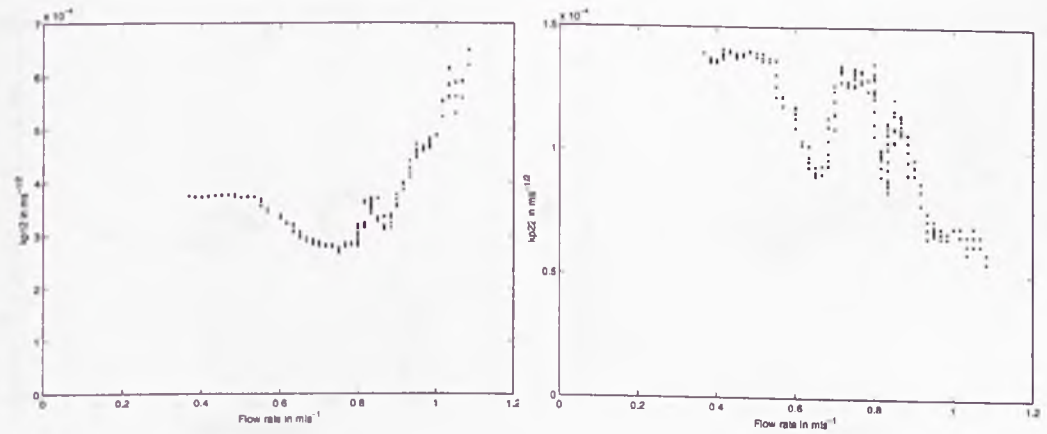


Figure 10.10: PCL based sensor diffusion front rate temperature and humidity dependence.



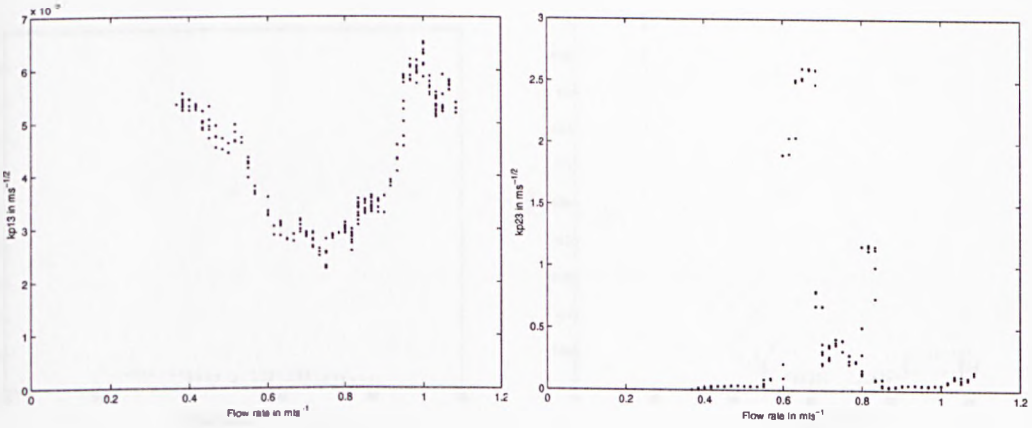
(a) Acetone exposures (b) Ethanol exposures

Figure 10.11: PSB based sensor diffusion front rate flow dependence.



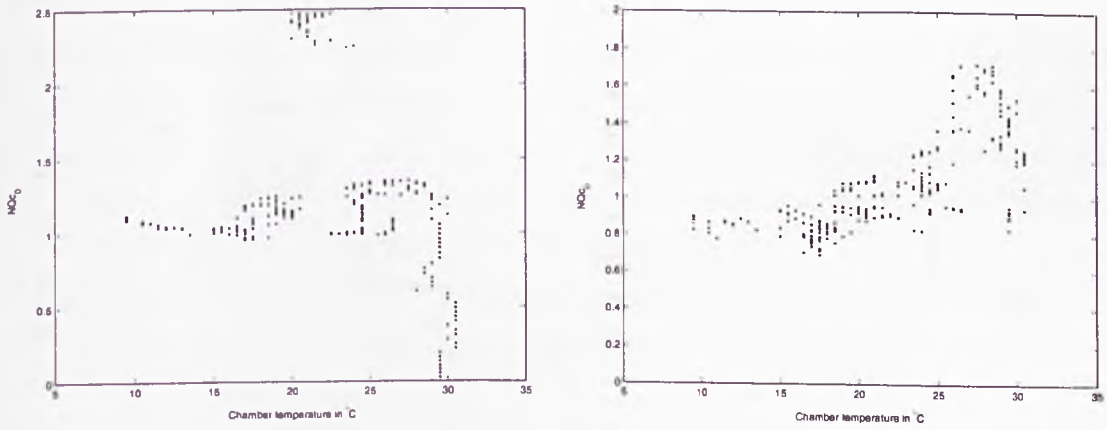
(a) Acetone exposures (b) Ethanol exposures

Figure 10.12: PVA based sensor diffusion front rate flow dependence.



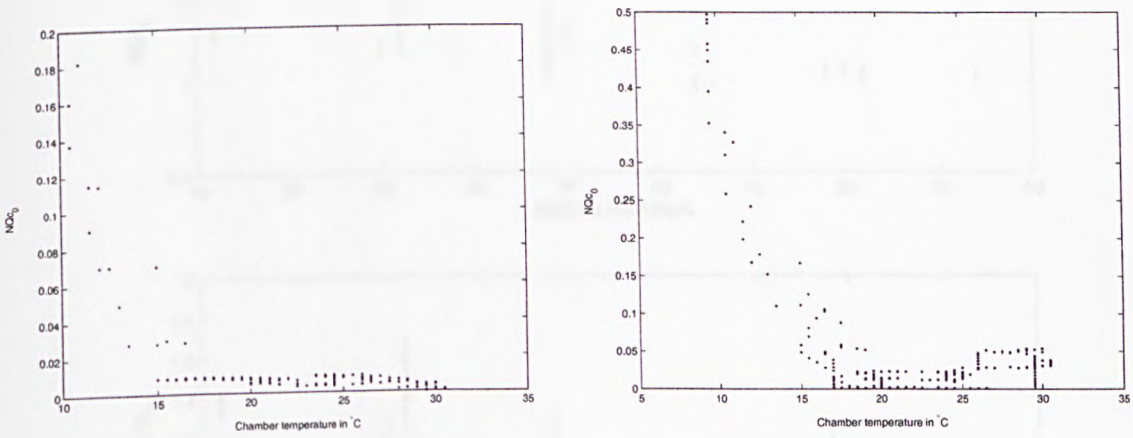
(a) Acetone exposures (b) Ethanol exposures

Figure 10.13: PCL based sensor diffusion front rate flow dependence.



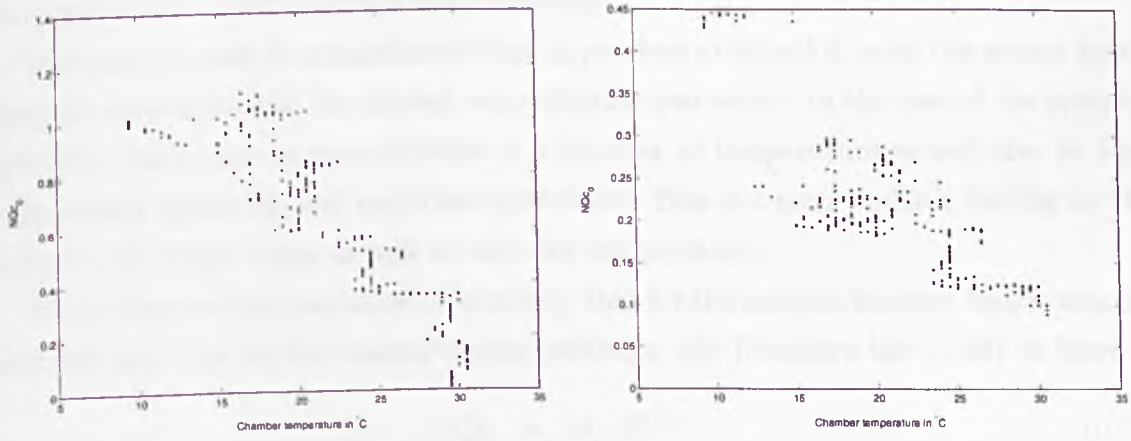
(a) Acetone exposures (b) Ethanol exposures

Figure 10.14: PSB based sensor temperature dependence of NQc_0.



(a) Acetone exposures (b) Ethanol exposures

Figure 10.15: PVA based sensor temperature dependence of NQc_0.



(a) Acetone exposures (b) Ethanol exposures

Figure 10.16: PCL based sensor temperature dependence of NQc_0.

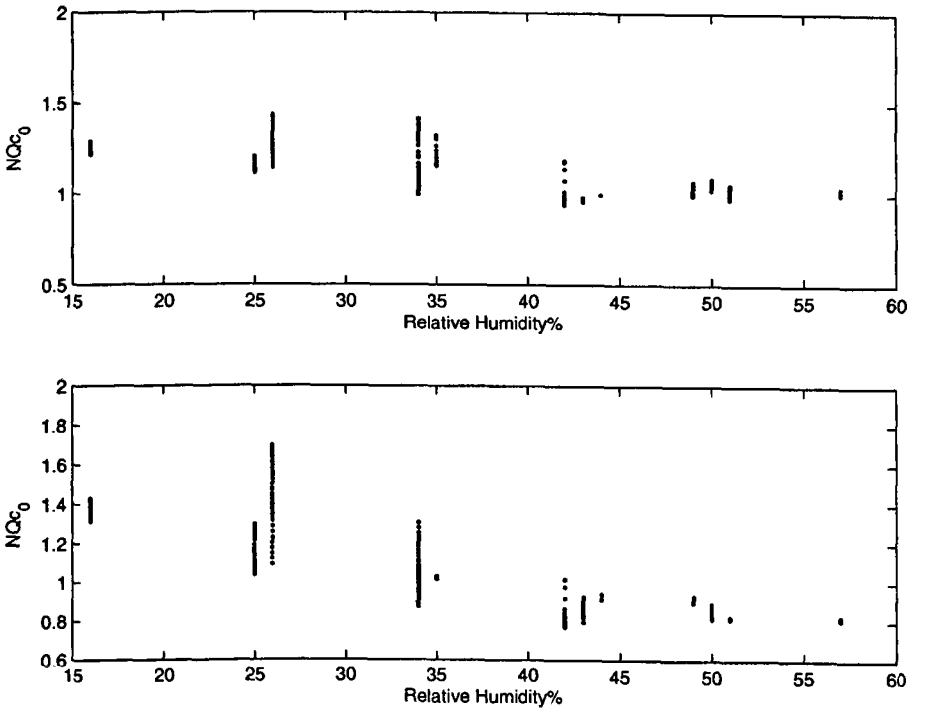


Figure 10.17: PSB based sensor pair humidity dependence of NQc_0.

ature. It can be seen that there is a consistent growth or reduction with temperature. Previously only a reduction has been observed.

However, it must be remembered that in previous studies[57], only the sensor temperature was varied and the solvent concentration was static. In the case of the system considered here, solvent concentration is a function of temperature as well, due to the temperature dependence of partition coefficients. This is a more realistic finding as 'in the field' electronic noses sample at ambient temperature.

Thus there are two partitions interacting: that for the sample chamber liquid/vapour partition and that for the sensor/vapour partition. By Trouton's law (8.78) we have

$$Q_L = A_L e^{B_L T} \tag{10.7}$$

$$Q_S = A_S e^{B_S T} \tag{10.8}$$

where Q_L, Q_S are the partition coefficients for the sample and sensor chambers respectively. Therefore $Q_S c_0$, the concentration at the atmospheric boundary is given

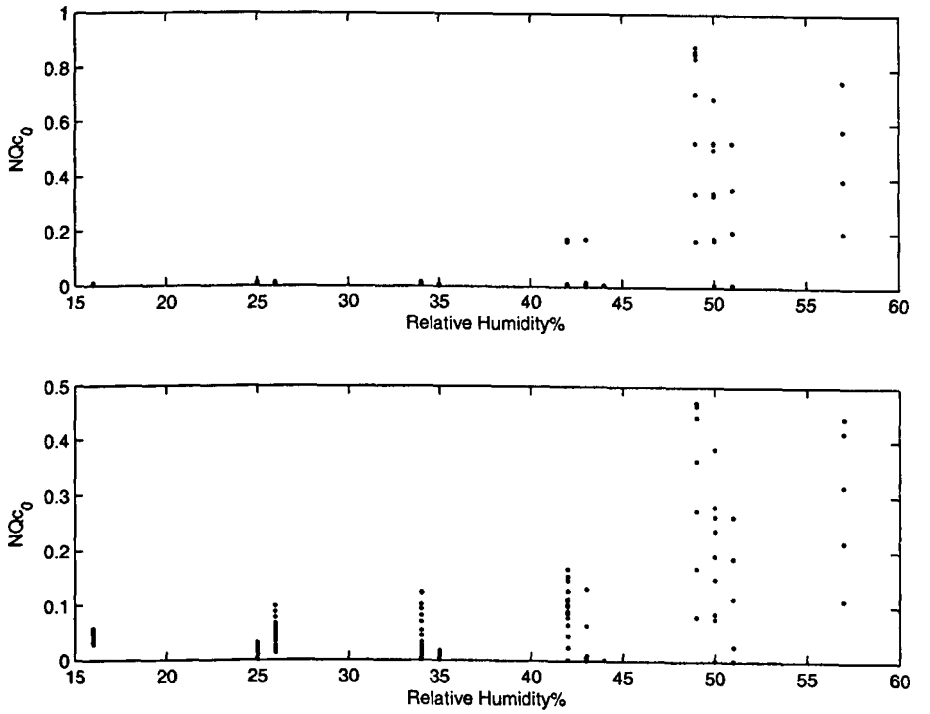


Figure 10.18: PVA based sensor pair humidity dependence of NQc_0.

by

$$Q_S c_0 = \frac{Q_S c_L}{Q_L} \tag{10.9}$$

$$= c_L \frac{A_S}{B_S} e^{(B_S - B_L)T} \tag{10.10}$$

where c_L is the sample chamber liquid concentration (which is assumed to be constant).

