Computational Optimization of Chemical Vapor Detector Arrays

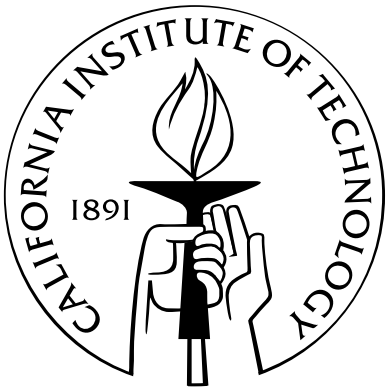
Thesis by

Brian Sisk

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2005

(Submitted February 14, 2005)

This work is dedicated to my wife Amanda. We all know I don't deserve you - thanks for having the grace to not point it out.

Acknowledgements

Thanks to the Lewis group for constant support, especially Alan Hopkins, Shawn Briglin, and Mike Koscho who were most helpful for getting my started on the project. Thanks also to my advisor, Nate Lewis, for pushing and directing me when I needed it, and letting me do things my way when it occasionally worked. I'd like to thank my collaborators, especially Alan Hopkins for collecting the data used as the foundation for Chapter 4. Thanks to Tom Vaid and Mike Burl for beginning a study that became my first publication at Caltech and formed Chapter 2, as well as Matteo Pardo (University of Brescia) with whom I worked on a study that became Chapter 5. Many thanks to Mario Blanco of the Goddard group who introduced me to Cerius^2 and other useful software packages.

Of a personal nature, thanks to everyone who helped me accomplish what I have, beginning of course with my parents, who always encouraged me to view things with a skeptical eye. Thanks also to my undergraduate/Masters' advisor Wei-Ping Pan as well as John Riley at Western Kentucky University for giving me my start in research. My greatest debt is due of course to my wife, Amanda, for always being there these four years, especially at the end when I know I was quite irritable. And thanks to her parents for letting me take her 2000 miles away.

Abstract

Arrays of broadly responsive, chemically sensitive detectors have been used for many years as a means of detecting a wide array of vapors. These systems have been used in fields ranging from analysis of wines and coffees to land mine and nerve agent detection to disease diagnosis. Despite their successes, these systems have been plagued by problems, namely a lack of sensor diversity, miniscule libraries of previously-recognized analytes, significant sensor drift, and weak signal processing capabilities compared to the mammalian olfactory system.

This work details progress toward the alleviation of those problems with regard to arrays of polymer/carbon-black composite chemiresistor detectors developed at Caltech. Specifically, it has been determined that larger sensor arrays allow the suitable recognition of more analytes, and a greater chance of successful discrimination between a given analyte others to which it is similar. Additionally, new classes of percolative, low carbon-black sensors have been developed that yield far higher sensitivities and stronger responses than traditional sensors, as well as responses that are exponential with concentration. Such sensors allow for recognition of analytes using lower precision electronics than was previously realizable. A method for calibrating the system with few analyte exposures has also been developed from an analysis of the correlations between sensor/analyte response changes with time over groups of analytes and sensors.

Further work has allowed algorithmic optimizations to assign functional group identities and certain physiochemical information such as molar volume and octanol/water partition coefficients to analytes that are completely unknown to the system, using a model built upon other known analytes. Additionally, a comparison of linear and nonlinear classifiers is performed to identify data characteristics that might be more suited to linear classifiers such as Fisher's Linear Discriminant or nonlinear ones such as Artificial Neural Networks.

These improvements to chemical vapor detector arrays and the processing of their data allow the extraction of more useful information and the minimization of time spent training and calibrating the system. By constructing more appropriate sensor arrays, establishing non-comprehensive but extensive analyte response libraries, choosing useful algorithmic classifiers, and performing timely and mimimal calibration, the utility of detector systems can be maximized while minimizing maintenance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 History

Detection of chemical vapors has been performed analytically since the development of early vapor-specific sensors for the selective detection of light gases such as water vapor or carbon monoxide. Such sensors have been made for a number of different vapors, are often of very low cost, but have the obvious disadvantage of detecting only one analyte. Thus, one would need in principle one sensor for each analyte of interest. Clearly, for situations in which one would prefer to analyze a larger number of analytes, the number of sensors required to do so would grow very quickly.

Alternatively, one can use a finite number of sensors, each of which responds to a wide variety of analytes. While each sensor in such a broadly-responsive system shows some measurable response to practically all analytes, the sensor array as a whole shows a slightly different response to each analyte it encounters. The original method, as detailed by Persaud and Dodd,[1] employs a series of sintered, polycrystalline semiconductor metal oxide detectors. Oxygen in the air oxidizes the semiconductor material and remains at the surface, forming various oxygen species (0^-, 0_2^-, 0^{2-}). This process is reversible, and reaches an equilibrium state, at which point the semiconductor is effectively p-type doped relative to the same semiconductor in a vacuum, resulting in a greater deal of band-bending at the surface of the semiconductor particles, and thus reduced resistance.[2] The introduction of reducing analyte vapors (such as hydrocarbons) results in consumption of the surface oxide species, changing the equilibrium at the surface, and a resulting increase in resistance.

Later methods have typically used a series of thin polymer films to absorb vapor, which is then transduced by any of a variety of methods. This approach has been used by a number of groups starting in the early 1980s.[1,3–6] Transduction methods used have included sensitive quartz crystal resonator, and since then methods that measure optical changes,[7–10] resistance changes,[5,11] and mass changes[3,12,13] in sensing materials have been developed (among others). In these cases, the sensing medium is typically a polymer, and as such all of these systems effectively measure the partition coeffecient for a given polymer/vapor combination for each detector in an array. This

partition coefficient is measured indirectly by the specific measurement scheme chosen, whether it be changes in optical absorption upon analyte uptake of a dye impregnated in the material[9] or a change in deflection of a microcantilever upon a sorption-based mass change.[13]

1.2 Implementation of Chemical Vapor Detection System at Caltech

Work in the Lewis group at Caltech has focused on conductive composites of carbon-black and polymers.[14,15] The carbon-black, which is conductive, allows current to pass across the sensor enabling resistance measurements to be made. Because the polymeric component expands when it absorbs vapor, the carbon-black particles necessarily grow farther apart. As such, the resistance of the composite increases upon vapor exposure. This change is measured as $\Delta R_{eq}/R_b$, where ΔR_{eq} represents the equilibrium resistance change upon exposure to vapor, and R_b indicates the baseline resistance before exposure.[14,16] The $\Delta R_{eq}/R_b$ metric has been shown to be linear with concentration and mass uptake over a wide range of vapor concentrations,[17] and is fairly consistent over different carbon-black loadings in the composite.[14] Analysis of the response data from such systems can be accomplished with any standard multivariate tool; among those used most frequently in our laboratory are Principal Components Analysis (PCA), Linear Discriminant Analysis (LDA), and Artificial Neural Networks (ANNs).[18–20] Each of these methods is used to connect an unknown, measured data cluster to information that has been previously collected by the detector array. These techniques of sensor fabrication and data analysis are described at length in later chapters in this thesis.

1.3 Requirements for Use of Chemical Vapor Detection Systems

Using systems like that developed at Caltech, or any of the other similar systems mentioned, allows for the detection and recognition of any analyte, assuming a number of restrictive criteria are met. First, a detector array must contain at least a few sensors that are sensitive to any analyte of interest, and selective enough to distinguish between it and all other analytes.[16,21] Additionally, these sensors must be stable enough to elicit consistent responses from the detector array during both training phases and later use. Finally, one need have trained the detector array on all potential analytes that the detector array might encounter, and employ a pattern-classification algorithm that is suitable for the analyte signatures that will be generated. For example, some classifiers work well for binary separations of data that is largely multivariate Gaussian (for example, LDA), while others work well for clustering many groups of highly nonlinear data (for example, ANN's).[18]

1.4 Limitations of Broadly-responsive Detector Array Systems

These conditions dramatically reduce the usability of any sensor array system, particularly for situations in which one does not know what analytes might be encountered "in the field." Without this knowledge, it is not feasible to tailor a detector array with a suitable number of task-appropriate detectors, and impossible to train the detector array on literally every analyte. Even if it were possible to do so, retraining would be necessary occasionally due to drift in detector responses, requiring the entire exercise to be frequently repeated. Finally, assuming one could attain all of this, choosing the wrong classification algorithm could result in extremely poor results, either because the algorithm lacks sufficient plasticity to model nonlinear data, or because a highly plastic model was overtrained on an insufficient amount of training data.

This situation prevents the widespread adoption of portable, broadly-responsive detector array systems, and has restricted their applications to more mundane uses in which the user knows approximately what to search for. As such, these systems are used less as the "electronic noses" they were designed to be, but more like the earlier single-analyte detectors, with the exception that the "single analyte" can vary. Rarely, however, have they been used in situations in which the user does not know what analyte may be encountered, because of the sensor selection and training required.

1.5 Outline of This Thesis

Methods are needed to determine appropriate detector array compositions and perform some degree of analyte identification with little or no *a priori* knowledge of the set of unknown analytes and concentrations that may be encountered. Additionally, some means of calibrating the device experimentally is necessary to reduce the overall training burden. This work seeks to address each of these burdens in turn. In this work, it is determined whether the size of the detector array can be reduced, or alternatively whether having a larger detector array can overcome problems encountered by lacking *a priori* knowledge of analyte identity. It will also be demonstrated that a wealth of information pertaining to analyte physiochemical characteristics, including functionality, can be assigned to an unknown analyte that has never been encountered by the detector array. Additionally, the development of a new, low carbon-black detector will be detailed that shows extremely high sensitivities at low concentration, to better distinguish among low-concentrations of similar analytes. Next, a method of recalibrating the detector array system using a small number of calibrant exposures is described. Finally, linear and nonlinear classification algorithms will be evaluated and compared on a wide variety of datasets. Ultimately, the results of this work should increase the usability of arrays of broadly-responsive chemical vapor detectors.

Bibliography

[1] Persaud, K.; Dodd, G. *Nature* **1982,** *299,* 352–355.

[2] Albert, K.; Lewis, N.; Schauer, C.; Sotzing, G.; Stitzel, S.; Vaid, T.; Walt, D. *Chemical Reviews* **2000,** *100,* 2595–2626.

[3] Ballantine, D.; Rose, S.; Grate, J.; Wohltjen, H. *Analytical Chemistry* **1986,** *58,* 3058–3066.

[4] Rose-Pehrsson, S.; Grate, J.; Ballantine, D.; Jurs, P. *Analytical Chemistry* **1988,** *60,* 2801–2811.

[5] Bartlett, P.; Archer, P.; Ling-Chung, S. *Sensors and Actuators* **1989,** *19,* 125–140.

[6] Shurmer, H.; Gardner, J.; P., C. *Sensors and Actuators B - Chemical* **1990,** *1,* 256–260.

[7] Ronot, C.; Archenault, M.; Gagnaire, H.; Goure, J.; Jaffrezicrenault, N.; Pichery, T. *Sensors and Actuators B - Chemical* **1993,** *11,* 375–381.

[8] White, J.; Kauer, J.; Dickinson, T.; Walt, D. *Analytical Chemistry* **1996,** *68,* 2191–2202.

[9] Dickinson, T.; White, J.; Kauer, J.; Walt, D. *Nature* **1996,** *382,* 697–700.

[10] Rakow, N.; Suslick, K. *Nature* **2000,** *406,* 710–713.

[11] Freund, M.; Lewis, N. *Proceedings of the National Academy of Sciences of the United States of America* **1995,** *92,* 2652–2656.

[12] Patrash, S.; Zellers, E. *Analytical Chemistry* **1993,** *65,* 2055–2066.

[13] Lang, H.; Baller, M.; Berger, R.; Gerber, C.; Gimzewski, J.; Battiston, F.; Fornaro, P.; Ramseyer, J.; Meyer, E.; Guntherodt, H. *Analytica Chimica Acta* **1999,** *393,* 59–65.

[14] Lonergan, M.; Severin, E.; Doleman, B.; Beaber, S.; Grubb, R.; Lewis, N. *Chemistry of Materials* **1996,** *8,* 2298–2312.

[15] Koscho, M.; Grubbs, R.; Lewis, N. *Analytical Chemistry* **2002,** *74,* 1307–1315.

[16] Doleman, B.; Lonergan, M.; Severin, E. *Analytical Chemistry* **1998,** *70,* 4177-4190.

[17] Severin, E.; Doleman, B.; Lewis, N. *Analytical Chemistry* **2000,** *72,* 658–668.

[18] Duda, R.; Hart, P. *Pattern Classification and Scene Analysis;* John Wiley and Sons: New York, 1984.

[19] Jurs, P.; Bakken, G.; McClelland, H. *Chemical Reviews* **2000,** *100,* 2649–2678.

[20] Vaid, T.; Burl, M.; Lewis, N. *Analytical Chemistry* **2001,** *73,* 321–331.

[21] Park, J.; Groves, W.; Zellers, E. *Analytical Chemistry* **1999,** *71,* 3877–3886.

Chapter 2

Classification Performance of Polymer/Carbon-Black Composite Vapor Detector Arrays as a Function of Array Size and Detector Composition

2.1 Abstract

The vapor classification performance of arrays of conducting polymer composite vapor detectors has been evaluated as a function of the number and type of detectors in an array. Quantitative performance comparisons were facilitated by challenging a collection of detector arrays with vapor discrimination tasks that were sufficiently difficult that at least some of the arrays did not exhibit perfect classification ability for all of the tasks of interest. Specific discrimination tasks involved differentiating between low concentration (< 1% of the vapor pressure) exposures to 1-propanol versus 2-propanol, low concentration exposures to n-hexane versus n-heptane, and differentiating between compositionally similar mixtures of closely related analytes, such as 9.37 ppm m-xylene with 10.2 ppm p-xylene versus 7.67 ppm m-xylene with 12.4 ppm p-xylene. A decision boundary was developed using a cross-validated Fisher linear discriminant algorithm on a training set of analyte presentations and the resulting chemometric model was then used to classify a subsequent collection of test analyte presentations to the array being evaluated. In other cases, classification performance was evaluated using the Fisher linear discriminant and a leave-one-out (LOO) cross-validation procedure. For nearly all of the discrimination tasks investigated in this work, classification performance either increased or did not significantly decrease as the number of chemically different detectors in the array increased. Any given subset of the full array of detectors, selected because it yielded the best classification performance at a given array size for one particular task, was invariably outperformed by a different subset of detectors, and by the entire array of 20 chemically diverse detectors when used in at least one other vapor discrimination task.

Arrays of detectors were nevertheless identified that yielded robust discrimination performance between compositionally close mixtures of 1-propanol and 2-propanol, n-hexane and n-heptane, and m-xylene and p-xylene, attesting to the excellent analyte classification performance that can be obtained through the use of such semi-selective vapor detector arrays.

2.2 Introduction

A significant issue in the use of arrays of semi-selective vapor detectors[1,2] is the dependence of analyte classification performance on the number of detectors in the array. Patel et al. have claimed that four conducting polymer composite detectors are sufficient to provide a "universal solvent detector".[3] Park et al. found that for certain analyte classification tasks, various arrays of three to six surface acoustic wave (SAW) detectors provided mutually comparable results, leading these authors to conclude that increasing the number of detectors in an array did not significantly improve classification performance.[4] However, measurements using conducting polymer composite vapor detectors have indicated that the performance in certain vapor classification tasks can improve as the number of different detectors in the array is increased.[5] The relationship between the number of detectors and overall system performance is important because significant engineering tradeoffs are faced for surface acoustic wave devices,[6–10] quartz crystal microbalances,[9,11,12] conducting polymer detectors,[13,14] and dye-impregnated optical beads[15,16] or optical fibers[17,18] as the number of detectors is increased. For example, more SAW detectors require increased power consumption; increased numbers of optical detectors require higher pixel counts in the focal plane of the detecting camera; and increased numbers of chemiresistors require more rapid analog-to-digital convertors and/ or readout circuitry parallelization to maintain the same individual detector data acquisition rate. Increasing the number of detectors in an array also requires more computational power to preprocess, process, and analyze the resulting data stream. The focus of thework reported herein is to address in a quantitative fashion the performance of differently sized arrays of semi-selective vapor detectors for selected vapor detection tasks.

The vapor detectors that we have used for this purpose are chemically sensitive resistors fabricated from composites of conductors and insulating organic polymers.[3,19,20] Sorption of an analyte into these materials produces a swelling of the film that affects the properties of the percolative network of conductive particles in the composite. The swelling produces a change in the dc electrical resistance of the detector that is readily read with a multiplexing dc ohmmeter.[19] The steady-state relative differential resistance response of these chemiresistive detectors has been shown to be related linearly to the concentration of analyte in the vapor phase and resulting mass uptake in the polymer phase over a reasonable range of analyte concentrations and compositions,[21,22] thereby simplifying the data analysis and signal processing requirements on the detector array output. Fur-

thermore, it is relatively easy to fabricate a large chemically diverse array of such detectors in which each detector film has a common conductor, typically carbon-black, but has a different insulating organic polymer supportmatrix, such as polycaprolactone or polyvinylpyrrolidone, as the resistive component of the composite.[20,23]

For data produced by an array of d detectors, with one descriptor per detector (in our case the steady-state relative differential response value of the composite), the response, X, to each analyte presentation can be described as a d-dimensional vector:

$$X = \sum_{i=i}^{d} c_i x_i \tag{2.1}$$

with the coefficient of the i^{th} dimension of X, c_i, equaling the response of the i^{th} detector in the array. At the system level, the analyte classification performance of an array of semiselective vapor detectors is intimately related to the signal processing algorithm used to analyze the data. Many different signal processing algorithms have been used for vapor detector array data, including principal components analysis, knearest neighbor analysis, neural networks, and the visualempirical region-of-influence method (VERI) .[24–28]

In previous work from our laboratory, a statistical approach was used to define a quantitative metric for evaluating the resolving power of a detector array in various vapor detection tasks.[5,23] A statistically defined metric is especially informative when the detector response is a linear function of analyte concentration because in such a case the concentration-normalized analyte patterns do not change with the concentration of analyte. To evaluate the magnitude of this metric, the points in a d-dimensional space are projected orthogonally onto a line, reducing the classification problem from d dimensions to one dimension. When the data are projected onto one dimension, it is desirable to maximize the distance between the means of the two classes being separated, while minimizing their within-class variation. Such a ratio can be expressed as a resolution factor, RF (eqn. 2.2), where d is the distance between the two-class means, and s_1 and s_2 are the standard deviations of the two classes, respectively.[24]

$$RF = \frac{\delta}{\sqrt{\sigma_1^2 + \sigma_1^2}} \tag{2.2}$$

One quantitative metric that has been proposed to standardize vapor detector array performance is a Euclidean distance along the vector defined by the means of the two response clusters of interest divided by the standard deviation of the clusters along this vector.[29] When such a metric is used, noisy detectors can actually produce a decrease in resolution factor when detectors are added to the array, because even though new detectors add to the overall information content of the array

output signal, the additional detectors contribute noise along a direction that is defined by the vector that runs through the mean response points of a specific analyte pair. If the noise along this direction is greater than the increased separation along this direction, then the computed resolution factor will decrease. The use of such a metric thus possesses the drawback that the best line for separating two clusters may not be the line that connects their mean response points. This drawback is mitigated by use of the Fisher linear discriminant analysis method, which instead chooses a projection vector such that the array resolving power is maximized for each analyte pair. Thus, artifacts from new detectors transferring noise to certain directions in detector space are minimized, and detectors which provide significant noise, rather than resolving power, for a given task are not strongly weighted along the optimum vector that is identified by the Fisher linear discriminant methodology. The RF value produced by Fisher's linear discriminant is also an inherent property of the data and is independent of the algorithm that might be used subsequently to assign any individual data point to a class.[24] This resolution factor is basically a multi-dimensional analogue to the separation factors used to quantify the resolving power of a column in gas chromatography, with the Fisher RF value serving as a quantitative indication of how distinct two patterns are from each other considering the optimal direction to analyze the signals and simultaneously taking into account the distribution of responses upon exposure to the analytes that comprise the solvent pair of concern.