It can be seen that an increase or decrease with temperature of the boundary condition Qc_0 is dependent upon the relative sizes of B_S and B_L.

It cannot be ascertained whether the swelling coefficient alters with temperature or not.

Figures 10.17, 10.18 and 10.19 show the dependence of NQc_0 on humidity. Again a variety of different behaviours can be seen between the sensors. The decrease shown in Figure 10.17 seems to suggest that PSB is the most porous to water; the low NQc_0 means that the response of the sensor has been lowered due to water adsorption.

These results in particular demonstrate the complex interaction of an electronic

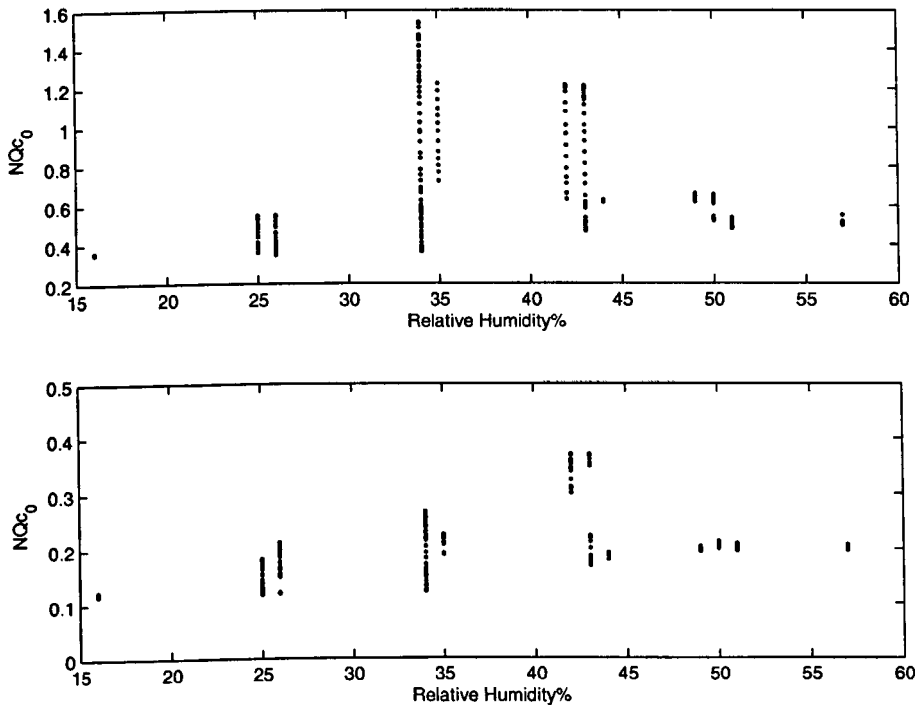


Figure 10.19: PCL based sensor pair humidity dependence of NQc_0.

nose with its environment.

10.5.4 Conductance

The temperature dependence of the conductance of composite sensors is well known and has already been discussed. The estimated parameter dependencies are shown in Figures 10.20, 10.21 and 10.22. Notice the consistency in the behaviour of the conductance for each pair of sensors.

The PSB sensors display the previously observed reduction with temperature of the baseline conductivity[135] as discussed in Chapter 8. The other two sensor types display a two stage behaviour. However for higher temperatures a reduction with temperature can be seen. This suggests that the sensors are expanding with temperature.

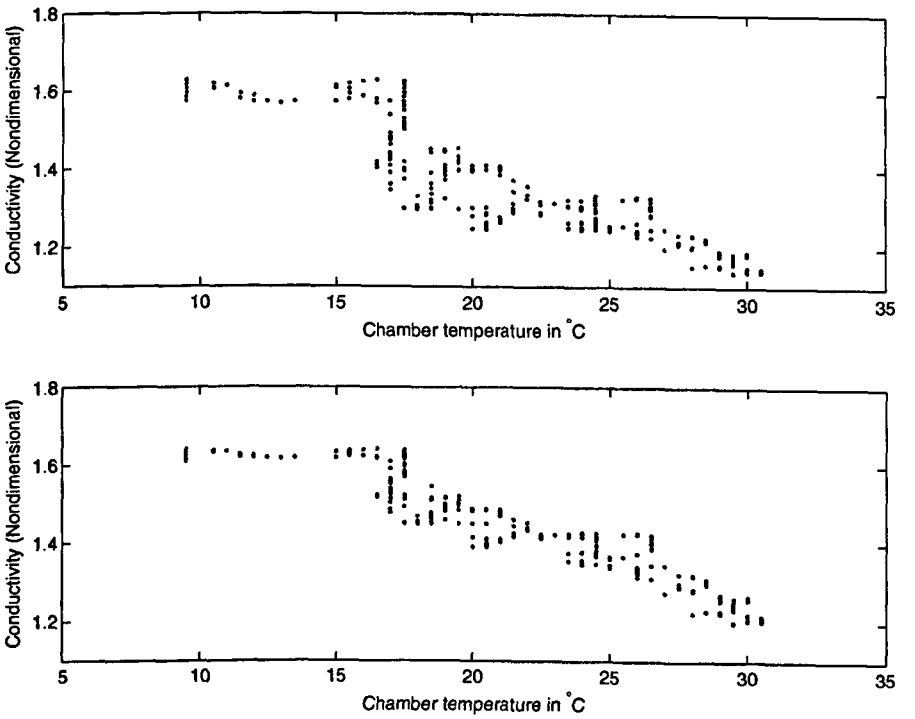


Figure 10.20: PSB based sensor pair temperature dependence of baseline conductance.

10.5.5 Film Thickness

There are no obvious dependencies of the film thickness upon the temperature. This might be because the film thickness interacts with the initial conductivity to give the baseline current. However the plots (Figures 10.23, 10.24 and 10.25) do show that the film thickness changes by a relatively small amount.

10.6 Discussion

A novel model proposed in this thesis has been fitted to a large number of experimental data sets. The resulting estimated parameter values have then been analysed for their environmental dependencies. The object of this analysis is to understand more clearly how the dynamics of an electronic nose system are affected by the environment in which it is operating. The model was successfully fitted to data obtained over a wide range of environmental conditions.

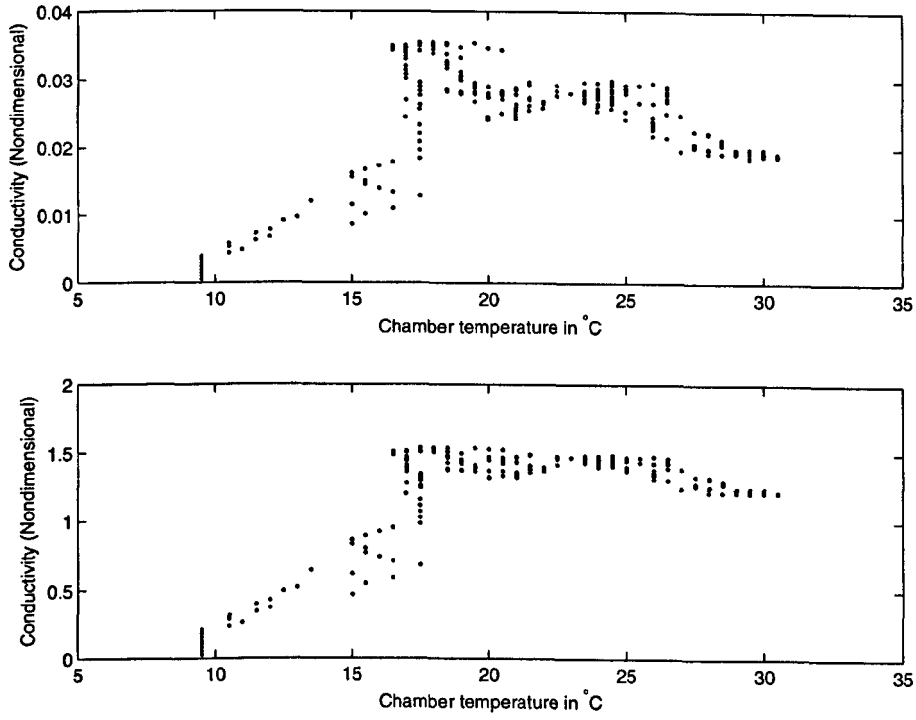


Figure 10.21: PVA based sensor pair temperature dependence of baseline conductance.

The rate at which the sensors react is clearly environmentally dependent. The increase in diffusion rate is very much temperature dependent. There are two possible effects contributing to this. The first is that mass action is temperature dependent and so reactions will tend to equilibrium more quickly at a higher temperature. Secondly it was demonstrated in Chapter 8 that diffusion in a polymer is concentration dependent. It is also known that the sample concentration might increase with temperature, so this could also explain the observed behaviour.

The flow dependence is extremely interesting with regard to the design of electronic nose systems. Previous work has suggested that increasing the flow rate greatly increases the rate at which the sensor reacts. However, the observed behaviour here suggests that the situation is more complex. The 'dead zone' for intermediate flow rate might be explained by the interaction of several conflicting processes:

- High flow exhausts initial sample headspace more quickly.
- High flow generates headspace more rapidly via sparging

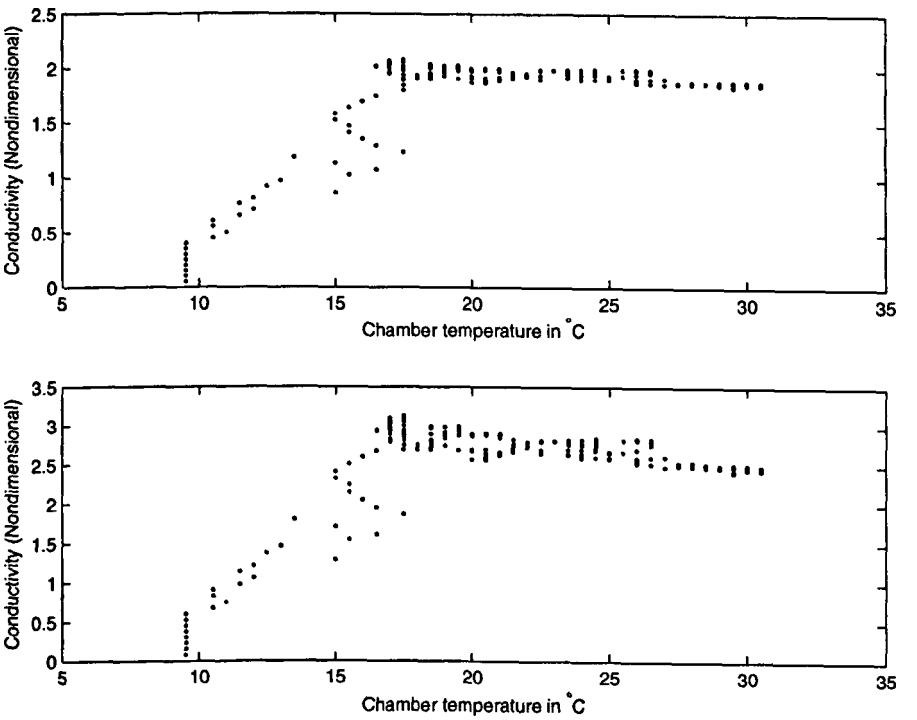


Figure 10.22: PCL based sensor pair temperature dependence of baseline conductance.

- High flow rate generates a greater response, possibly due to the changing fluid flow boundary layer at the sensor surface.

As shown in Chapter 9, the sample concentration will move between equilibrium with and without fluid flow, the equilibrium with flow being lower. At lower flow rates the difference will be very small, so the observed concentration will be high. With a higher flow rate (exhausting the headspace in 1 or 2 seconds) the concentration will drop by a much larger amount and so the 'time average' concentration will be lower.