In previous work, when the performance of various vapor detector arrays was evaluated by assessing the mean pairwise Fisher resolution factor of the median-performing array of k detectors out of a total of N detectors $({}_NC_k) = 20!/[k!(N\text{-}k)!]$ on a broadly construed set of test vapors, the system performance increased as the number of chemically different detectors increased. Thus, the average Fisher resolution factor between pairs of a broadly construed test set of analytes increased as more detector choices were available to produce the optimally performing vector for resolution of a specific analyte pair.[5] A similar result was obtained when the resolution factor of the worst-resolved pair of analytes in the test set was evaluated.[5]

Although the Fisher RF value provides a reasonable metric to assess the performance of different detector arrays that contain the same total number of detectors,[5] it is not optimal for evaluation of changes in performance with array size, because the Fisher RF value will never decrease with the addition of more detectors, no matter how "noisy".[24] Additionally, the resolution factor is a useful tool for assessing the separation between clusters inherent in the response of the detector array, but the magnitude of the RF value does not directly translate to analyte classification performance unless certain distributional assumptions are satisfied. A good comparative measure of the performance of arrays of varying sizes is the ability to classify an unknown test analyte, and in fact the frequency with which a model correctly classifies analyte exposures can decrease with the addition of noisy detectors. Fisher's linear discriminant was the most accurate classifier for our array for the tasks

examined in a closely related previous study,[23] so it was chosen as the discriminant for this work as well. A simple threshold was set using the assumption that the projected (one-dimensional) distributions for each class are Gaussian, in accordance with expectations for the situation in which the variability in detector response arises from external variation in the state of the system (random fluctuations in temperature, analyte concentration, etc. during data collection) as opposed to detector-related drift. Analysis of the actual population distributions of the detector responses resulting from multiple exposures to a given analyte validated this assumption in that that the observed population distributions were in accord with expectations for a Gaussian distribution, within the limits of agreement expected for relatively small sample populations. A hyperplane orthogonal to w, placed at the proper threshold along the vector, is then used to discriminate between a pair of analytes. The merit of using the Fisher linear discriminant is that it automatically weights detectors according to their signal/ noise-based resolving power for the task of concern, as opposed to forcing a noisy detector to transfer its noise into the performance of the remainder of the system without realizing a concomitant gain in signal/noise performance for analyte resolution.

To compare quantitatively the relative performance of various detector arrays, the collection of arrays must be presented with pairs of analytes that will not be perfectly classified by at least some of the arrays. This was not the case with pairs of single-component organic vapors presented at relatively high concentration, all of which were perfectly (or nearly so) separated from each other, including structural isomers, such as o-xylene and m-xylene.[3,5,20,30] As part of this work, we have challenged a polymer/carbon-black composite detector array with a pair of compounds that are very chemically similar, H_2O and D_2O. In addition, it is useful to consider the classification performance between mixtures of analytes. The responses of the polymer/carbon-black composite detectors are linear with analyte concentration, and the response to a binary mixture of analytes is the response of the pure analytes weighted by the mole fraction of analytes in the mixture.[30] Thus, the Euclidean distance between a binary vapor mixture that is 0.50 mole fraction of each constituent and a binary mixture that is a 0.60:0.40 mole fraction mixture of these same analytes should be approximately one-tenth of the Euclidean distance between the array responses to the individual pure analytes. Several different binary mixtures of 1-propanol and 2-propanol, of n-hexane and n-heptane, and of m-xylene and p-xylene were therefore utilized as part of the present work. Another method to decrease the discriminating ability of a detector array is to decrease the signal-to-noise ratio of the individual detectors. Low concentrations of analytes will decrease the detector signal and therefore reduce the signal-to-noise ratio, broadening the clusters of data in detector space relative to the separation of their mean values. A number of low concentration (<1% of the vapor pressure) exposures to 1-propanol, 2-propanol, n-hexane, and n-heptane were therefore studied and the performance of a variety of detector arrays was also assessed for these specific sensing tasks.

Table 2.1: Polymeric component of carbon-black composite vapor detectors.

Detector	Set A	Set B
1	Poly(ethylene-co-vinyl acetate), 70% vinyl acetate	Polycaprolactone
2	Poly(ethylene oxide)	Poly(ethylene-co-vinyl acetate), 40% vinyl acetate
3	Poly(vinyl pyrrolidone)	Poly(ethylene oxide)
4	1,2-Polybutadiene	Poly(ethylene glycol)
5	Polycaprolactone	Poly(styrene-co-butadiene), ABA block copolymer, 30% styrene
6	Poly(4-vinyl phenol)	Poly(methyloctadecylsiloxane)
7	Poly(vinyl acetate)	Poly(vinyl stearate)
8	Cellulose acetate	Ethyl cellulose
9	Poly(4-vinyl pyridine)	Kraton D-1102
10	Poly(methyl methacrylate)	Kraton G-1652M
11	Poly(styrene-co-maleic anhydride)	Poly(4-vinyl phenol)
12	Poly(vinyl butyral)	Poly(vinyl acetate)
13	Hydroxypropyl cellulose	Poly(vinyl pyrrolidone)
14	Ethyl cellulose	Polycarbonate
15	Poly(ethylene-co-acrylic acid), 86% ethylene content	Polystyrene
16	Poly(methyloctadecylsiloxane)	Polysulfone
17	Poly(ethylene glycol)	Poly(methyl methacrylate)
18	Poly(ethylene-co-vinyl acetate) 18% vinyl acetate	Poly(vinyl butyral)
19	Polystyrene	Hydroxypropyl cellulose
20	Poly(styrene-co-acrylonitrile)	Poly(styrene-co-isoprene), 14% styrene

2.3 Experimental

The acquisition and initial treatment of some of the data analyzed in this paper have been described in a prior article.[23] One detector set used in this collection of experiments has been described previously and is designated herein as detector Set A (Table 2.1).[23] A first data set collected with these detectors consisted of exposures to two chemically similar analytes, H_2O and D_2O, each at $P/P^\circ = 0.050$ where P° is the vapor pressure of the analyte at 21±1°C and P is the partial pressure of the analyte during the exposure period. Another collection of data consisted of exposures to n-hexane, n-heptane, 1-propanol, and 2-propanol, each at $P/P^\circ = 0.010$, 0.0075, 0.0050, and 0.0025. Data were also collected for mixtures of n-hexane/n-heptane and for mixtures of 2-propanol/1-propanol at partial pressure ratios of P/P° = 0.025:0.025, 0.027:0.023, 0.021:0.029, and 0.035:0.015 for each solvent pair of interest, with the first P/P° value corresponding to the first member of the solvent pair listed. Each analyte was exposed to the detectors 140 times for 300 s per exposure, with a separation of 600 s between exposures. The background gas was oil-free laboratory air that contained 1.10±0.15 parts per thousand (ppth) of water vapor.

The initial exposures of each run tended to give responses that varied more than those later in the run. For this reason, the first 40 exposures to each analyte in this data set were rejected. The

Table 2.2: Description of tasks in series I and II data sets.

Task	Analyte 1				Analyte 2			
	Component 1	P/P°	Component 2	P/P°	Component 1	P/P°	Component 2	P/P°
1	H_2O	0.050			D_2O	0.050		
2	1-Propanol	0.010			2-Propanol	0.010		
3	1-Propanol	0.0075			2-Propanol	0.0075		
4	1-Propanol	0.0050			2-Propanol	0.0050		
5	1-Propanol	0.0025			2-Propanol	0.0025		
6	n-Heptane	0.010			n-Hexane	0.010		
7	n-Heptane	0.0075			n-Hexane	0.0075		
8	n-Heptane	0.0050			n-Hexane	0.0050		
9	n-Heptane	0.0025			n-Hexane	0.0025		
10	1-Propanol	0.0025	2-Propanol	0.0025	1-Propanol	0.0023	2-Propanol	0.0027
11	1-Propanol	0.0025	2-Propanol	0.0025	1-Propanol	0.0029	2-Propanol	0.0021
12	1-Propanol	0.0025	2-Propanol	0.0025	1-Propanol	0.00157	2-Propanol	0.0035
13	1-Propanol	0.0023	2-Propanol	0.0027	1-Propanol	0.0029	2-Propanol	0.0021
14	1-Propanol	0.0023	2-Propanol	0.0027	1-Propanol	0.0015	2-Propanol	0.0035
15	1-Propanol	0.0029	2-Propanol	0.0021	1-Propanol	0.0015	2-Propanol	0.0035
16	n-Heptane	0.0025	n-Hexane	0.0025	n-Heptane	0.0023	n-Hexane	0.0027
17	n-Heptane	0.0025	n-Hexane	0.0025	n-Heptane	0.0029	n-Hexane	0.0021
18	n-Heptane	0.0025	n-Hexane	0.0025	n-Heptane	0.0015	n-Hexane	0.0035
19	n-Heptane	0.0023	n-Hexane	0.0027	n-Heptane	0.0029	n-Hexane	0.0021
20	n-Heptane	0.0023	n-Hexane	0.0027	n-Heptane	0.0015	n-Hexane	0.0035
21	n-Heptane	0.0029	n-Hexane	0.0021	n-Heptane	0.0015	n-Hexane	0.0035
22-3	1-Propanol	0.0011	2-Propanol	0.00090	1-Propanol	0.00090	2-Propanol	0.0011
24-5	n-Heptane	0.0011	n-Hexane	0.00090	n-Heptane	0.00090	n-Hexane	0.0011
26-7	m-Xylene	0.0011	p-Xylene	0.00090	m-Xylene	0.00090	p-Xylene	0.0011
28	1-Propanol	0.0075			2-Propanol	0.0083		
29	1-Propanol	0.075			2-Propanol	0.083		
30	Benzene	0.0065			Toluene	0.00567		
31	Benzene	0.065			Toluene	0.056		

remaining 100 exposures to each analyte, which were mutually similar within a run, were used for analysis of this data set. Each task was run separately, resulting in all of the exposures from task 1 being run before any of the exposures of task 2, for example. Within a single task, the analytes were presented in random order over the 200 exposures (100 to each analyte) of the task. Randomization was performed to minimize any effects of detector drift that might artificially aid in discrimination between analytes. Table 2.2 presents a summary of the 21 tasks (1-21) that are collectively denoted as the Series I data runs.