This effect is counteracted partly by sparging, which 'bubbles' the carrier air through the solvent liquid driving off more volatiles into the atmosphere of the sample chamber. The greater the flow rate, the greater the number of solvent molecules driven into the gas phase and so the initial concentration is increased, as well as the ongoing flux of molecules into the gas phase.

The flow rate will also affect the concentration of molecules reaching the sensor film skin. This may be because of the interaction of fluid flow and diffusion of the solvent

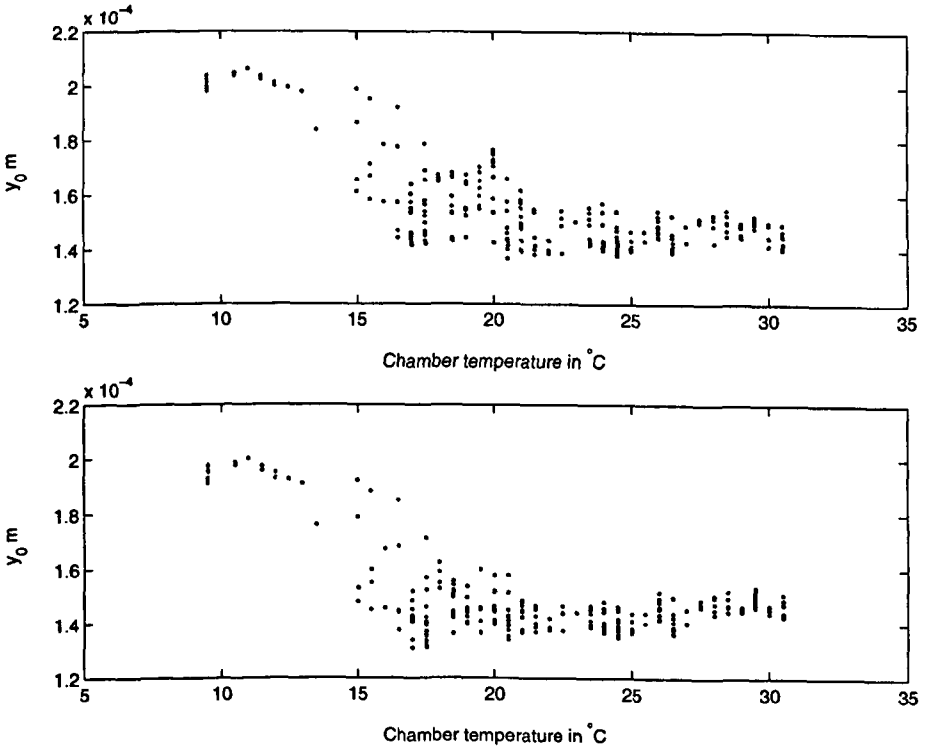


Figure 10.23: PSB based sensor pair temperature dependence of initial film thickness.

in the carrier gas. As noted in the last chapter, the velocity profile in the smart nose is parabolic. The flow rate is at a maximum at the centre of the channel and is, in theory, static at the surface of sensor. This means that diffusion across the profile will play a part in delivery of the solvent vapour to the sensor. At low velocities, the width of the 'stagnant' area near the sensor surface will be wider than at high velocities. This means that diffusion will be the main transport effect. The result of this might be a lower observed concentration at the sensor.

There is some weak humidity dependence. One might have expected a greater dependence given that diffusion is concentration dependent. This possibly means that water and the solvents interact in a nonideal manner; there is no linear superposition of the effects of the water and the solvent.

The effects of temperature on NQc_0 are very interesting and have served to highlight the interaction of processes on the periphery of the electronic nose system. It can be seen that a solvent-polymer interaction can be characterised by the temperature

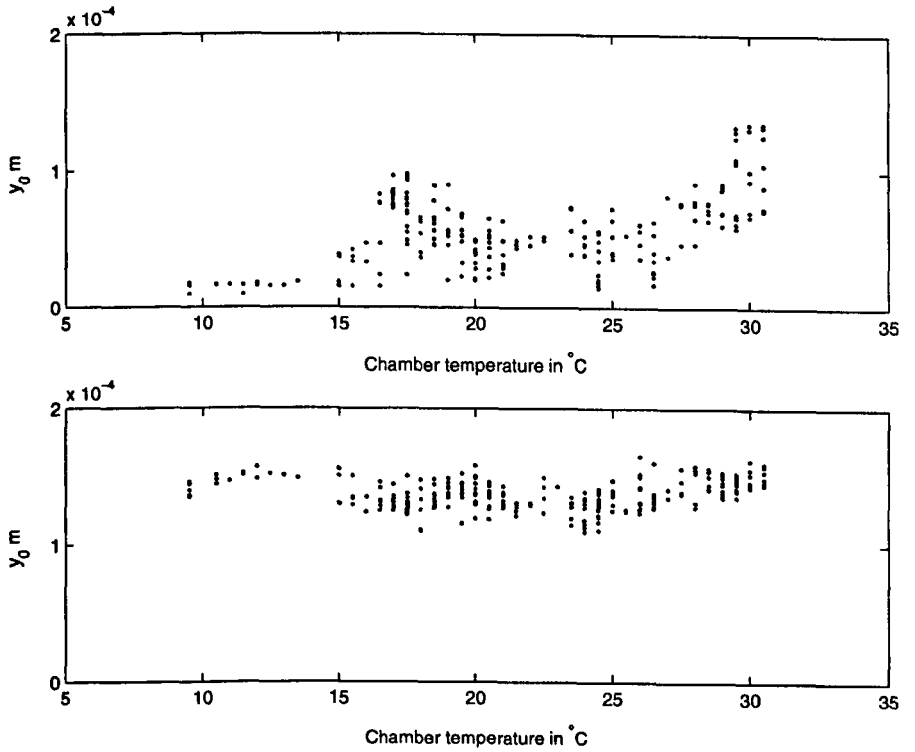


Figure 10.24: PVA based sensor pair temperature dependence of initial film thickness.

dependence of the current change. This gives possible design strategies for new electronic nose systems. This has already been investigated in part in [173]. This approach could be broadened to flow aspects [13] and humidity variations. This work has found fruition already in temperature modulated devices [174].

These results also point to the problem of an 'optimal' sensor array temperature, which, it would appear, is solvent and polymer dependent. The boiling point law (8.78) was not observed directly as Q could not be identified uniquely in the model.

Conductance results were roughly in line with what was thought to happen in the physical system, the film swelling as it becomes warm. There was some anomalous behaviour at lower temperatures, though this may be to due to the way that conductance and film thickness naturally interact to give the initial current. A relationship between these two variables means that there are not necessarily unique parameters for each. The anomalies may be as a result of the optimisation routine converging more easily to a different set of solutions to those preferred for higher temperature responses.

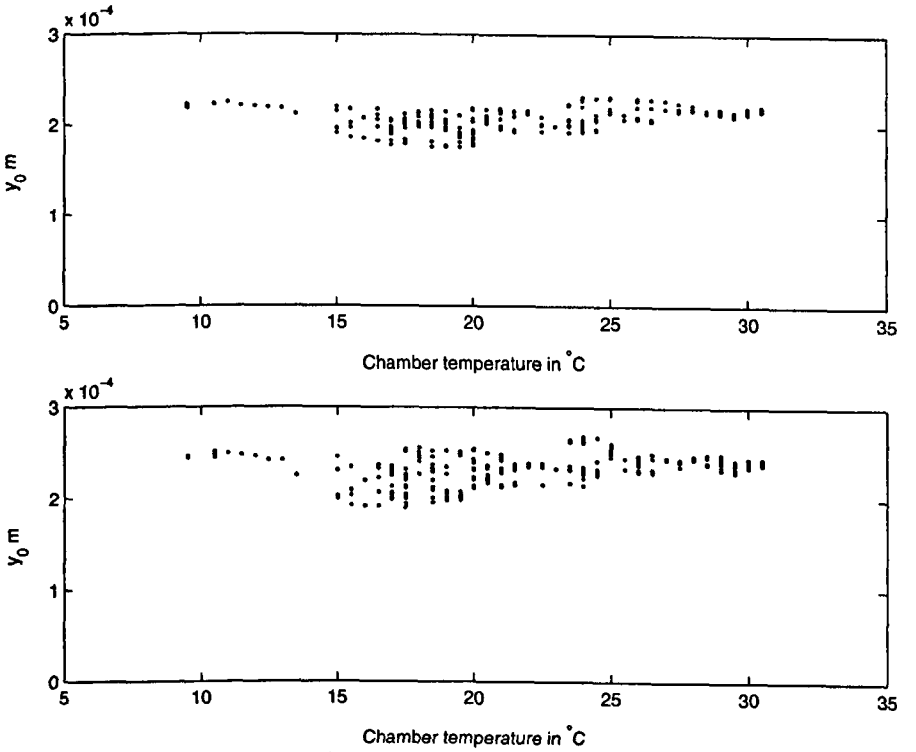


Figure 10.25: PCL based sensor pair temperature dependence of initial film thickness.

Broadly speaking, the observations made were consistent with previous studies. This means that a great deal of confidence can be placed on these results.

These experimental data do make full environmental analysis difficult, that is the uncontrolled environmental conditions. The result of this was a rather uneven grid of test conditions. However, the uncontrolled nature of the experiments was a major aim of the research. The distributions of the various environmental conditions are shown in Figures 10.26, 10.27 and 10.28.

It can be seen in Figure 10.26 that the majority of experiments were performed at $17°C$ and above. This would suggest that the relatively small number of experiments performed at temperatures lower than this are not to be held in much confidence. This is upheld in the observations above, with a different low temperature mode for $10 - 15°C$ than that for the larger range $15 - 30°C$. On the whole the distributions of flow rates and humidity are good.

This uncontrolled grid of experiments has meant that it is difficult to analyse the

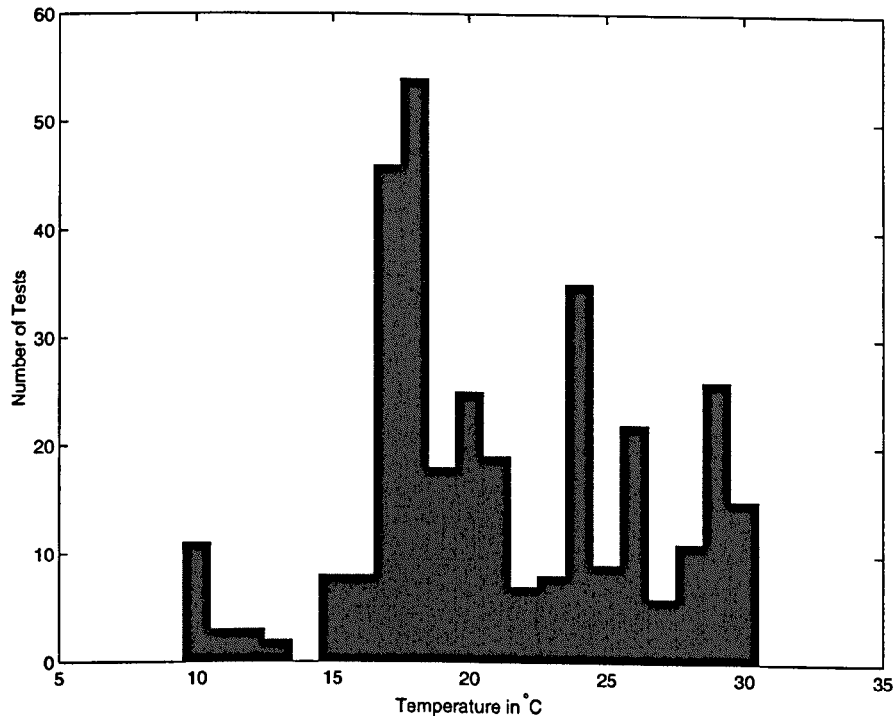


Figure 10.26: Distribution of Temperature at time of experiment.

separate effects provided by the environment. However the investigation has highlighted the complex nature of the temperature dependence of the observed concentration Qc_0. The model has shown itself to be valid over a wide range of conditions and to display the previously empirically observed behaviour.

It has been observed as well that the environmental dependencies characterise the sensor/solvent interaction. The degree of dependence varies from sensor to sensor. It can be seen that the rate constants of PSB and PCL type sensors are highly linear in their temperature dependence (Figures 10.5 and 10.7) whereas the relationship for PVA is not as well defined (Figure 10.6). The dependencies of the product NQc_0 are also interesting; the sign of the relationship slope alone provides a dichotomy of temperature limited and enhanced detection.