In a separate set of experiments, binary mixtures of nhexane/ n-heptane, 1-propanol/2-propanol, and m-xylene/pxylene were exposed to a different array of conducting polymer composite detectors. The detector array used in these runs, denoted as detector Set B, consisted of 80 detectors that were housed in two separate chambers. The 80 detectors were formed from four nominally identical copies of each of 20 compositionally distinct detector materials (Table 2.1). Detectors were fabricated using procedures described previously.[5,9,21,30] All copies of each detector type were formed in the

same detector fabrication run for each particular detector composition. The detector copies were placed sequentially along the direction of flow, with one full detector array placed after the other. Each array consisted of 20 compositionally different detectors that were placed in the chamber in the order listed in Table 2.1. Except where otherwise noted, only data from the first set of 20 compositionally different detectors were analyzed as detector Set B data.

In these runs, designated collectively as the Series II data runs, the fractional partial pressures for the first six runs (tasks 22-27 in Table 2.2) were set at $P/P^\circ = 0.011$ for one component of an analyte mixture and $P/P^\circ = 0.0090$ for the other component. The partial pressures of the analytes were permuted to generate the second analyte mixture that formed each discrimination task (Table 2.2). Each analyte was exposed 200 times to the detector array, with each individual exposure consisting of 70 s of exposure to laboratory air, 80 s of exposure to the analyte of interest, and then 60 s of exposure to laboratory air. In the last four runs of the Series II data (tasks 28-31 in Table 2.2), two pairs of analytes (1-propanol or 2-propanol, followed by benzene or toluene, respectively) were exposed separately to the detectors at relatively high and relatively low vapor phase concentrations. The specific analyte partial pressure values used in the low analyte concentration runs were chosen to be small enough that each analyte generated a strong response on approximately five detectors, while producing a less discernable signal on the other detectors. The partial pressures used in the high concentration runs were then set to be a factor of ten greater than those used in the low concentration exposures (Table 2.2). In these runs, each analyte was exposed 200 times to a single set of 20 detectors, with each exposure consisting of 40 s of flowing background laboratory air (having a water vapor concentration of 1.10±0.15 ppth), 70 s of analyte, then followed by 40 s of laboratory air. As with the Series I data runs, P° values were calculated for a laboratory temperature of 21±1°C. All exposures of an analyte pair that comprised a particular classification task were run consecutively, but the analyte type was randomized within the 400 vapor exposures that comprised a particular vapor classification task. Values of P° used to calculate analyte concentrations for hexane, heptane, 1-propanol, 2-propanol, benzene, toluene, m-xylene, p-xylene, H_2O and D_2O were 0.192, 0.0446, 0.0209, 0.0565, 0.102, 0.0304, 0.00852, 0.0113, 0.0295, and 0.0269 atmospheres, respectively.

For each exposure in the Series I Data Set, the baseline resistance, R_b, of each detector was calculated from the average of the resistance readings for the 60 s immediately prior to the beginning of the analyte exposure. The equilibrium response, R_{eq}, was calculated from the average of the resistance readings for the last 60 s of the exposure to analyte vapor. For runs in the Series II Data Set, which used shorter exposure times, R_b was determined from the average of the resistance readings in the last 30 s immediately prior to each exposure, and R_eq was determined from data obtained in the last 30 s of each analyte vapor exposure. Baseline correction of Series II data was performed by fitting a separate regression line to the pre-exposure resistance readings for each analyte exposure, and correcting all subsequent data points by the difference in the value of the

regression fit at the time of the measurement of that data point and at t=0. The quantity used in analysis of both Series I and II data was the steady-state relative differential resistance change, $\Delta R_{eq}/R_b$, where $\Delta R_{eq} = R_{eq} - R_b$. Data were converted to $\Delta R_{eq}/R_b$ form in Microsoft Excel, Labview, and original C++ code, while all subsequent data analysis was performed in Matlab using original Matlab code. The $\Delta R_{eq}/R_b$ data were evaluated in unnormalized form in the analysis discussed herein. The signal-to-noise (S/N) ratio for each detector at the analyte concentration of interest was computed by determining the ratio between the ΔR_{eq} value determined from a single exposure to the analyte of interest and three times the standard deviation of the corrected baseline resistance.[31,32]

The classification performance, $\mathcal{P}$, of the detector arrays was evaluated using different methods for the different data series. In all cases, once a Fisher discriminant decision boundary was determined and detector weights were set for a particular task and array, each test data point for that task and detector array was assigned to the class member of the pair-wise discrimination task that resided on the same side of the decision boundary as the test data point. Because only 100 exposures to each analyte were available in the Series I Data Set, the array classification performance from these data was evaluated using the leave-one-out (LOO) cross-validation procedure. This process provides an approach in which both training set data and test set data can be obtained from a data set in which the classes of all members are known.[24] In the LOO approach, one member is removed from a data set of n members, the classifier is developed using the $n - 1$ remaining members of the data set, and the resulting classifier is used to assign the withheld member to a class. A second member of the data set is then withheld, a new classifier is constructed from the $n - 1$ remaining members, and this classifier is then used to assign that member to a class. The procedure is repeated through all n members of the data set. The fraction of correct assignments out of the n possible cases provides an evaluation of the LOO classification performance $\mathcal{P}_l$ of a given array towards a specific pair of analytes utilizing a given discriminant algorithm.

For the Series II data, 200 exposures were available for each analyte. This larger data set facilitated reliable use of independently constructed test and training sets to evaluate the classification performance of the various arrays of interest. First, 100 of the 200 total exposures to each analyte were randomly selected to form the training set for that analyte. The remaining exposures formed the test set for that particular analyte. The training set was used to formulate a Fisher linear discriminant decision boundary (a hyperplane orthogonal to the Fisher discriminant vector) which was then evaluated with respect to analyte classification performance on the exposures in the test set. The classification performance on the training set is denoted as $\mathcal{P}_{trn}$ and that on the test set is denoted as $\mathcal{P}_{tst}$. The capability of the model with respect to a given training set was determined by the Fisher RF value; the capability with regard to a test set was determined by the fraction of the exposures that were correctly classified using the decision boundary developed on the training set. The Series II data were also analyzed separately using the LOO procedure to facilitate direct

comparison with the classification performance resulting from analysis of the Series I data.

A Beowulf cluster, comprised of 64 Pentium III machines, was used for the more processor-intensive analyses, particularly for the comparison of the performance of 20-detector arrays with all possible combinations of arrays having fewer than 20 detectors on all of the analyte classification tasks. A Pentium III single-processor machine was used for the more routine calculations, such as identifying which five-detector sets performed the best for a specific analyte classification task.

2.4 Results

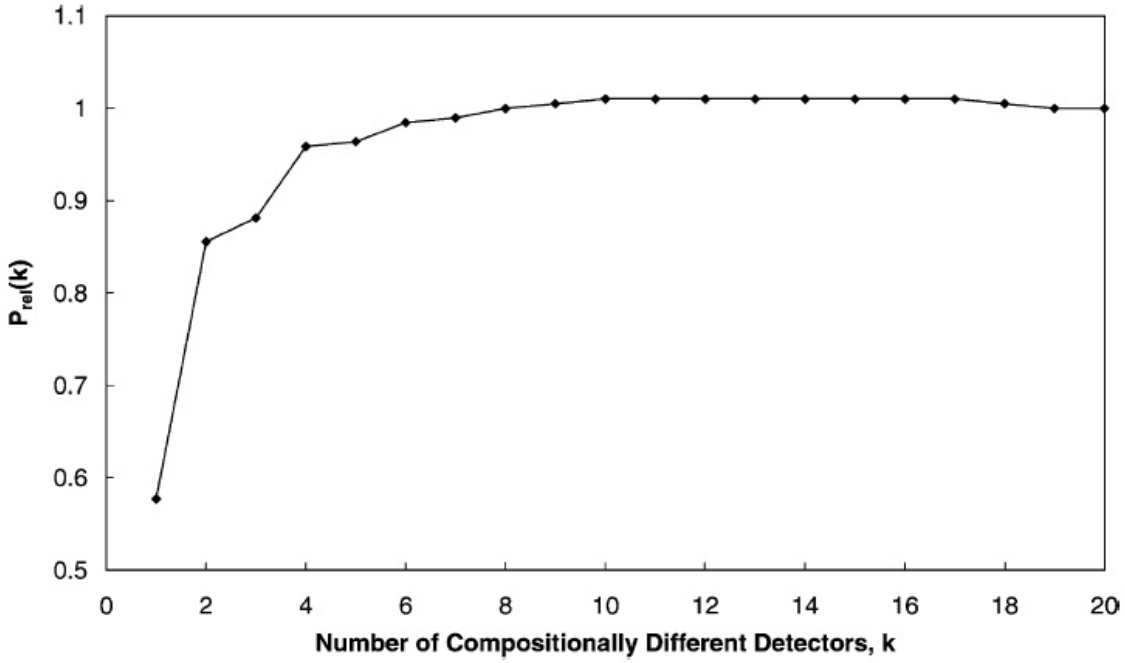
2.4.1 Classification of N-hexane/N-heptane and 1-Propanol/2-Propanol Mixtures as a Function of the Number of Different Detectors in the Array

Figure 2.1 displays the classification performance for detector Set A discriminating a mixture of n-hexane at P/P°=0.025 and n-heptane at P/P°=0.025 from a mixture of n-hexane at P/P°=0.027 and n-heptane at $P/P\circ$=0.023 (task 16,Table 2.2). The classification performance for this task was calculated using the LOO cross-validation procedure on the entire set of 100 Series I exposures to each analyte mixture. The full 20-detector Set A array yielded 78% correct classification for the 200 exposures of this discrimination task. This value is denoted as $\mathcal{P}_l\{[\Omega(20) \Rightarrow 16\}$ where the argument of Ω indicates the number of unique detectors in the array, Ω, and the $\Rightarrow$ 16 notation indicates that the LOO classification performance, $\mathcal{P}_l$, of this array was evaluated for classification task 16 in Table 2.2. For each value of k in the range 1< k <20, where k is the number of compositionally different detectors in the array, an exhaustive search of all possible 20!/[k!(20-k!]) k-detector combinations from the 20-detector Set A array was performed to identify the array having k detectors that had the best LOO classification performance for the task of concern. During this process, new detector weights and a new decision boundary were determined for every k-detector combination evaluated. The best-performing array, denoted $[\Omega^{16}_{max}(k)]$ for each value of k, had a classification performance $\mathcal{P}_l\{[\Omega^{16}_{max} \Rightarrow 16\} = \mathcal{P}^{16}_{l,max}(k)$. Fig. 2.1 depicts the relative classification performance, $\mathcal{P}^{16}_{rel}(k)$, of these optimally performing array subsets of detector Set A as a function of array size, with the relative classification performance of each selected subset representing the classification performance of $[\Omega^{16}_{max}(k)]$ normalized relative to that of the full 20-detector array (Eq. 2.3):

$$\mathcal{P}^{16}_{rel}(k) = \frac{\mathcal{P}^{16}_{rel}\{[\Omega^{16}_{max}(k)] \Rightarrow 16\}}{\mathcal{P}^{16}_{rel}\{[\Omega(20)] \Rightarrow 16\}} = \frac{\mathcal{P}^{16}_{l,max}(k)}{\mathcal{P}^{16}_{rel}\{[\Omega(20)] \Rightarrow 16\}} \tag{2.3}$$

Table 2.3a lists the best-performing detector sets for each value of k in the range 1< k <10 as well as the LOO classification performance of these detector sets for this particular analyte separation

Figure 2.1: Classification performance vs. array size for distinguishing a mixture of n-hexane at P=P0 (r) 0:025 and n-heptane at P/P°=0.025 from a mixture of n-hexane at P/P°=0.027 and n-heptane at P/P°=0.023. The classification performance was calculated using the LOO cross-validation procedure on the entire set of 100 exposures to each analyte mixture. The full Set A detector array, consisting of 20 compositionally different detectors, had a 78% correct classification performance for this task. For each value of k in the range 1< k <20 where k is the number of compositionally different detectors in the array, an exhaustive search of all possible k-detector combinations from the 20-detector Set A array was performed to identify the array having k detectors that had the best LOO classification performance for the task of concern. The classification performance for this task of the best-performing k-member detector set is plotted relative to that of the full 20-detector array.



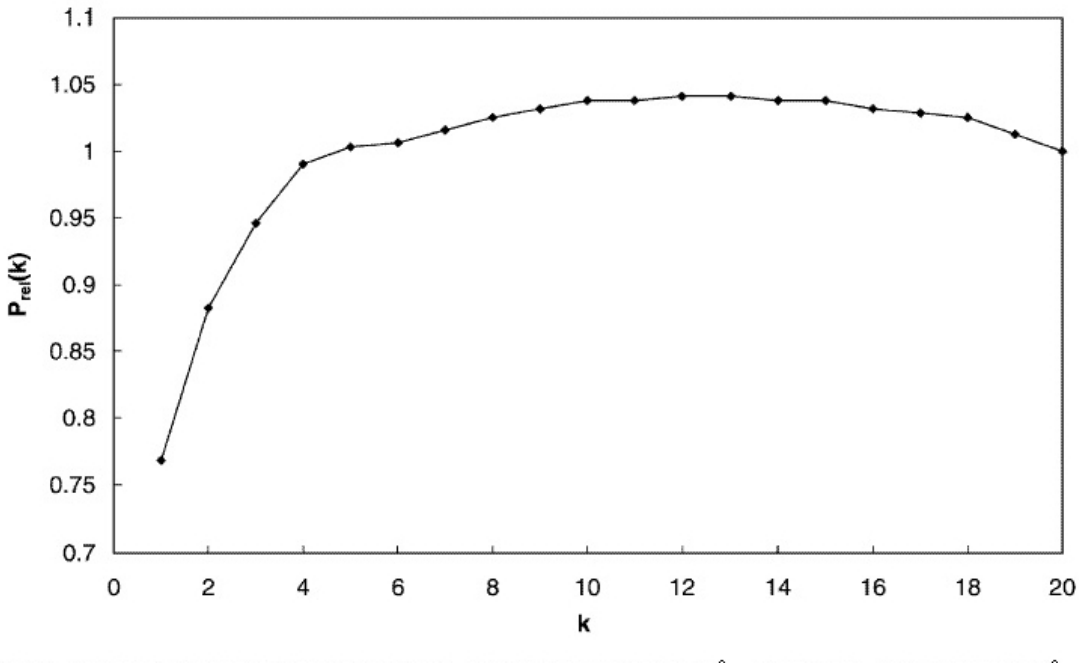


Figure 2.2: Classification performance vs. array size for distinguishing a mixture of n-hexane at $P/P°$=0.0090 from n-heptane at $P/P°$=0.011 from a mixture of n-hexane at $P/P°$=0.011 and n-heptane at $P/P°$=0.0090. The classification performance was calculated using the LOO cross-validation procedure on the entire set of 200 exposures to each analyte mixture. The first set of 20 compositionally different detectors of the Set B detector array had a 95% correct classification performance in the test set data for this task. For each value of k in the range 1< k <20, an exhaustive search of all possible k-detector combinations from the 20-detector Set B array was performed to identify the array having k detectors that had the best LOO classification performance for the task of concern. The classification performance for this task of the best-performing k-member detector set is plotted for 1< k <20 relative to that of the full 20-detector array.

task.

Figure 2.2 displays analogous data for detector Set B, in which a mixture of n-hexane at $P/P°$=0.011 and n-heptane at $P/P°$=0.0090 is separated from a mixture of n-hexane at $P/P°$=0.0090 and n-heptane at $P/P°$=0.011 (task 24, Table 2.2). The classification performance $\mathcal{P}_l\{[\Omega(20)] \Rightarrow 24\}$ was evaluated for the full 20-detector Set B using the LOO procedure (to aid in comparison with Data Set I results), and this array yielded 95% correct classification for this task. An exhaustive search was then performed to identify the subset of detectors of a given array size, designated as $[\Omega^{24}_{l,max}(k)]$ that yielded the best LOO cross-validation performance for each value of k. The classification performance, $\mathcal{P}^{24}_{l,max}(k)$, normalized relative to the performance of the full 20-detector array for this task is plotted for 1< k <20. Table 2.3b lists the best-performing detector sets for each value of k in the range 1< k <10 as well as the LOO classification performance of these detector sets for this particular analyte separation task.

Table 2.3: Detectors that provided maximum classification performance at various array sizes.

	Array size	Detectors										Performance
(a)For task 16												
	1	9										0.5700
	2	12	13									0.6200
	3	9	13	14								0.6700
	4	2	11	12	13							0.7000
	5	1	2	8	11	13						0.7300
	6	1	2	8	11	13	17					0.7650
	7	1	2	3	8	11	13	18				0.7850
	8	2	9	11	12	13	14	15	17			0.7950
	9	2	3	9	11	12	13	14	15	17		0.7950
	10	1	2	3	7	11	12	13	16	17	18	0.8100
(b) For task 24												
	1	7										0.6375
	2	19	20									0.8750
	3	8	19	20								0.9175
	4	8	14	19	20							0.9225
	5	8	11	15	19	20						0.9350
	6	1	7	8	15	19	20					0.9375
	7	1	2	7	8	15	19	20				0.9400
	8	1	2	6	7	8	15	19	20			0.9450
	9	1	2	3	6	7	8	15	19	20		0.9475
	10	1	2	3	7	8	9	12	15	19	20	0.9500
(c) For task 10												
	1	19										0.5600
	2	3	9									0.8300
	3	3	9	19								0.8550
	4	3	6	14	18							0.9300
	5	3	6	14	15	18						0.9350
	6	3	9	11	12	14	18					0.9550
	7	2	3	9	11	12	14	18				0.9600
	8	3	6	8	9	11	12	14	18			0.9700
	9	3	6	9	11	12	14	17	18	19		0.9750
	10	3	6	8	9	11	12	14	17	18	20	0.9800
(d) For task 22												
	1	3										0.6050
	3	8	18	19								0.7450
	4	8	11	18	19							0.7800
	5	8	10	11	18	19						0.7900
	6	2	8	10	11	18	19					0.7925
	7	1	2	7	8	12	18	19				0.8000
	8	1	2	6	8	11	14	18	19			0.8075
	9	1	2	6	8	10	11	15	18	19		0.8125
	10	1	2	6	8	9	11	12	15	18	19	0.8175

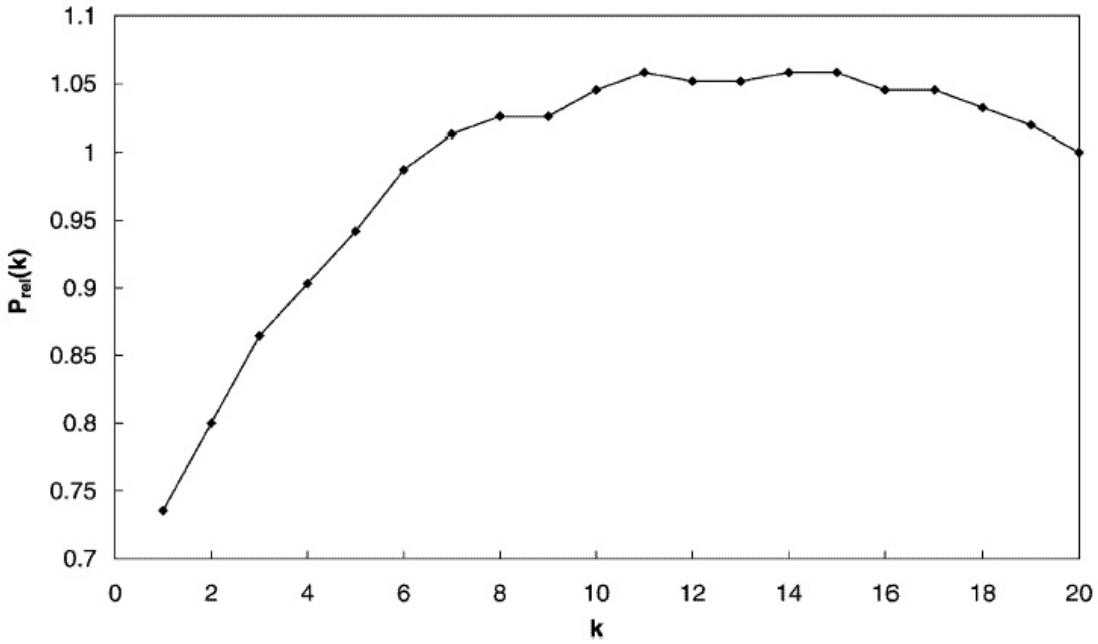


Figure 2.3: Classification performance vs. array size for distinguishing between a mixture of 1-propanol at $P/P°$=0.025 and 2-propanol at $P/P°$=0.025 and another mixture of 1-propanol at $P/P°$=0.023 and 2-propanol at $P/P°$=0.027. The classification performance was calculated using the LOO cross-validation procedure on the entire set of 100 exposures to each analyte mixture. The full Set A detector array, consisting of 20 compositionally different detectors, had a 97% correct classification performance in the test set data for this task. For each value of k in the range $1< k <20$, an exhaustive search of all possible k-detector combinations from the 20-detector Set A array was performed to identify the array having k detectors that had the best LOO classification performance for the task of concern. The classification performance for this task of the best-performing k-member detector set is plotted for $1 < k < 20$ relative to that of the full 20-detector array.