There are some disparities between the model and experimental responses. The model predicts more rapid initial dynamics than are actually observed. The reason for this is probably the nature of the error in the steady state approximation, which

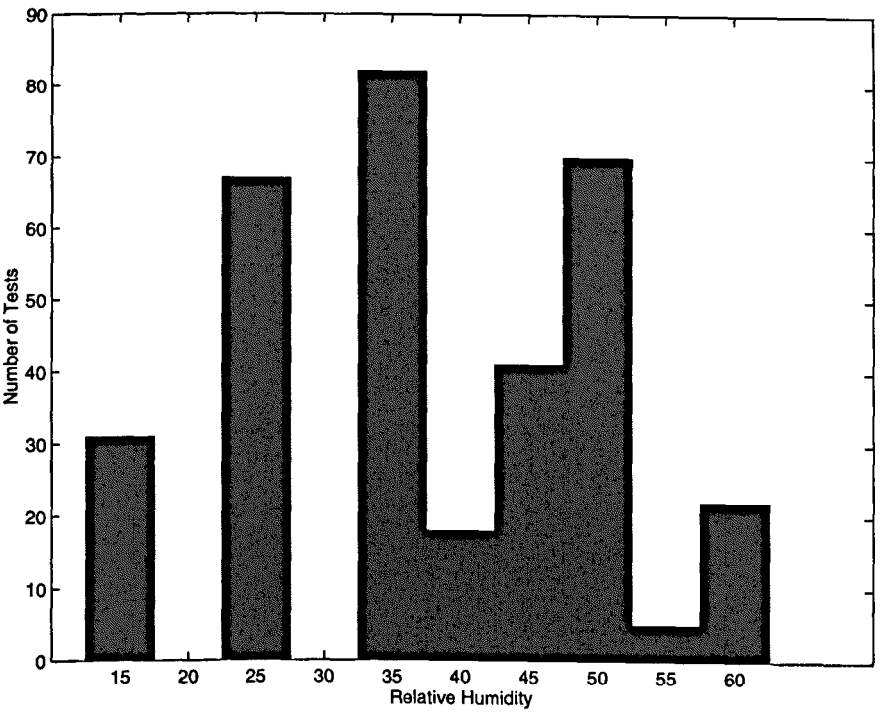


Figure 10.27: Distribution of Humidity at time of experiment.

naturally assumes rapid dynamics. The error plot shown in Figure 8.8 shows an error in the form of a surplus concentration diffusing into the polymer film. This would account for this rapid change in current at the beginning of the simulation.

Eradicating this means eradicating the steady-state assumption. Using the Neumman solution of the form (8.29)(8.30) and (8.31) would result in great difficulties in obtaining an analytical solution. However it could be simulated numerically. The ultimate goal would be to simulate the original nonlinear PDE (8.1). However, this would necessitate some Finite Element or Finite Difference approach which would be extremely computationally expensive with regard to parameter estimation.

The current changes observed in the experimental data are not monotonic with respect to time; there appears to be some superimposed oscillation. This may be due to the pump supplying an intermittent flow. However, it was placed half way between the sample and sensor chambers, 300mm from each, which should have eradicated this. The oscillation, however, demonstrates the rapid reversible dynamics of these sensors.

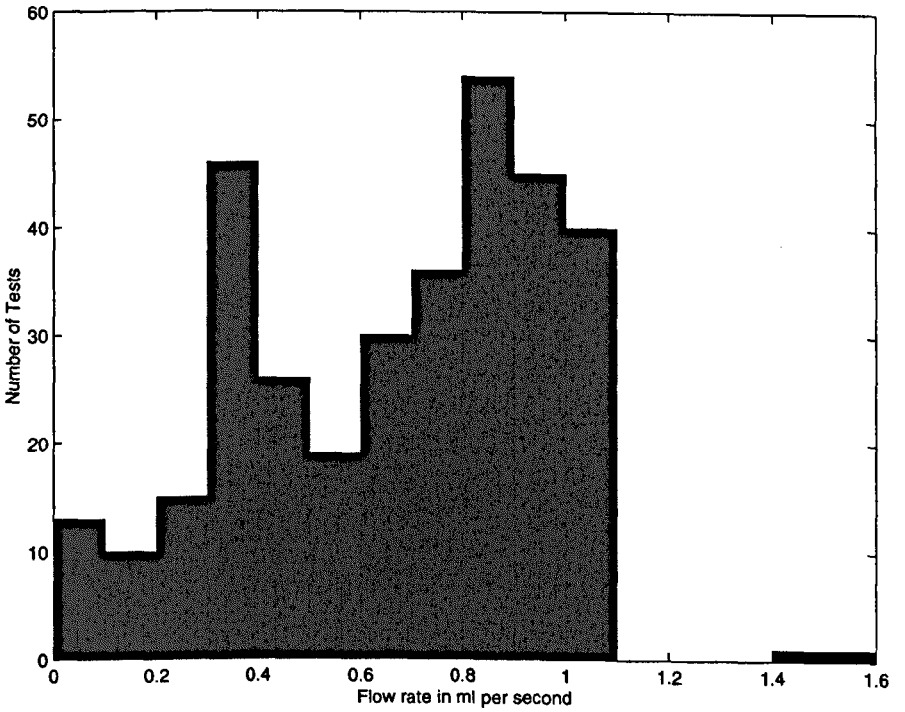


Figure 10.28: Distribution of Flow rate at time of experiment.

There are other sources of error. There is still some variation in parameter values for a fixed temperature and humidity. Air pressure was not monitored, but this will have an effect on the dynamics of the system [175]. The partition coefficients will vary with air pressure, and this may have an effect on the estimates for the rate constants and the product NQc_0, as discussed above. There will also be a change in the volume of the headspace in the sample chamber over time, due to the slow evaporation of the liquid solvent. This will alter the amount of solvent vapour available at static equilibrium concentration, thus changing the 'time averaged' concentration observed.

It is also suspected that the pump characteristics changed with time. The pump ran for 336 hours of testing, plus the several weeks it took to calibrate the experimental rig. In this time it came into contact with a great deal of solvent vapour that may have changed the characteristics of the pump's rubber diaphragm.

It is possible that, after so many exposures, the properties of the sensors themselves may have changed. This, however, is difficult to judge due to all the other factors

involved in the sensor responses.

To reiterate: the model fit in Figure 10.2 demonstrates the model predicts the behaviour of the real system well; the fitting error over all experiments, Figure 10.3, demonstrates good fitting over the time span of the experiments; the environmental dependencies of the model parameters appear, on the whole, to be coherent and corroborated by previous work. The model also demonstrates how physical fabrication dependent characteristics can be separated from the chemical, polymer/solvent dependent, characteristics by using mechanistic modelling. This will help to reduce

- Inconsistency of sensor response between sensor arrays due to fabrication variations.
- The effects of long term sensor drift.

It can thus be concluded that generally the model is a valid approximation of the real system and has successfully analysed the experimental data produced.

Chapter 11

Conclusion

Black box and mechanistic modelling of electronic nose systems has been investigated. The motivation for this work was to bring about a greater understanding of these systems. The exposition was illustrated with examples using both synthesised and real data obtained in the laboratory.

The aims of this thesis were threefold:

- To analyse the behaviour/output of electronic nose systems.
- To model the internal mechanisms of carbon black-polymer sensors.
- To investigate the effects of environmental conditions on the input-output behaviour of an electronic nose system.

11.1 The Empirical Nature of Black Box Modelling

A number of different models were proposed to analyse real electronic nose data. The problems of small data sets and tuning complex models have been highlighted. An investigation was made into the relationship between *model complexity*, the amount of data available and model parameter tuning. It was concluded that the number of model parameters as well as the smoothness of the target function are important. The smoothness of the target function is largely influenced by the scale of separation

between two data classes. This therefore dictates the amount of data needed to robustly identify a good approximation of the target discriminate model. Therefore the pitfalls of the empirical data driven modelling technique were investigated.

The focus of the investigation then turned to resolving this problem by data set reduction. The interaction between different black box model types and two data reduction techniques were investigated. It was concluded that the choice of the two was very important. A judicious choice of preprocessing method can lead to an increase in model accuracy of some 20-30% from that obtained from the raw data.

Novel methods of parameter identification were investigated; namely genetic algorithms and a batch gradient search method. Various methods of implementing these algorithms were compared to the Support Vector Method for the tuning of Radial Basis Function Neural Networks. It was concluded that the gradient search method was too sensitive to the error curve in the solution space and was prone to oscillation, while the genetic algorithm was much better for the generation of 'initial guesses' for parameter estimation algorithms. The SVM overall was far more robust for this application.

These new methods were then applied to several data sets of biomedical processes. They were all challenging in nature as they attempted to detect signals with a large amount of background noise present. The conclusions of this work were that electronic nose systems show great promise in a medical diagnosis context.

11.2 Mechanistic Modelling

In the second half of the thesis a different approach was taken; mechanistic modelling of a particular type of gas sensor. Carbon black-polymer sensors were modelled using a nonlinear diffusion model. This was coupled with an electron hopping conduction model to give an analytic solution for the time dependent change of current through a sensor. This assumed a steady state solution and rapid dynamics were assumed. A model for six such sensors in three pairs responding to two different solvent samples was produced by using the single sensor solution twelve times with common parameters representing the relationships between sensors of the same type and the same sensor responding to two different solvent samples.

The model fits demonstrate that the assumptions made on the model are sufficient to generate the real system behaviour, except for the steady state solution adopting initial dynamics that were too rapid. This still only resulted in an error of approximately 5% in most cases.

The model was fitted to a series of experimental sensor exposures, where environmental conditions were recorded. The data were produced using a newly constructed test rig designed specifically for the task. The resulting dependence of estimated model parameter values on these conditions was then investigated. The dependence of parameter values largely agreed with previous empirical observations. Most interesting were the temperature dependencies of the product of parameters NQc_0 which had previously been observed to reduce with temperature. Here, however, it was observed to increase in some cases. This is because the temperature also dictates the gas phase concentration. Thus it was concluded that the temperature dependence of the interaction between solvent liquid phase evaporation and condensation into the sensor was a particular characteristic of a particular solvent/polymer pair.

It is concluded that the models give us a greater understanding of the behaviour of electronic nose systems. They also allow us to investigate the time dependence present within the system, how environmental factors affect the behaviour and which characteristics are important when solving the inverse model problem.

The results of this work indicate that electronic noses may become important tools, especially if more of a dynamical systems, rather than a statistical, modelling approach is taken. This will allow us to make important decisions at the design stage of the physical system and to analyse the resulting data more efficiently.

11.3 Future Work

The directions of future work have been discussed implicitly throughout this thesis. These may be summarised as:

- Further data set reduction techniques
- Larger data sets

- Development of more sophisticated genetic algorithms
- Numerical simulation of solvent transport models
- Sample analysis and identification by analysis of the electronic nose system dynamics

11.3.1 Data set reduction

It is apparent from the findings that more sophisticated data preprocessing techniques lead to more rigorous model building. The essential aim of the work documented in this thesis was to understand how the modelling of electronic nose systems may be improved.

It is therefore suggested that a more in-depth analysis of pre-processing and information extraction from electronic nose data is carried out. More general, possibly *ad hoc*, techniques are required that can be incorporated into the data processing modules of electronic nose systems in order to extract information from gas sensor signals.

Feature extraction is the most promising technique especially when considering the cross-sensitivity of an array of gas sensors. The sensor output to feature mapping might lead to some novel system design, for example a set of the same type of sensor but with different film thicknesses.

It is apparent that more sophisticated data gathering is required. Data sets need to be as large as possible so that outliers can be identified and eliminated and so that models may be identified as robustly as possible.

11.3.2 Larger data sets

The investigations reported here resulted in a few hundred samples using the Automatic Headspace Sampler of a mass spectrometer. It would be interesting to investigate more extreme systems modelling where there are several thousand samples to analyse. It would be expected that a number of issues, including computer speed, would then be manifested.

These data would have to be collected as part of a much larger scale project. It would be preferable to source the samples from a hospital or clinic as these, for example the urine samples of Chapter 7, provide a much more challenging classification problem than laboratory produced samples.

11.3.3 Development of more sophisticated genetic algorithms

The genetic algorithms discussed in this thesis have proven themselves useful for black box *and* mechanistic model identification. This is not electronic nose system specific. The major advantage of this algorithm has been the selection of good start values for more conventional minima search algorithms.

It is apparent though from the results in Chapters 4 and 5 that the algorithms need to be more sophisticated. The oscillations and loss of good solutions exhibited suggest that emphasis should be placed upon the crossover, or mating, aspects of the algorithm. A number of issues need to be addressed, all coming under the umbrella of 'symmetry of solutions'.

It became apparent in Chapter 5 that the parameters representing a neural network model may be permuted to give the same input-output response. For a mechanistic model it is possible that there are several sets of parameter vector solutions. The consequence of this is that the genetic algorithm will attempt to cross chromosomes that represent points close to two different minima; resulting in offspring that represent poorer solutions.

It seems necessary then to investigate ways of checking whether a pairing of individuals is 'good' or not. A definition of 'good' might be based upon some broad idea of the individuals being part of convergence to the same minimum.

11.3.4 Numerical simulation of solvent transport models

The mechanistic model presented in this thesis was an approximation of a nonlinear model that was developed based upon established properties of solvent polymer interactions. The model that was produced gave more insight into the dynamics of an electronic nose system.

However a simulation of the original model could provide more important information. This would probably mean a finite element based simulation to numerically solve the nonlinear PDE (8.2). Coupling this with a conductance model such as in (8.61) to give the observed current change dynamics would be extremely costly computationally. This could rule out parameter estimation because of the large number of repeat simulations that such a process would require.

11.3.5 Sample analysis and identification by analysis of the electronic nose system dynamics

Further dynamic analyses involving parameter estimation are still required. The first recommendation would be to carry out a second environmental dependency study using a more controlled environment. This would mean that different conditions, such as humidity and temperature, could be altered separately.

A second recommendation would be to investigate whether the estimated parameters might be used to give more accurate discrimination between odours. A secondary aim would be to see whether a greater understanding of the environmental dependencies might allow the building of system models that accurately model the real system under different environmental conditions.

11.4 Concluding Remarks

A systems modelling approach has been applied to electronic nose systems. Techniques have been proposed and tested for robust data analysis using black box models. Data set reduction and novel optimisation algorithms have proved advantageous for model building. This has been done with the aim of extending the biomedical applications of such systems. It has been demonstrated how mechanistic modelling may be used in the future to characterise electronic nose systems by the environmental dependencies of its outputs. The models and modelling techniques presented in this thesis demonstrate the advantage of applying rigorous systems modelling techniques to electronic nose systems in order to gain a greater understanding of these systems.

Appendix A

Derivation of Thermodynamic Model

A.1 Diffusion into Polymer

The standard partial differential equation for diffusion, as given in [176], is

$$Lc = \{\frac{\partial}{\partial t} - D\nabla^2\}c(\mathbf{r}, t) = \rho(\mathbf{r}) \tag{A.1}$$

where c is the concentration of particles, and ρ is the number injected into the system per unit volume per unit time. The variables t and $\mathbf{r}$ correspond to time and space respectively. D is the diffusion constant with dimension $[L^2T^{-1}]$, using assumption (i) of Chapter 8 . Of course here, in theory, each unit volume can contain any arbitarily large amount of solvent.

Diffusion of a substance can be discussed in terms of the flux of transport, J. It is often called the diffusion current density. Hence equation A.1 can be written,

$$\frac{\partial c}{\partial t} = -\nabla J + \rho \tag{A.2}$$

Equations (A.1) and (A.2) imply Fick's law

$$J = -D\nabla c \tag{A.3}$$

Due to this, however, the model proposed is unfaithful to the true situation. Solutions to (A.1) allow for an arbitrarily large capacity in all volume elements. Indeed, the larger the concentration gradient, the greater the magnitude of the flux of transport. However, in polymers there is 'reluctance' to accommodate more solvent. This can be illustrated by considering the related processes in Gas Chromatography. A gas chromatograph is essentially a long narrow hollow column packed with fine particles. As a sample, initially in its mobile vapour state, is introduced it goes through a continuous sorption-desorption cycle between the mobile phase and a stationary phase within the column material. This set up results in a diffusion rate that varies across chemical species. Three different diffusion processes are present:

1. **Ordinary Diffusion.** This is the process driven by a concentration gradient.

2. **Eddy Diffusion.** There is a variation in the size of particle sizes in the column. Therefore each molecule 'sees' a different path, resulting in a variation in the speed of individual molecules.

3. **Local Nonequilibrium** As a 'plug' travels along the column it spreads out due to varying diffusion constants. Hence, at every point the stationary phase tries to reach equilibrium with a time dependent mobile phase. This makes the sorption-desorption process time dependent.

This two phase diffusion is relevant to polymer diffusion. It would be expected to be a pure concentration gradient driven flux. However, as has been noted above and in [177] the process is non-Fickian. Homogeneity is assumed, so Eddy Diffusion is neglected. However the local non-equilibrium effect will be the major factor in the reluctance to absorb. The stationary phase will cause the thermodynamic changes discussed below.

A.1.1 Modelling using Chemical Potential

First, consider the pore model which is a standard diffusion model for polymers [147]. Consider a pore of length $2L \gg R_p$ where R_p is the pore radius, and assume the

solid phase is immobile. The pore can then be considered as a line segment oriented along the z axis. Keeping faithful to assumption (iv) in Chapter 8, consider a single component bulk fluid. The flux can then be expressed as

$$J_f = -\frac{n_f D_{fm}}{k_b T}\frac{d\mu_f}{dz} \tag{A.4}$$

Then for the situation where the bulk external phase is in equilibrium with the pore fluid we have that

$$\mu_f = \mu_{fB} = \mu_0(T) + k_B T \log_e \zeta_B \tag{A.5}$$

where ζ_B is the chemical activity of the bulk fluid. It is assumed that the pore fluid is whatever the sensor was purged with, and the bulk external fluid is the solvent.

The chemical activity should be defined at this point. In [175] the chemical activity is defined in terms of an equation similar to (A.5). It is then deduced that the activity has units of pressure. For an ideal gas the assumption is made that $\zeta_B \longrightarrow n_{fB}$ as $n_{fB} \longrightarrow 0$. Hence, equations (A.5) and (A.4) together give

$$J_f = -D_{fm} K_f \frac{d\ln\zeta_B}{d\ln n_{fB}}\frac{dn_{fB}}{dz} \tag{A.6}$$

or, for an ideal gas, in the limit as $n_{fB} \longrightarrow 0$

$$J_f = -D_{fm} K_f \frac{dn_{fB}}{dz} \tag{A.7}$$

Here K_f is the partition coefficient of the fluid f, defined by

$$K_f = \frac{n_f}{n_{fB}} \tag{A.8}$$

Note that (A.7) together with (A.2) yields (A.1) in one dimension.

A.1.2 Polymer Cross-Linking

To calculate the chemical potential it is neccesary to calculate the energy changes in the polymer body as it dissolves into a solvent. Most polymers exhibit an elastic behaviour due to two effects;

- Polymer chains, in an undisturbed state, rest in a minimum energy configuration. When this configuration is altered there is an increase in potential and thus there is a force produced in reaction to this.
- There is an interaction between chains, often referred to as cross-linking.

A.2 Thermodynamic modelling of solvent sorbtion

An expression is now derived for the chemical activity of a polymer dissolving in a polymer with terms representing osmotic pressure and the elasticity of the polymer chains. By osmotic pressure is meant the self-diffusion due to a concentration gradient. As above an expression is sought for the chemical potential as a function of space. To achieve this the definition of chemical potential is first given [178],

$$\mu_1 = \left(\frac{\partial(nG)}{\partial n}\right)_{T,P} \tag{A.9}$$

Here G is the molar energy and n is the number of moles. A modified version is considered which replaces this measure with total energy and number of molecules. Here we follow the working of [133], which is actually a static model considering the polymer as a whole. However it may be extended to a dynamic model as the expressions are in terms of volume fractions and the number of solvent molecules.

The internal energy is decomposed into two independent terms,

$$\Delta G = \Delta G_{mix} + \Delta G_{el} \tag{A.10}$$

where G_{mix} is the free energy of mixing and the elastic free energy due to expansion of the polymer network is denoted by G_{el}.

A.2.1 Free energy of mixing

It is assumed that the entropic contribution is due to change of configuration. For a solvent-polymer mix, the change in entropy is

$$\Delta S^*_{mix} = -k(n_1 \ln v_1 + n_2 \ln v_2) \tag{A.11}$$

and the heat of mixing (enthalpy) is

$$\Delta H = \chi_1 kT n_1 v_2 \tag{A.12}$$

The mixing term is first derived; for an ideal mix,

$$\begin{aligned} \Delta G_{mix} &= \Delta H_{mix} - T\Delta S_{mix} \\ &= \Delta H_{mix} - T\Delta S^*_{mix} \end{aligned} \tag{A.13}$$

where H_{mix} is the enthalpy due to mixing and S_{mix} is the entropy of mixing. It is assumed that the entropic contribution is due to the change of configuration of the polymer links. For a solvent-polymer mix, the change in entropy is:

$$\Delta S^*_{mix} = -k(n_1 \ln v_1 + n_2 \ln v_2) \tag{A.14}$$

and the enthalphy is given by

$$\Delta H = \chi_1 kT n_1 v_2 \tag{A.15}$$

Here n_i is the number of molecules and v_i is the volume fraction of substance i. In the case of a single component solvent, the solvent is set to be substance 1 and the polymer is substance 2. The constant χ_1 is a constant which characterises the solvent-polymer interaction

By substituting Equations (A.11) and (A.12) into (A.13) then the *Flory-Huggins* expression for the free energy of mixing are obtained:

$$\Delta G_{mix} = kT(n_1 \ln v_1 + n_2 \ln v_2 + \chi n_1 v_2) \tag{A.16}$$

In the case of solvent sorption in a polymer $n_2 = 0$ as it is assumed that there are no individual polymer molecules in the network structure.

A.2.2 Free energy due to elastic deformation

The free energy due to elastic deformation is now considered. Similar to Equation A.13 the free energy change due to elastic deformation is expressed as

$$\Delta G_{el} = \Delta H_{el} - T\Delta S_{el} \tag{A.17}$$

It is assumed that the change in enthalpy is zero, i.e. that is there is no change in the internal energy of the system. A change in free energy occurs because the polymer chains are perturbed from their preferred state of lowest energy. The expression put forward in [133] is

$$\Delta G_{el} = \left(\frac{kTv_e}{2}\right)(3\alpha_s^2 - 3 - \ln\alpha_s^3) \tag{A.18}$$

where α_s is the linear factor of deformation and v_e is the density of crosslinks.

In calculating the change in entropy a statistical model of the polymer chain configuration must be considered. Below the configuration of chains is represented as being normally distributed. This is adequate for small extensions of the polymer chains [133], however at large extensions other alternative distributions, such as the inverse Langevin function should be considered.

A network of v chains is considered.

The end-to-end vectors of the polymer chains are considered to be distributed as

$$W(x,y,z) = \left(\frac{\beta}{\pi^{\frac{1}{2}}}\right)^3 e^{-\beta^2 r^2} \tag{A.19}$$

where $r^2 = x^2 + y^2 + z^2$ and $\beta = \frac{\sqrt{\frac{3}{2}}}{n^{\frac{1}{2}}l}$. Let the network now be subjected to a homogeneous strain described by the factor α_s in the x, y and z directions. Hence the mean position of a cross link relative to another must change by the same factor. Hence, any chain characterised by the vector $r_i = (x_i, y_i, z_i)$ after extension must have had components $(\frac{x_i}{\alpha_s}, \frac{y_i}{\alpha_s}, \frac{z_i}{\alpha_s})$. Thus the number of chains having their end-to-end vector components changing from x_i to $x_i + \Delta x$, etc. after deformation is given by:

$$v_i(x_i, y_i, z_i) = vW\left(\frac{x_i}{\alpha_s}, \frac{y_i}{\alpha_s}, \frac{z_i}{\alpha_s}\right)\frac{\Delta x \Delta y \Delta z}{\alpha_s^3} \tag{A.20}$$

where v_i is for the i-th configuration. The configuration entropy change involved in the formation of a network structure in a deformed state specified by α_s is now calculated.

The probability Ω that the unlinked polymer occurs in a configuration that is consistent with that of a deformed network with v_i chains having end-to-end vector (x_i, y_i, z_i) as required by the deformation factor α_s needs to be calculated. Let us suppose that

u of the units of the polymer molecules have been designated to participate in crosslinking. For these chains to be acceptable in the formation of the network specified by the deformation factor, the configuration of the u chains must fulfil the following conditions:

1. The chains must occur in the correct distribution specified by the v_i.

2. The units designated for crosslinking must occur in the correct juxtaposition.

The probabilities that these conditions are fulfilled are designated Ω_1 and Ω_2 respectively. Hence we have $\Omega = \Omega_1\Omega_2$, as it is assumed these are independent events. For convenience we denote the probability that a given chain has components within Δx, Δy, Δz of (x_i, y_i, z_i)

$$\omega_i = W(x_i, y_i, z_i)\Delta x \Delta y \Delta z \tag{A.21}$$

Thus the probability that each chain in the network complies with condition 1 is the product of all such terms. Since which particular chain takes up which configuration is of no importance, by multiplying by the number of possible permutations we arrive at

$$\Omega_1 = u! \prod_i \frac{\omega_i^{v_i}}{v_i!} \tag{A.22}$$

On taking logs and introducing Stirling's approximation for the factorials,

$$\ln \Omega_1 = \sum v_i \ln \frac{\omega_i u}{v_i} \tag{A.23}$$

After substituting (A.19) into (A.20), we obtain

$$\ln\left(\frac{\omega_i u}{v_i}\right) = \beta^2 \left(\frac{1}{\alpha_s^2} - 1\right)[x_i^2 + y_i^2 + z_i^2] + \ln(\alpha_s^3) \tag{A.24}$$

Substituting (A.24) into (A.23) and replacing the summation by an integration over $\mathbb{R}^3$ the following is obtained

$$\ln \Omega_1 = -u\left[\frac{3\alpha_s^2 - 3}{2} - \ln \alpha_s^3\right] \tag{A.25}$$

The second probability factor, Ω_2, is now considered. Condition 2 specifies that all the designated polymer units must be paired up (hence we assume that v is even). The probability that one of the u units has another such unit neighbouring it in the required volume δV is $(u-1)\frac{\delta V}{V}$. Hence, by requiring that all the units to be paired up, the probability is

$$\begin{aligned} \Omega_2 &= (u-1)(u-3)\ldots(1)\left(\tfrac{\delta V}{V}\right)^{\frac{u}{2}} \\ &\cong \left(\tfrac{u}{2}\right)!\left(\tfrac{2\delta V}{V}\right)^{\frac{u}{2}} \end{aligned} \tag{A.26}$$

Replacing V with $\alpha_s^3 V_u$ where V_u is the volume of the undeformed sample, the result

$$\ln \Omega_2 = -\left(\frac{u}{2}\right)\ln \alpha_s^3 + \text{const} \tag{A.27}$$

is obtained. The constant term in Equation (A.27) is independent of the deformation and so it will soon become clear it is unimportant. Substituting (A.25) and (A.27) into (A.17) results in (A.18).