2.4.2 Classification of N-hexane/N-heptane and 1-Propanol/2-Propanol Mixtures as a Function of the Number of Different Detectors in the Array

Figures 2.3 and 2.4 and Table 2.3c and d display analogous data for classification of mixtures of 1-propanol and 2-propanol (tasks 10 and 22, respectively). Detector Set A was challenged to discriminate a mixture of 1-propanol at $P/P°$=0.025 and 2-propanol at $P/P°$=0.025 from a mixture of 1-propanol at $P/P°$=0.023 and 2-propanol at $P/P°$=0.027, while detector Set B was challenged to discriminate a mixture of 1-propanol at $P/P°$=0.011 and 2-propanol at $P/P°$=0.0090 from a mixture of 1-propanol at $P/P°$=0.0090 and 2-propanol at $P/P°$=0.011.

Figures 2.1-2.4 clearly indicate that for these detector arrays, the classification performance for an individual task either increased or did not decrease significantly as the number of chemically different detectors increased. Similar trends were observed for essentially all of the other tasks evaluated in this work, with the exception of the xylenes separation tasks (tasks 25 and 26 in

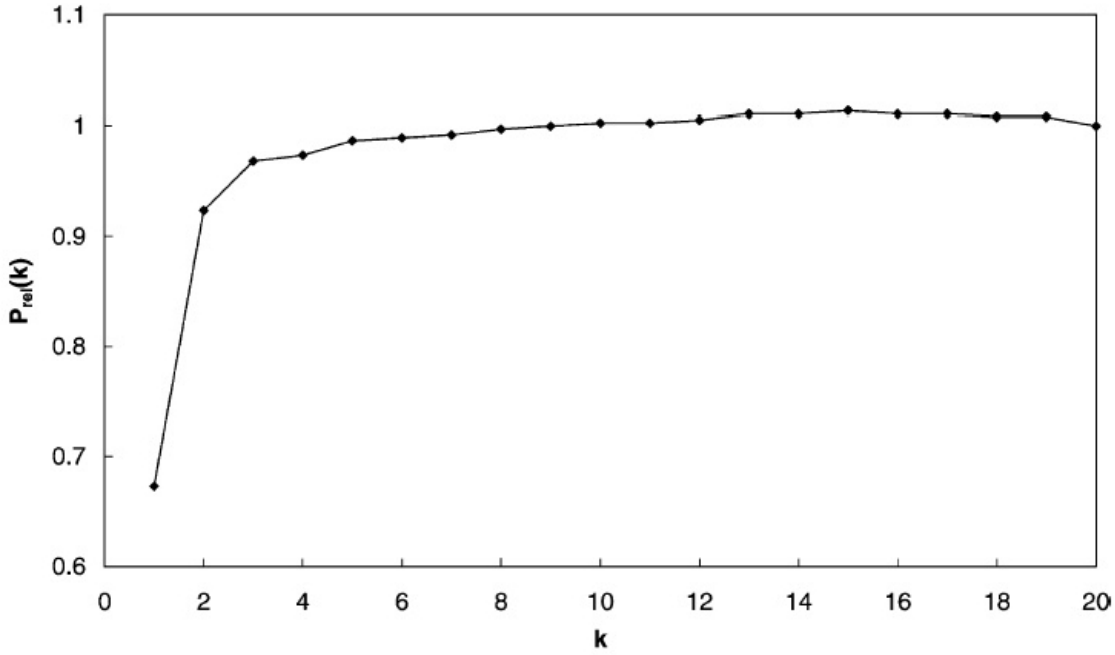


Figure 2.4: Classification performance vs. array size for distinguishing between a mixture of 1-propanol at $P/P°$=0.0090 and 2-propanol at $P/P°$=0.011 and another mixture of 1-propanol at $P/P°$=0.011 and 2-propanol at $P/P°$=0.0090. The classification performance was calculated using the LOO cross-validation procedure on the entire set of 200 exposures to each analyte mixture. The first set of 20 compositionally different detectors of the Set B detector array had a 79% correct classification performance in the test set data for this task. For each value of k in the range $1< k <20$, an exhaustive search of all possible k-detector combinations from the 20-detector Set B array was performed to identify the array having k detectors that had the best LOO classification performance for the task of concern. The classification performance for this task of the best-performing k-member detector set is plotted for $1< k <20$ relative to that of the full 20-detector array.

Table 2.2) of Data Set II.

2.4.3 Relative Classification Performance of "Optimal" Five-detector Arrays in Vapor Discrimination Tasks

Figures 2.1-2.4 also indicate that nearly optimal classification performance could be obtained in many cases by selecting subsets of 4-10 detectors, and then only using response data from that subset of detectors to classify presentations of that particular pair of analytes to the detector array. However, as shown in Table 3, the collection of detectors in the array of a given size that produced the best classification performance for the hexane/heptane mixture classification task was not the same as the collection of detectors that produced the best classification performance, at the same array size, for the 1-propanol/2-propanol mixture classification task. To assess the generality of this observation, an extensive comparison was performed of the classification performance of selected detector array subsets in a diverse collection of different analyte classification tasks.

For specificity, this analysis concentrated on evaluating the performance of arrays that consisted of five compositionally different detectors obtained from the full collection of 20 Set B detectors. The performance of such five-detector arrays was evaluated for the first six separate classification tasks in the Series II Data Set (tasks 22-27, Table 2.2). For each five-detector subset, a Fisher discriminant boundary was developed on the training set of a given task, yielding a Fisher RF value for each subset. For each specific classification task, an exhaustive search was performed through all 20!/[(15!)(5!)] possible five-detector subsets in the Set B detector array to identify the five-detector combination that maximized the Fisher RF value resulting from the training data set for the task of interest. These five-detector arrays are designated as $[\Omega^{J}_{trn,max}(5)]$ where the subscript *trn* indicates that the arrays were selected based on their abilities to maximize the RF values of the training set data, and the index J designates the task for which the array was identified as providing the maximum Fisher RF value.

The classification performance of each $[\Omega^{J}_{trn,max}(5)]$ array was then evaluated on an independent test set of exposures that included six analyte classification tasks (tasks 22-27)of the Series II data. In this process, for each combination of a five-detector array $[\Omega^{J}_{trn,max}(5)]$ and an individual pair-wise separation task K of the Series II data, new detector weights and a new Fisher discriminant boundary were calculated based on the 200 analyte exposures that comprised the training set for task K presented to the detector set $[\Omega^{J}_{trn,max}(5)]$. The classification performance $\mathcal{P}_{tst}\{[\Omega^{J}_{trn,max}(5)] \Rightarrow K\}$ was then evaluated by determining the fraction of a subsequent, separate 200 exposure task K test data set that were on the correct side of the Fisher discriminant boundary. The process was repeated for all J and K in the range $22< J <27$ and $22< K < 27$.

The resulting detector performance data are summarized in the two 6×6 matrices of Table 4. The

entries in each column designate the performance, $\mathcal{P}_{tst}\{[\Omega^{J}_{trn,max}(5)] \Rightarrow K\}$, of each $[\Omega^{J}_{trn,max}(5)]$array when used in each of the six tasks (22< K <27) evaluated in the Series II data run. Two matrices are presented, one of which displays the absolute classification performance of the detector sets, $\mathcal{P}_{tst}[\Omega^{J}_{trn,max}(5)] \Rightarrow K\}$, 22< J <27, 22< K <27, and the other of which displays the classification performance of the detector sets for each task K when normalized to the value of $\mathcal{P}_{tst}\{[\Omega^{J}_{trn,max}(5)] \Rightarrow K\}$ obtained when $K = J$. In other words, for each test set task K, the test set classification performance of the various five-detector arrays $[\Omega^{J}_{trn}(5)]$ was normalized relative to the test set classification performance of the array, $[\Omega^{J}_{trn,max}(5)]$, that yielded the maximum training set Fisher RF value for that particular analyte classification task. The normalization accounts for inherent differences in the difficulties of the various classification tasks being evaluated. Furthermore, because the task is a two-class classification problem, a classification performance value of 0.5 is equal to that of random chance, so normalization was performed using the formula:

$$\frac{\mathcal{P}_{tst}\{[\Omega^{J}_{trn,max}(5) \Rightarrow K\} - 0.5}{\mathcal{P}_{tst}\{[\Omega^{J}_{trn,max}(5) \Rightarrow J\} - 0.5} \tag{2.4}$$

The entries along the diagonal of the normalized matrix are all 1.0, because by definition they are the values to which the test set classification performance for a given task have been normalized.