A.2.3 The final PDE

The above thermodynamic model is used to derive a diffusion PDE. A one dimensional case is considered, given the inherent symmetries of a solvent diffusing into a polymer surface.

It is desirable to write a PDE in terms of the concentration of the solvent. This is so that the initial conditions may be written in terms of the concentration at the surface of the polymer (here it is assumed that the sensor is in a sufficiently large chamber, so that there is no drop in concentration as the solvent vapour is absorbed into the polymer.

Using the definition in (A.9)

$$\mu_1 - \mu_0 = \frac{\partial G_{mix}}{\partial n} + \frac{\partial G_{el}}{\partial n} \tag{A.28}$$

The result for the mixing component is assumed to be as in Section A.1.1. The elastic term is therefore analysed and it is assumed that flux due to osmotic pressure and deformation combine linearly.

In [133] the following expression was derived:

$$\mu_{el}^{1} - \mu_{el}^{0} = KTv_1 \left(\frac{v_e}{v_u}\right)\left(v_2^{\frac{1}{3}} - \frac{v_2}{2}\right) \tag{A.29}$$

where the variables have been previously defined. Note this equation is entirely in terms of volume fractions and the density of the polymer chain network. To introduce the density function $u(x,t)$ it is noted that $v_2 + v_1 = 1$ and the following assumption is made:

Assumption vi The volume occupied by x solvent molecules is given by:

$$v = Nx \tag{A.30}$$

The following is obtained

$$\mu_{el}^{1} - \mu_{el}^{0} = KTNc\left[\frac{v_e}{v_u}\right]\left((1-Nc)^{\frac{2}{3}} - \frac{(1-Nc)}{2}\right) \tag{A.31}$$

If a one dimensional medium is considered then equations (A.2) and (A.4) imply that the second derivative with respect to space will give the rate of change of density with respect to time.

$$\frac{\partial c}{\partial t} = N\left[\frac{v_e}{v_u}\right]\left[\frac{\partial^2 c}{\partial x^2}\left((1-Nc)^{\frac{1}{3}} - \left(\frac{1-Nc}{2}\right)\right) + N\frac{\partial c}{\partial x}\left(1 - \frac{2}{3}(1-Nc)^{-\frac{2}{3}}\right)\right.$$
$$\left. + Nc\left(-\frac{2N}{9}\left(\frac{\partial c}{\partial x}\right)^2(1-Nc)^{-\frac{5}{3}} - \frac{\partial^2 c}{\partial x^2}\left(\frac{1}{2} - \frac{1}{3}(1-Nc)^{-\frac{2}{3}}\right)\right)\right] \tag{A.32}$$

This expression is immensly complex and would defy any analysis. Therefore, assume small concentrations of the solvent and use $(1-x)^{\epsilon} \approx (1-\epsilon x)$ for small x. This yields

$$\frac{\partial c}{\partial t} = \left[\frac{v_e}{v_0}\right]\left[\frac{\partial^2 c}{\partial x^2}\left(D + \frac{N}{3}c - \frac{2N^2}{9}c^2\right) + N\left(\frac{\partial c}{\partial x}\right)^2\left(\frac{1}{3} - \frac{2N}{9}c + \frac{10N^2}{27}c^2\right)\right] \tag{A.33}$$

for some D.

References

[1] http://images.encarta.msn.com/xrefmedia/aencmed/targets/illus/ilt/T041371A.gif. December, 2004.

[2] http://soma.npa.uiuc.edu/courses/bio303/Image226.jpg. December, 2004.

[3] www.chemistry.adelaide.edu.au/external/soc-rel/content/quad-sch.png. December 2004.

[4] www.rohmhaas.com/company/plabs.dir/images/GasChro1.gif. December, 2004.

[5] P. Niyogi and F. Girosi. On the relationship between generalization error, hypothesis complexity, and sample complexity for radial basis functions. *Neural Computing*, 8(4):819–842, 1996.

[6] S.S.Schiffman and T.C. Pearce. Introduction to olfaction: Perception, anatomy, physiology and molecular biology. In T.C. Pearce, editor, *The Handbook of Olfaction*. Wiley,Weinheim, 2001.

[7] J.A. Maruniak, W.L. Silver, and D.G. Moulton. Olfactory receptors respond to blood-borne odorants. *Brain Research*, 265:312–316, 1983.

[8] S. Sachse, A. Rappert, and C.G. Galizia. The spatial representation of chemical structures in the antennal lobe of honey bees: Steps towards the olfactory code. *European Journal of Neuroscience*, 11:3970–3982, 1999.

[9] J.J. Rodriguez-Bencomo, J.E. Conde, F. Garcia-Montelongo, and J.P. Perez-Trujillo. Determination of major compounds in sweet wines by headspace solid-

phase microextraction and gas chromatography. *Journal of Chromatography A*, 991:13–22, 2003.

[10] F. Mayer and W. Grosch. Aroma simulation on the basis of the odourant composition of roasted coffee headspace. *Flavour and Fragrance*, 16:180–190, 2000.

[11] C.E. Reilly. Odourant intensity and valance are processed by distinct human brain regions. *Journal of Neuroscience*, 250:643–645, 2003.

[12] J.W. Gardner and P.N. Bartlett. A brief history of electronic noses. *Sensors and Actuators B*, 18-19:211–220, 1994.

[13] S.E. Stitzel, D.R. Stein, and D.R. Walt. Enhancing vapor sensor discrimination by mimicking a canine nasal cavity flow environment. *Journal of the American Chemical Society*, 125:3684–3685, 2003.

[14] H.A. Schultens and D. Schild. Biophysical properties of olfactory receptor neurones. In J.W. Gardner and P.N. Bartlett, editors, *Sensors and Sensory Systems for an Electronic Nose*. Kluwer Academic Publishers, London, 1992.

[15] A.M. Taurino, D. Delllo Monaco, S. Capone, M. Epifani, R. Rella, P. Siciliano, L. Ferrara, G. Maglione, A. Basso, and D. Balzarano. Analysis of dry salami by means of an electronic nose and correlation with microbiological methods. *Sensors and Actuators B*, 95(1-3):123–131, 2003.

[16] V.R. Kinton, J. Christenson, and B. Kolahgar. Fast determination of beverages using a mass spectrometry based chemical sensor. *Liquid Chromatography- Gas Chromatography in North America 2003*, 2003:42–43, 2003.

[17] B.P.J.D. Costello, R.J. Ewen, H. Gunson, N.M. Ratcliffe, P.S. Sivanand, and P.T.N. Spencer-Phillips. A prototype sensor system for the early detection of microbially linked spoilage in stored wheat grain. *Measurement, Science and Technology*, 14(4):397–409, 2003.

[18] K.D. Bartle. Introduction to the theory of chromotographic separations with reference to gas chromatography. In P.J. Baugh, editor, *Gas Chromatography, A Practical Guide*. Oxford University Press, Oxford, 1993.

[19] L.F. Di Cesare, E. Forni, D. Viscardi, and R.C. Nani. Changes in the chemical composition of basil caused by different drying procedures. *Journal of Agricultural and Food Chemistry*, 51:3575–3581, 2003.

[20] F.J. Senoráns, A. Ruiz-Rodríguez, E. Ibánez, J. Tabera, and G. Reglero. Isolation of brandy aroma by countercurrent supercritical fluid extraction. *Journal of Supercritical Fluids*, 26:129–135, 2003.

[21] R.C. Mejías, R.N. MarínM. de Valme, Garcia Moreno, and C.G. Barroso. Optimisation of headspace solid-phase microextraction for the analysis of volatile phenols in wine. *Journal of Chromatography A*, 995:11–20, 2003.

[22] C.-H. Wu, M.-N. Lin, C.-T. Feng, K.-L. Yang, Y.-S. Lo, and J.-G. Lo. Measurement of toxic volatile organic compounds in indoor air of semiconductor foundries using multisorbent adsorption/thermal desorption coupled with gas chromatography-mass spectrometry. *Journal of Chromatography A*, 996:225–231, 2003.

[23] K. Osada, K. Yamazaki, M. Curran, J. Bard, B.P.C. Smith, and G. Beauchamp. The scent of age. *Proceedings of the Royal Society of London Series B*, 270(1518):929–933, 2003.

[24] X.S. Chai and J.Y. Zhu. Determination of the solubility of inorganic salts by headspace gas chromatography. *Journal of Chromatograhy A*, 996:157–161, 2003.

[25] M.M. Baum, S. Kumar, , and A.M. Lappas. Measurement of acetylene in breath by ultraviolet absorption spectroscopy: Potential for noninvasive cardiac output monitoring. *Review of Scientific Instruments*, 74(6):3104–3110, 2003.

[26] T. Chen, M. Rimpiläinen, R. Luukkainen, T. Möttönen, T. Yli-Jama, J. Jalava, O. Vainio, and P. Toivanen. Bacterial components in the synovial tissue of pa-

tients with advanced rheumatoid arthritis or osteoarthritis: Analysis with gas chromatography-mass spectrometry and pan-bacterial polymerase chain reaction. *Arthritis and Rheumatism (Arthritis Care and Research)*, 49:328–334, 2003.

[27] J.S. Buyer. Improved fast gas chromatography for FAME analysis of bacteria. *Journal of Microbiological Methods*, 54:117–120, 2003.

[28] B. Toso, G. Procida, and B. Stefanon. Determination of volatile compounds in cows' milk using headspace GC-MS. *Journal of Dairy Research*, 69(4):569–577, 2002.

[29] A. Accorsi, A. Barbieri, G.B. Raffi, and F.S. Violante. Biomonitoring of exposure to nitrous oxide, sevoflurane, isoflurane and halothane by automated GC/MS headspace urinanalysis. *International Archives of Occupational and Environmental Health*, 74(8):541–548, 2002.

[30] P.N. Bartlett and J.W. Gardner. Odour sensors for an electronic nose. In P.N. Bartlett and J.W. Gardner, editors, *Sensors and Sensory Systems for an Electronic Nose*. Kluwer Academic Publishers, Dordrecht, 1992.

[31] J.W. Gardner and P.N. Bartlett. Pattern recognition in odour sensing. In J.W. Gardner and P.N. Bartlett, editors, *Sensors and Sensory Systems for an Electronic Nose*. Kluwer Academic Publishers, London, 1992.

[32] W. Göpel, K-D. Schierbaum, S. Vaihinger, and U. Weimar. Fine tuning of electron- and ion-conducting materials for sensor arrays. In J.W. Gardner and P.N. Bartlett, editors, *Sensors and Sensory Systems for Electronic Noses*. Kluwer Academic, Dordrecht, 1992.

[33] H.V. Shurmer, J.W. Gardner, and H.T. Chan. The application of discrimination techniques to alcohols and tobaccos using tin-oxide sensors. *Sensors and Actuators*, 18(3-4):361–371, 1989.

[34] W. Schmid, N. Bârsan, and U. Weimar. Sensing of hydrocarbons with tin-oxide sensors: Possible reaction path as revealed by consumption measurements. *Sensors and Actuators B*, 89:232–236, 2003.

[35] J.Klöber, M.Ludwig, and H.A.Schneider. Effects of thickness and additives on thin-film sno_2 gas sensors. *Sensors and Actuators B*, 3:69–74, 1991.

[36] G. Bidan. Electroconducting conjugated polymers: New sensitive matrices to build up chemical or electrochemical sensors. a review. *Sensors and Actuators B*, 6:45–56, 1992.

[37] T.A. Nguyen, J.N. Barisci, A. Partridge, and G.G Wallace. Investigation of conducting polymer material for sensor arrays. *Synthetic Materials*, 137:1445–1446, 2003.

[38] B. Dufour, P. Rannou, D. Djurado, M. Zagorska, I. Kulszewicz-Bajer, and A. Pron. The role of chain and dopant engineering in the preparation of processible conducting polymers with desired properties. *Synthetic Metals*, 135-136:63–68, 2003.

[39] M. Ando, S. Murakami, and H. Naito. Charge carrier transport properties of poly(9,9-diocytlfluorene) thin films. *Synthetic Metals*, 135-136:285–286, 2003.