The data indicate that different detector subsets yielded the best classification performance for different tasks. An equivalent statement is that the combination of five detectors which produced the best classification performance for one task was significantly outperformed by another five-detector set in another, different task. For example, the combination of five detectors, $[\Omega^{27}_{trn,max}(5)]$, that produced the best Fisher RF value in a training set run of the pair of m-xylene/pxylene mixtures yielded only 34-38% of the normalized classification performance on 1-propanol/2-propanol test set mixtures (tasks 22, 23) relative to the 1-propanol/2-propanol mixture test set classification performance (tasks 22, 23) of the five-detector combination, $[\Omega^{22}_{trn,max}(5)]$ or $[\Omega^{23}_{trn,max}(5)]$, that produced the best Fisher RF value for each 1-propanol/2-propanol mixture discrimination task (Table 2.4). Similarly, the five-detector set that yielded the best Fisher RF value for the other training set of m-xylene/pxylene exposures, $[\Omega^{26}_{trn,max}(5)]$, only yielded 22-40% of the classification performance on 1-propanol/2-propanol test set mixtures (tasks 22, 23) relative to the 1-propanol/2-propanol mixture test set performance of the five-detector set $[\Omega^{22}_{trn,max}(5)]$ or $[\Omega^{23}_{trn,max}(5)]$ that yielded the best Fisher RF value on the training sets of these 1-propanol/2-propanol mixture classification tasks. Similarly, the five-detector sets that were found to yield the best Fisher RF value for either of the n-hexane/n-heptane training set mixture tasks, $[\Omega^{24}_{trn,max}(5)]$ and $[\Omega^{25}_{trn,max}(5)]$, yielded much worse test set classification performance in the 1-propanol/2-propanol classification tasks (22, 23) than was produced when either of the five-detector sets $[\Omega^{22}_{trn,max}(5)]$ and $[\Omega^{23}_{trn,max}(5)]$ that produced the best

Fisher RF value on the 1-propanol/2- propanol training set tasks (22, 23) were used on the test set of 1-propanol/2-propanol mixture tasks (22, 23).

Good classification performance in the m-xylene/pxylene test set mixture classification tasks was observed for many (although not all) of the five-detector collections that yielded the best Fisher RF training set values for any of the other binary mixture classification tasks. However, the five-detector combinations that produced the best Fisher RF training set values for the m-xylene/p-xylene tasks yielded inferior performance for the 1-propanol/2-propanol and n-hexane/n-heptane test set mixture classification tasks relative to detector collections that yielded the best Fisher RF training set values for those specific tasks (Table 4). As might be expected, five-detector sets that yielded the best training set Fisher RF values for any given task yielded approximately the same test set classification performance for that task as they did for a separate test set run that represented a nominally identical task. For example, $\mathcal{P}_{tst}\{[\Omega^{24}_{trn,max}(5)] \Rightarrow 24\} \approx \mathcal{P}_{tst}\{[\Omega^{25}_{trn,max}(5)] \Rightarrow 24\}$

2.4.4 Classification Performance as a Function of the Number of Compositionally Different Detectors in the Array

Given that different k-detector subsets were observed to produce the best classification performance for different tasks, and that classification performance in general either increased or did not decrease significantly as the number of chemically different detectors increased (Figures 1-4), it is of interest to compare the performance of subsets of detectors to the performance of a full array of 20 compositionally different detectors. This comparison was performed for all of the different tasks and data sets collected during the course of this work.

First, the LOO classification performance of the array of 20 chemically different detectors in detector Set A was evaluated for each of the 21 tasks for which these detectors were used (Table 2.2, tasks 1-21). This returned a set of performance values $\mathcal{P}_l\{[\Omega(20)] \Rightarrow J\}$ for each task ($1< J <21$). The LOO classification performance for every individual combination of k-detectors, where $1< k <20$, for each of the 21 tasks was then also evaluated. For every task, the classification performance for each individual combination of k-detectors,$\mathcal{P}_l\{[\Omega_i(k)] \Rightarrow J\}$ for $1\leq i \leq [20!/k!(20-k)!]$, was then compared to that of the full 20-detector array, $\mathcal{P}_l\{[\Omega(20)] \Rightarrow J\}$ for that same task. For each task, the classification performance for any k-detector array, $\mathcal{P}_l\{[\Omega_i(k)] \Rightarrow J\}$, was then normalized by dividing by the classification performance of the full 20-detector Set A array on that same task, $\mathcal{P}_l\{[\Omega(20)] \Rightarrow J\}$. These performance ratios were tabulated and used to create a function $g(k)$ for which, by definition, no combination of k detectors does strictly better than $g(k)$ relative to the full 20- detector Set A array on all 21 tasks of the Series I data run. Therefore, when an array containing k detectors is used to perform a set of tasks, at least one task among the set will yield a classification performance no better than $g(k)$ relative to performance of the full 20-detector array, regardless

Table 2.4: Classification performance of arrays formed from five compositionally different detectors.

Array	Detectors	Task					
		22	23	24	25	26	27
Absolute Performance							
$[\Omega^{22}_{trn,max}(5)]$	8, 9, 11, 18, 19	0.735	0.755	0.755	0.815	0.850	0.880
$[\Omega^{23}_{trn,max}(5)]$	8, 11, 13, 18, 19	0.760	0.765	0.755	0.795	0.810	0.865
$[\Omega^{24}_{trn,max}(5)]$	1, 8, 15, 19, 20	0.620	0.560	0.930	0.930	0.760	0.850
$[\Omega^{25}_{trn,max}(5)]$	1, 3, 4, 19, 20	0.670	0.565	0.890	0.925	0.655	0.695
$[\Omega^{26}_{trn,max}(5)]$	4, 8, 9, 12, 18	0.595	0.560	0.705	0.655	0.865	0.835
$[\Omega^{27}_{trn,max}(5)]$	1, 8, 9, 12, 19	0.590	0.600	0.705	0.740	0.815	0.880
Normalized Performance[a]							
$[\Omega^{22}_{trn,max}(5)]$		1.00	0.962	0.593	0.741	0.959	1.00
$[\Omega^{23}_{trn,max}(5)]$		1.11	1.00	0.593	0.694	0.849	0.961
$[\Omega^{24}_{trn,max}(5)]$		0.511	0.226	1.00	1.01	0.712	1.01
$[\Omega^{25}_{trn,max}(5)]$		0.723	0.245	0.907	1.00	0.425	0.513
$[\Omega^{26}_{trn,max}(5)]$		0.404	0.226	0.477	0.365	1.00	0.882
$[\Omega^{27}_{trn,max}(5)]$		0.383	0.337	0.477	0.565	0.863	1.00

[a]Due to differences in inherent task difficulty, prediction abilities for each task are normalized, with 1.00 representing a task being applied to all 20 sensors. Normalization was accomplished from the formula $value_{norm} = \frac{value_{raw}-0.5}{norm-0.5}$, where $norm$ represents the number by which $value_{raw}$ is to be normalized

of how the k detectors are chosen. Interpreted slightly differently, $g(k)$ yields the highest LOO classification performance relative to that of the full 20-detector array that was met or exceeded on each of the 21 tasks when using any single detector array of size k.

As displayed in Fig. 2.5, no combination of k detectors with $1< k <20$ performed as well as the full 20-detector array on all of the tasks evaluated. Specifically, for every combination of k detectors with $1< k <20$, at least one task was identified for which that detector combination exhibited LOO classification performance that was less than the performance obtained by using the full 20-detector array (i.e. $g(k) < 1.0$ for $k < 20$). Additionally, the value of $g(k)$ more closely approached that of the full array as k increased. Also, $g(k) > g(k-1)$ for $2< k <20$, hence increasing the size of the detector array always resulted in an increase in the value of $g(k)$.

The same procedure was performed for the Set B detectors with Series II data. As displayed in Fig. 6, the value of $g(k)$ was less than unity for all values of $k <20$, so for every k-member subset of the Set B array there was at least one task for which that k-member array was outperformed by the full 20-detector Set B array. Again $g(k)$ increased monotonically with increases in k. Note that the $g(k)$ function is somewhat biased in favor of the k-detector subsets because it reports the maximum LOO classification performance of all possible subsets having k detectors. If another nested cross-validation were used to choose the best subset of size k based on training set performance and this subset were then applied to an independent test set of data, the normalized classification performance would likely be lower than that reported in Figs. 2.5 and 2.6.

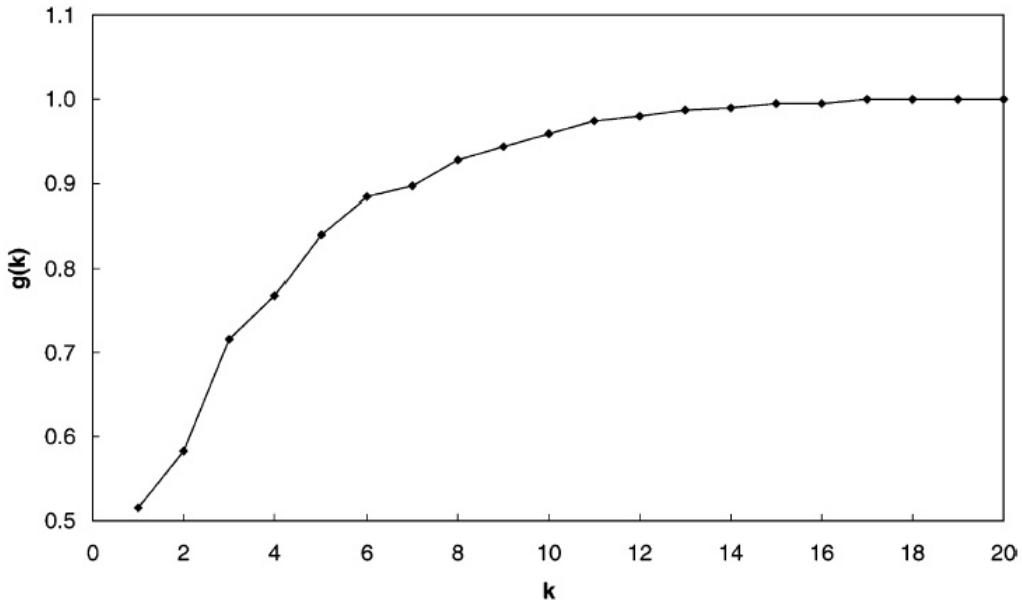


Figure 2.5: Classification performance vs. array size k. For each value of k in the range $1< k <20$, an exhaustive search of all possible k-detector combinations from the 20-detector Set A array was performed to identify the array having k detectors that had the best LOO classification performance for each of the 21 tasks in the Series I Data Set. For each task, the classification performance for any k-detector array was then compared to that of the full 20-detector Set A array. No combination of k detectors does strictly better than $g(k)$ relative to the full 20-detector Set A array on all 21 tasks of the Series I data run.