[40] D.E. Tallman, C.K. Vang, M.P. Dewald, G.G. Walace, and G.P. Bierwagen. Electron transfer mediated deposisiton of conducting polymers on active metals. *Synthetic Metals*, 135-136:33–34, 2003.

[41] M-A. De Paoli and W.A. Gazotti. Conductive polymer blends: Preparation, properties and applications. *Macromolecular Symposia*, 189:83–103, 2003.

[42] M.C. Lonergan, E.J. Severin, B.J. Doleman, S.A. Beaber, R.H. Grubs, and N.S. Lewis. Array-based vapour sensing using chemically sensitive carbon black-polymer resistors. *Chemistry and Material*, 8:2298–2312, 1996.

[43] Z.C. Feng and C. Chicone. A delay differential model for surface acoustic wave sensors. *Sensors and Actuators A*, 104:171–178, 2003.

[44] Z. Ali, K.Pavey, and E. Robens. Survey on mass determination with oscillating systems: Part III. Acoustic wave mass sensors for chemical and biological sensing. *Journal of Thermal Analysis and Calorimetry*, 71:31–35, 2003.

[45] M. Penza and G. Cassano. Application of principal component analysis and artificial neural networks to recognize the individual VOCs of methanol/2-propanol in a binary mixture by SAW multi-sensor array. *Sensors and Actuators B*, 89:269–284, 2003.

[46] E.J. Severin, R.D. Sanner, B.J. Doleman, S.A. Beaber, R.H. Grubs, and N.S. Lewis. Differential detection of enatiometric gaseous analytes using carbon black-chiral polymer composite, chemically sensitive resisitors. *Analytical Chemistry*, 70:1440–1443, 1998.

[47] M.C. Burl, B.J. Doleman, A. Schaffer, and N.S. Lewis. Assessing the ability to predict human percepts of odor quality from the detector reponses of a conducting polymer composite-based electronic nose. *Sensors and Actuators B*, 72:149–159, 2001.

[48] H.G. Byun K.C. Persaud, S.M. Khaffaf, P.J. Hobbs, and T.H. Misselbrook. Application of unsupervised clustering methods to the assessment of malodour in agriculture using an array of conducting polymer odour sensors. *Computers and Electronics in Agriculture*, 17:233–247, 1997.

[49] J.E. Haugen, O. Tomic, F. Lundby, K. Kvaal, E. Strand, L. Svela, and K. Jorgensen. Analysis of off-flavours in raw cow's milk with a commercial gas-sensor system. In J.W. Gardner and K.C. Persaud, editors, *Electronic Noses and Olfaction 2000*. Institute of Physics Publishing, 2000.

[50] G. Ólafsdóttir, Á. Högnadóttir, E. Martinsdóttir, and H. Jónsdóttir. Application of an electronic nose to predict total volatile bases in capelin (mallotus villosus) for fishmeal production. *Journal of Agricultural and Food Chemistry*, 48:2353–2359, 2000.

[51] S.H. Hahn, M. Frank, and U. Weimar. Rancidity investigation on olive oil: a comparison of multiple headspace analysis using an electronic nose and GC/MS. In J.W Gardner and K.C. Persaud, editors, *Electronic Noses and Olfaction 2000.* Institute of Physics Publishing, 2000.

[52] C. Di Natale, A. Macagno, R. Paolesse, E. Tarizzo, A. Mantini, and A. D'Amico. Human skin odour analysis by means of an electronic nose. *Sensors and Actuators B*, 65:216–219, 2000.

[53] J.W. Gardner, H.W. Shin, and E.L. Hines. An electronic nose system to diagnose illness. *Sensors and Actuators B*, 70:19–24, 2000.

[54] J.W. Gardner, H.W. Shin, E.L. Hines, and C.S. Dow. An electronic nose system for monitoring the quality of potable water. *Sensors and Actuators B*, 69:336–341, 2000.

[55] C. Delpha, M. Siadat, and M. Lumbreras. An electronic nose using time reduced modelling parameters for a reliable discrimination of forane 134a. *Sensors and Actuators B*, 77(1-2):517–524, 2001.

[56] D. Dumitrescu, B. Lazzerini, and F. Marcelloni. A fuzzy hierarchial classification system for olfactory. *Pattern Analysis and Applications*, 3(4):325–334, 2000.

[57] J.A. Covington, J.W. Gardner, C. Toh, P.N. Bartlett, D. Briand, and N.F. de Rooij. Characterisation of an electrodeposited conducting polymer FET array for vapour and odour sensing. In J.W. Gardner and K.C. Persaud, editors, *Electronic Noses and Olfaction 2000.* Institute of Physics Publishing, 2000.

[58] C.Pinheiro, T.Scäfer, C.M. Rodrigues, A.Barros, S.Rocha, I.Delgadillo, and J.G. Crespo. Integrating pervaporation with electronic nose for monitoring the muscatel aroma production. In J.W. Gardner and K.C. Persaud, editors, *Electronic Noses and Olfaction 2000.* Institute of Physics Publishing, 2000.

[59] F. Bender, N. Barie, G. Romoudis, A. Voigt, and A. Rapp. Development of a preconcentration unit for a SAW sensor micro array and its use for indoor air quality monitoring. *Sensors and Actuators B*, 93(1-3):135–141, 2003.

[60] C.R. Tamanaha, L.J. Whitman, and R.J. Colton. Hybrid macro-micro fluidics system for a chip-based biosensor. *Journal of Micromechanics and Microengineering*, 12(3):347, 2002.

[61] J.W. Gardner. A non-linear diffusion-reaction model of electrical conduction in semiconductor gas sensors. *Sensors and Actuators B*, 1:166–170, 1990.

[62] T. Bohlin. *Interactive System Identification: Prospects and Pitfalls.* Springer-Verlag, Berlin, 1991.

[63] M. Casdagli. Nonlinear prediction of chaotic time series. *Physica D*, 35:335–356, 1989.

[64] P. Davis. *Interpolation and Approximation.* Blaisdale, New York, 1963.

[65] M. Gevers and G. Li. *Parameterizations in Control, Estimation and Filtering Problems.* Springer-Verlag. Berlin, 1993.

[66] L. Lyung. *System Identification: Theory for the user.* Prentice Hall Information System Science Series, New Jersey, 1987.

[67] T. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, 14:326–334, 1965.

[68] S. Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall, New Jersey, 1999.

[69] A.J. Smola, B. Schölkopf, and K-R. Müller. The connection between regularization operators and support vector kernels. *Neural Networks*, 11:637–649, 1998.

[70] V Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, 1995.

[71] M.A. Aizerman, E.M. Braverman, and L.I. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

[72] I.J. Leontaritis and S.A Billings. Input-output parametric models for non-linear systems. Part I: Deterministic non-linear systems. *International Journal of Control*, 41(2):303–328, 1985.

[73] S. Chen and S.A. Billings. Representations of non-linear systems: The NARMAX model. *International Journal of Control*, 49(3):1013–1032, 1989.

[74] D.E. Rumelhart and eds J.L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol1.* MIT press, Cambridge MA, 1986.

[75] T. Bachinger and Carl-Frederik Mandenius. Searching for process information in the aroma of cell cultures. *Trends in Biotechnology*, 18(12):494–500, 2000.

[76] A. Ortega, S. Marco, T. Sundic, and J. Samitier. New pattern recognition systems designed for electronic noses. *SENSOR ACTUAT B*, 69(3):302–307, 2000.

[77] V.N. Vapnik. *Statistical Learning Theory.* Wiley InterScience, New York, 1998.

[78] A.K. Pavlou, N. Magan, D. Sharp, J. Brown, H. Barr, and A.P.F. Turner. An intelligent rapid odour recognition model in discrimination of *Helicobacter Pylori* and other gastroesphageal isolates in vitro. *Biosensors and Bioelectronics*, 15:333–342, 2000.

[79] M.O. Balaban, F. Korel, A.Z. Odabasi, and G. Folkes. Transportability of data between electronic noses: Mathematical methods. *Sensors and Actuators B*, 71(3):203–211, 2000.

[80] P.L. Bartlett. The sample complexity of pattern classification with neural networks: The size of the weights is more important then the size of the network. *IEEE Transactions on Information Theory*, 44(2):525–536, 1998.

[81] H.-E. Endres, W. Göttler, H.D. Jander, S. Drost, G. Sberveglieri, G. Faglia, and C. Perego. A systematic investigation on the use of time-dependent sensor signals in signal-processing techniques. *Sensors and Actuators B*, 24-25:785–789, 1995.

[82] X. Vilanova, E. Llobet, R. Alcubilla, J.E. Sueiras, and X. Correig. Analysis of the conductance transient in thick-film tin oxide gas sensors. *Sensors and Actuators B*, 31:175–180, 1996.

[83] J.W. Gardner. A diffusion-reaction model of electrical conduction in tin oxide gas sensors. *Semiconductor Science and Technology*, 4:345–350, 1989.

[84] J.W. Gardner, P.N. Bartlett, and K.F.E. Pratt. Modelling of gas-sensitive conducting polymer devices. *IEE Proceedings: Circuits, Devices and Systems*, 142(5):321–333, 1995.

[85] P.N. Bartlett and J.W. Gardner. Diffusion and binding of molecules to sites within homogenous thin films. *The Philosophical Transactions of the Royal Society of London*, 354:35–57, 1996.

[86] J.W. Gardner, E.L. Hines, F. Molinier, P.N. Bartlett, and T.T. Mottram. Prediction of health of dairy cattle from breath samples using neural network with parametric model of dynamic response of array of semiconducting gas sensors. *IEE Proceedings: Science, Measurement and Technology*, 146(2):102–106, 1999.

[87] K.R. Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge.* Routledge and Kegan Paul, London and Henley, 1968.

[88] S. Geman, E. Bienestock, and Rene Dourstat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58, 1992.

[89] M.J.D. Powell. Radial basis functions for multivariable interpolation: A review. In J.C. Mason and M.G. Cox ., editors, *IMA Conference on Algotithms for the Approximation of Functons and Data.* IMA publishing, 1987.

[90] B. Kosko. *Fuzzy Engineering.* Prentice Hall, International Edition, 1997.

[91] S.V. Chakravarthy and J. Ghosh. Scale-based clustering using the radial basis function network. *IEEE Transactions on Neural Networks*, 7(5):1250–1261, 1996.

[92] F. Camastra and A. Vinciarelli. Estimating the intrinsic dimension of data with a fractal-based method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1404–1407, 2002.

[93] K.-C. Li and K. Shedden. Identification of shared components in large ensembles of time series using dimension reduction. *Journal of the American Statistical Association*, 97:759–765, 2002.

[94] C.A. Nicolaou, S.Y. Tamura, B.P. Kelley, S.I. Bassett, and R.F. Nutt. Analysis of large screening data sets adaptively grown phylogenetic-like trees. *Journal of chemical information and computer sciences*, 42:1069–1079, 2002.

[95] M. Hörnquist, J. Hertz, and M. Wahde. Effective dimensionality of large-scale expression data using principle components analysis. *Biosystems*, 65:147–156, 2002.

[96] F. Camastra and A. Vinciarelli. Intrinsic dimension estimation of data: An approach based on grassberger-procaccia's algorithm. *Neural Processing Letters*, 14:27–34, 2001.

[97] K.J. Falconer. *Fractal Geometry: Mathematical Foundations and Applications.* Wiley, 1990.

[98] P-C. Lo and W-P Chung. An efficient method for quantifying the multichannel EEG spatial-temporal complexity. *IEEE Transactions on Biomedical Engineering*, 48(3):394–397, 2001.

[99] P-C. Lo and W-P. Chung. An approach to quantifying the multi-channel EEG spatial-temporal feature. *Biometrical Journal*, 42(7):901–916, 2000.

[100] A. Potapov and M.K. Ali. Neural networks for estimating intrinsic dimension. *Physical Review E*, 65(4):046212, 2002.

[101] P. Boilot, E.L Hines, M.A. Gongora, and R.S. Folland. Electronic noses inter-comparison, data fusion and sensor selection in discrimination of standard fruit solutions. *Sensors and Actuators B*, 88(1):80–88, 2003.

[102] A.N. Chaudry, T.M. Hawkins, and P.J. Travers. A method for selecting an optimum sensor array. *Sensors and Actuators B*, 69:236–242, 2000.

[103] J.W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 18(5):401–409, 1969.

[104] A. Albert. *Regression and the Moore-Penrose Pseudoinverse.* Academic Press, London, 1972.

[105] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in FORTRAN, Second Edition.* Cambridge University Press, Cambridge, U.K., 1992.

[106] É. Walter and L. Pronzato. *Identification of Parametric Models.* Springer- Verlag. Berlin, 1997.

[107] I. Cha and S. Kassam. Time-series prediction using adaptive radial basis function networks. In *27th Annual Conference on Information Science and Systems*, pages 818–823. Johns Hopkins University, 1993.

[108] K.K.Shukla A.K.Srivastava and S.K.Srivastava. Exploring neuro-genetic processing of electronic nose data. *Microelectronics Journal*, 29:921–931, 1998.

[109] L. Kuncheva. Initializing of an RBF network by a genetic algorithm. *Neurocomputing*, 14:273–288, 1997.

[110] B.Burdsall and C.Giraud-Carrier. GA-RBF:a self-optimising RBF network. In *Third international conference of artificial neural networks and genetic algorithms*, pages 346–349. SPringer-Verlag, 1997.

[111] T.F. Coleman and Y. Li. A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, 6(4):1040–1058, 1996.

[112] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* Wiley Interscience, New York., 1991.

[113] C.S. Wallace and D.M. Boulton. An information measure for classification. *Journal of Computing*, 11:185–195, 1968.

[114] J. Rissanen. Modelling by shortest data description. *Automatica*, 14:456–471, 1978.

[115] J.R. Fitzgerald, D.E. Sturdevant, S.M. Mackie, S.R. Gill, and J.M. Musser. Evolutionary genomics of *Staphylococcus aureus*: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proceedings of the National Academy of Sciences of the United States of America*, 98:8821–8826, 2001.

[116] C Cucarella, M.A. Tormo, C. Ubeda, M.P. Trotonda, M. Monzon, C. Peris, B. Amorena, I. Lasa, and J.R. Penades. Role of biofilm-associated protein bap in the pathogenesis of bovine staphylococcus aureus. *Infection and Immunity*, 72(4):2177–2185, 2004.

[117] R. Esteves De Matos, D.J. Mason, C.S. Dow, and J.W. Gardner. Investigation of the growth characteristics of *e.coli* using headspace analysis. In J.W. Gardner and K.C. Persaud, editors, *Olfaction and Electronic Noses 2000.* IOP Publishing Ltd, 2000.

[118] W.M. Wong, S.K. Lam, H.H. Xia, V.S. Tang, K.C. Lai, W.H. Hu, C.K. Chan, K.L. Cheung, and B.C. Wong. Accuracy of a new near patient test for the diagnosis of *Helicobacter pylori* infection in chinese. *Journal of Gastroenterology and Hepatology*, 17:1272–1277, 2002.

[119] C.K. Eapen, S. Sugathan, M. Kuriakose, T. Abdoel, and K.L. Smits. Evaluation of the clinical utility of a rapid blood test for human leptospiros. *Diagnostic Microbiology and Infectious Disease*, 42:221–222, 2002.

[120] R.N. Kore, C.S. Dow, and K.M. Desai. A new automated system for urine analysis: A simple, cost-effective and reliable method for distinguishing between glomerular and nonglomerular sources of haematuria. *British Journal of Urology*, 84(4):454–460, 1999.

[121] R. De Maesschalck, D. Jouan-Rimbaud, and D.L. Massart. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50:1–18, 2000.

[122] Wei-Chien Chang. On using principal components before separating a mixture of two multivariate normal distributions. *Applied Statistics*, 32(3):267–275, 1983.

[123] S.V. Chakravarthy and J. Ghosh. Scale based clustering using the radial basis function network. *IEEE Transactions on Neural Networks*, 7(5):1250–1261, 1996.

[124] L. Oukhellou and P. Aknin. Hybrid training of radial basis function networks in a partitioning context of classification. *Neurocomputing*, 28:165–175, 1999.

[125] P-E Liss, O. Aspevall, D. Karlsson, and U. Forsum. Terms used to describe urinary tract infections- the importance of conceptual clarification. *APMIS*, 111:291–299, 2003.

[126] J.-L. Baudel, M. Alzieu, F. Barbutt, J. Guglielminotti, E. Maury, B. Guidet, , and G. Offenstadt. A monthly systematic bacteriological report improves accuracy of identification of hospital-acquired infections in an ICU. *Journal of Hospital Infection*, 53:14–17, 2002.

[127] D. Raveh, B. Rudensky, M. Huerta, Y. Aviv, and A.M. Yinnon. Use of time-trend analysis in the design of empirical antimicrobial treatment of urinary tract infection. *European Journal of Clincical Microbiology and Infectious Disease*, 22:158–164, 2003.

[128] E.J. Kuijper, J. van der Meer, M.D. de Jong, P. Speelman, and J. Dankert. Usefulness of gram stain for diagnosis of lower respiratory tract infection or urinary tract infection and as an aid in guiding treatment. *European Journal of Clincical Microbiology and Infectious Disease*, 22:228–234, 2003.

[129] S. Canbaz, Y. Peksen, A.T. Sunter, H. Leblebicioglu, and M. Sunbul. Antibiotic prescribing and urinary tract infection. *International Journal of Antimicrobial Agents*, 20:407–411, 2002.

[130] T.A. Russo and J.R. Johnson. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: Focus on an increasingly important endemic problem. *Microbes and Infection*, 5:449–456, 2003.

[131] T.M. Hooton and W.E. Stamm. Diagnosis and treatment of uncomplicated urinary tract infection. *Infectious Disease Clinics of North America*, 11(3):551–581, 1997.

[132] Cyrano Sciences. *Practical Guide for the Cyranose 320.* Cyrano Sciences, 2000.

[133] Paul J. Flory. *Principles of Polymer Chemistry.* Cornell University Press, New York, 1953.

[134] E.K. Sichel. *Carbon black-polymer composites: the physics of electrically conducting composite materials.* Dekker, New York, 1982.

[135] M. Omastová, J. Prokeš, S. Košina, and D. Hlavatá. Stability of electrical properties of conducting polymer composites. *Macromolecular Symposia*, 170:241–248, 2001.

[136] B. Doleman, M. Lonergan, E. Severin, T. Vaid, and N. Lewis. Quantitative study of the resolving power of arrays of carbon black-polymer composites in various vapor-sensing tasks. *Analytical Chemistry*, 70:4177–4190, 1998.

[137] E.J Severin, B.J. Doleman, and Nathan S. Lewis. An investigation of the concentration dependence and response to analyte mixtures of carbon black/ insulating organic polymer composite vapor detectors. *Analytical Chemistry*, 72:658–668, 2000.

[138] B. Doleman, R. Sanner, E. Severin, R. Grubbs, and N. Lewis. Use of compatible polymer blends to fabricate arrays of carbon black-polymer composite vapor detectors. *Analytical Chemistry*, 70:2560–2564, 1998.

[139] Y. Mamunya. Polymer blends filled with carbon black: Structure and electrical properties. *Macromolecular Symposium*, 170:257–264, 2001.

[140] M. Haug, K.D. Schierbaum, G. Gauglitz, and W. Göpel. Chemical sensors based upon polysiloxanes:comparison between optical, quartz microbalance, calorimetric, and capacitance sensors. *Sensors and Actuators B*, 11:383–391, 1993.

[141] S.J. Patrash and E.T. Zellers. Characterization of polymeric surface acoustic wave sensor coatings and semiempirical models of sensor responses to organic vapors. *Analytical Chemistry*, 65:2055–2066, 1993.

[142] S.P. Russell and D.H. Weinkauf. Vapour sorption in plasma polymerized vinyl acetate and methyl methacrylate thin films. *Polymer*, 42:2827–2836, 2001.

[143] A. Reynier, P. Dole, S. Humbel, and A. Feigenbaum. Diffusion coefficients of additives in polymers. I. Correlation with geometric parameters. *Journal of Applied Polymer Science*, 82:2422–2433, 2001.

[144] A. Reynier, P. Dole, and A. Feigenbaum. Additive diffusion coefficients in polyolefins. II. Effect of swelling and temperature on the d=f(m) correlation. *Journal of Applied Polymer Science*, 82:2434–2443, 2001.

[145] A Windie. *Polymer Permeability.* Elsevier, Amsterdam, 1984.

[146] M.A Parker and D. Vesely. Temperature and time-dependence of diffusion of PCL into PVC. *Journal of Polymer Science, Part B: Polymer Physics*, 8:1869–1878, 1986.

[147] Editor P. Neogi. *Diffusion in Polymers.* Marcel Dekker, Inc. New York, 1996.

[148] D. Vesely. Molecular sorption mechanism of solvent diffusion in polymers. *Polymer*, 42:4417–4422, 2001.

[149] A. Vergara, L. Paduano, G. Mangiapia, and R. Sartorio. Diffusion coefficient matrix in nonionic polymer-solvent mixtures. *Journal of Physical Chemistry*, 105:11044–11051, 2001.

[150] B. Amsden. Modelling solute diffusion in aqueous polymer solutions. *Polymer*, 43(1623-1630):285–286, 2001.

[151] S.E. Wood and R. Battino. *Thermodynamics of Chemical Systems.* Cambridge University Press, Cambridge, 1990.

[152] Y.C. Lee, S.Y. Lin, and H.S. Liu. Role of equation of state on studying surfactant adsorption kinetics. *Langmuir*, 17:6196–6202, 2001.

[153] B. Lundberg and B. Sundqvist. Resistivity of a composite conducting polymer as a function of temperature, pressure, and environment: Applications as a pressure and gas concentration transducer. *Journal of Applied Physics*, 60(3):1074–1079, 1986.

[154] D. Stauffer and A. Aharony. *Introduction to Percolation Theory: Second Edition.* Taylor and Francis, 1985.

[155] H. Böttger and V.V. Bryksin. *Hopping Conduction in Solids.* Akademie-Verlag, Berlin, 1985.

[156] P. Talik, M. Zabkowska-Waclawek, and W. Waclawek. Sensing properties of the CB-PCV composites for chlorinated hydrocarbon vapours. *Journal of Material Science*, 27:6807–6810, 1992.

[157] G.R. Ruschau, R.E. Newnham, J. Runt, and B.E. Smith. 0-3 ceramic/polymer composite chemical sensors. *Sensors and Actuators*, 20:269–275, 1989.

[158] H. Zois, L. Apekis, and M. Omastová. Electrical properties of carbon black-filled polymer composites. *Macromolecular Symposium*, 170:249, 2001.

[159] J. Crank. *The Mathematics of Diffusion.* Clarendon Press, Oxford, 1975.

[160] H.S. Carslaw and J.C. Jaeger. *Conduction of Heat in Solids.* Clarendon Press, Oxford, 1959.

[161] T. Ficker. Non-isotherm steady-state diffusion within glaser's condensation model. *International Journal of Heat and Mass Transfer*, 46:5175–5182, 2003.

[162] M.A. Islam, H. Buschatz, and D. Paul. Non-equilibrium surface reactions - a factor in determining steady state diffusion flux. *Journal of Membrane Science*, 204:370–384, 2002.

[163] A. Hill. *Proceedings of the Royal Society*, 104, 1928.

[164] J.W. Gardner. Electrical conduction in solid-state gas sensors. *Sensors and Actuators*, 18(3-4):373–387, 1989.

[165] A.B. Littlewood. *Gas Chromatography.* Academic Press, New York, 1970.

[166] R. Buchhold, A. Nakladal, G. Gerlach, K. Sahre, and K.-J. Eichhorn. Mechanical stress in micromachined components caused by humidity-induced in-plane expansion of thin polymer films. *Thin Solid Films*, 312:232–239, 1998.

[167] G. Searle. *Dynamic Modelling of Electronic Nose Systems.* PhD thesis, University of Warwick, 2002.

[168] F.M. White. *Viscous Fluid Flow, 2nd Edition.* McGraw-Hill, 1991.

[169] R.K. Shah and A.L. London. *Laminar Flow Forced Convection in Ducts.* Academic Press, New York, 1978.

[170] J.A. Covington, J.W. Gardner, D. Briand, and N.F. de Rooij. A polymer gate FET sensor array for detecting organic vapour. *Sensors and Actuators B-Chemical*, 77(1-2):155–162, 2001.

[171] J.V. Hatfield, J.A. Covington, and J.W. Gardner. GasFETs incorporating conducting polymers as gate materials. *Sensors and Actuators B-Chemical*, 65(1-3):253–256, 2000.

[172] S. Marsili-Libelli, S. Guerrizio, and N. Checchi. Confidence regions of estimated parameters for ecological systems. *Ecological Modelling*, 165:127–146, 2003.

[173] C. Arnold, M. Harms, and J. Goschnick. Air quality monitoring and fire detection with the Karlsruhe electronic micronose KAMINA. *IEEE Sensors*, 2(3):179–188, 2002.

[174] Z. Ankara, T. Kammerer, A. Gramm, and A. Schutze. Low power virtual sensor array based on a micromachined gas sensor for fast discrimination between H-2, CO and relative humidity. *Sensors and Actuators B*, 100(1-2):240–245, 2004.

[175] Clyde R. Metz. *Schaum's Outline Series: Physical Chemistry.* McGraw-Hill Book Company, 1976.

[176] G. Barton. *Elements of Greens's Functions and Propagation.* Oxford Science Publications, 1989.

[177] S. Kasargod, F. Adib, and P. Neogi. Diffusion in inhomogeneous polymer membranes. *Journal of Chemical Physics*, 103(16):7114–7119, 1995.

[178] M. M. Abbott and H. G. Van Ness. *Schaum's Outlines: Thermodynamics with Chemical Applications.* McGraw-Hill Book Company, 1989.