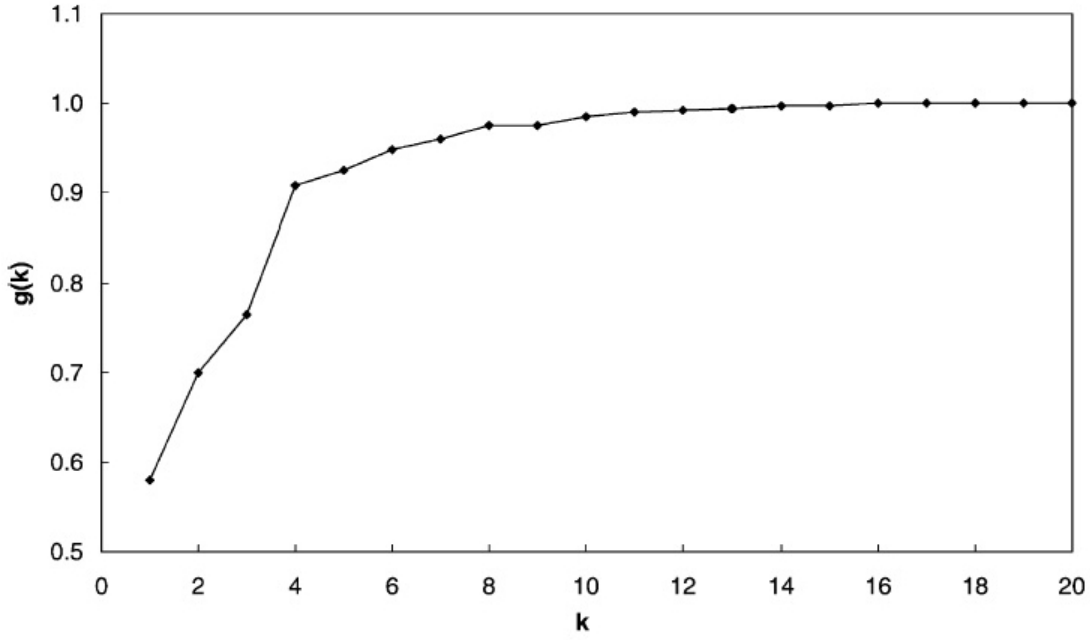


Figure 2.6: Classification performance vs. array size k. For each value of k in the range $1< k <20$, an exhaustive search of all possible k-detector combinations from the 20-detector Set B array was performed to evaluate the LOO classification performance of each array for each of the six tasks in the Series II Data Set. For each task, the classification performance for any k-detector array was then compared to that of the full 20-detector Set B array. No combination of k detectors does strictly better than $g(k)$ relative to the full 20-detector Set B array on all six tasks of the Series II data run.

2.4.5 Improvement in Classification Performance Upon Addition of Compositionally Different Detectors to an Array Relative to Addition of Nominally Identical Copies of Detectors to an Array

Some of the improvement in classification performance displayed by the full 20-detector array relative to the performance of k-detector subsets ($k < 20$) for either the Set A or Set B detectors could possibly result from the larger number of observations that are used in the analysis process for a given measurement task when a constant number of data points is acquired from more total detectors. Two methods were used to assess the differences in classification performance that resulted from obtaining more data from fewer compositionally different detectors relative to obtaining less data from a larger number of compositionally different detectors.

In the first approach, the Series I data were manipulated to generate a new data set, designated as Series I-A, in which two consecutive exposures of a single analyte to a detector within each task were taken as a single exposure of that analyte to two duplicate detectors. The full Series I-A Data Set therefore consisted of 40 descriptors (two consecutive $\Delta R_{eq}/R_b$ responses from 20 compositionally different Set A detectors) for 100 independent analyte exposures, as opposed to the 20 descriptors (one $\Delta R_{eq}/R_b$ value from each of the 20 compositionally different Set A detectors) for 200 independent analyte exposures that formed Data Set I. Descriptors obtained from consecutive analyte exposures to the same detector composition were linked computationally in the Series I-A Data Set so that the second data point in consecutive analyte exposures was required to be included in a classification task if the first data point was used in that classification task. The LOO classification performance for every individual combination of k compositionally different detectors, where $1< k <10$, for each of the 21 tasks was then evaluated. Because consecutive analyte exposure data points were coupled computationally and treated as arising from duplicate detectors of the same composition, this process was used to generate a $g(k')$ versus k curve, with $g(k')$ being defined as described above and with $k' = 2k$ being an even-numbered integer in the range $2 \leq k' \leq 20$. The $g(k')$ performance was then compared to the $g(k)$ performance of this same Set A array that has already been depicted in Fig. 2.5.

As displayed in Fig. 2.7, classification performance could be improved for at least one classification task for $1< J <21$ at every array size tested by including additional data from compositionally different detectors into the analysis as opposed to including the same amount of additional data from detectors which had the same chemical composition as those in the original array. This analysis reinforces the conclusion demonstrated above that in general at a given array size achieving optimal classification performance for different classification tasks requires use of the data produced by different collections of detectors (cf. Tables 3 and 4). Hence, addition of a compositionally different detector will improve the overall array classification performance for at least those tasks

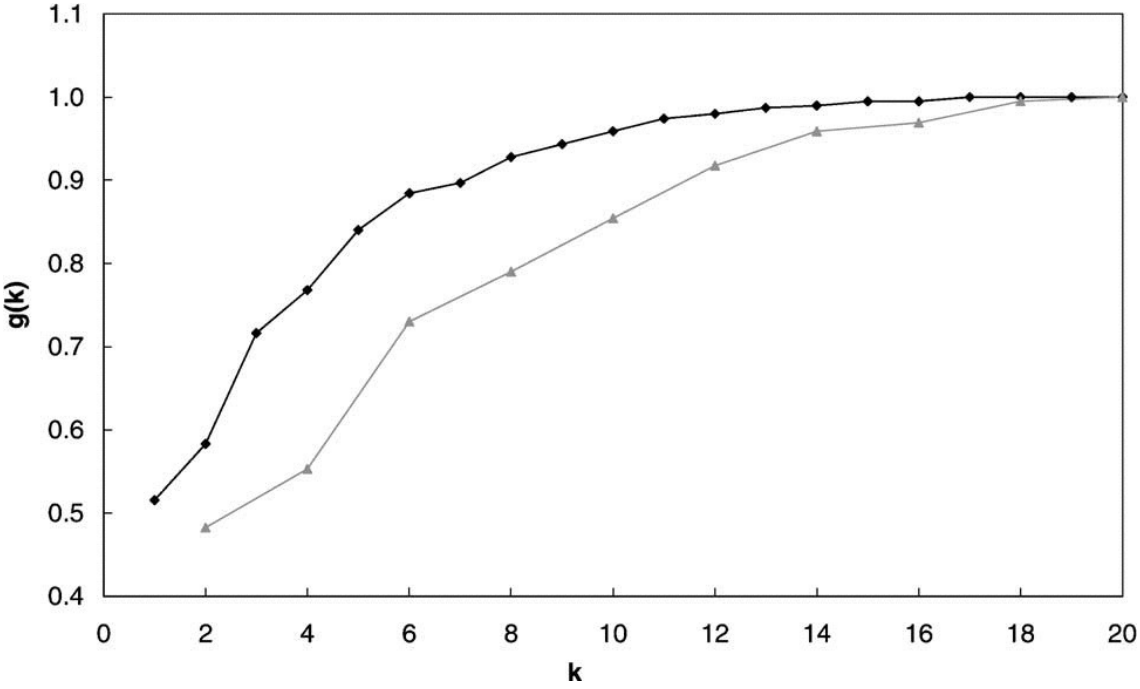


Figure 2.7: Classification performance vs. total detector array size k for detector Set A in conjunction with Series I and I-A data. The $g(k)$ curve for Series I data (◆) is identical to that displayed in Figure 2.5. The Series I-A data had two consecutive exposures of a single analyte to a detector within each task taken as a single exposure of that analyte to two duplicate detectors. Descriptors obtained from consecutive analyte exposures to the same detector composition were linked computationally in the Series I-A Data Set so that the second data point in consecutive analyte exposures was required to be included in a classification task if the first data point was used in that classification task. The LOO classification performance for every individual combination of k compositionally different detectors, where $1< k <10$, for each of the 21 tasks was then evaluated. Because consecutive analyte exposure data points were coupled computationally and treated as arising from duplicate detectors of the same composition, this process was used to generate a $g(k')$ vs. k curve (▲) with k' (=$2k$) being an even-numbered integer in the range $2< k' <20$. No combination of two copies of k compositionally different detectors does strictly better than $g(k')$ relative to the full 20-detector Set A on all 21 tasks of the Series I data run.

for which the added detectors provide important analyte classification information, relative to the classification performance that is obtained upon addition of detectors having nominally identical compositions as those present the original detector array (which by definition must not be optimal for at least some members of a diverse set of analyte discrimination tasks).

In the second approach, the six five-detector sets $[\Omega^{J}_{trn,max}(5)]$ for $22< J <27$ (Table 2.4) that yielded the best Fisher discriminant RF value for the training set data of each task for $22< J <27$, respectively, were used to generate six 20-detector arrays, $[\Omega^{J}_{trn,max}(5x4)]$ for $22< J <27$, by analyzing additionally the data produced by the three nominally identical copies of each detector composition that were available in the full 80-detector array of the Set B detectors. The weights for each detector in these 20-detector arrays were then independently determined using the training set data for each task of interest ($22< J <27$). The test set classification performance of these six 20-detector

Table 2.5: Classification performance of arrays formed from four copies of five compositionally different detectors.

Array	Detectors	Task					
		22	23	24	25	26	27
Absolute Performance							
$[\Omega^{22}_{trn,max}(5x4)]$	8, 9, 11, 18, 19	0.815	0.830	0.875	0.925	0.915	0.935
$[\Omega^{22}_{trn,max}(5x4)]$	8, 11, 13, 18, 19	0.805	0.815	0.825	0.910	0.875	0.945
$[\Omega^{22}_{trn,max}(5x4)]$	1, 8, 15, 19, 20	0.690	0.640	70.965	0.975	0.895	0.925
$[\Omega^{22}_{trn,max}(5x4)]$	1, 3, 4, 19, 20	0.725	0.640	0.925	0.955	0.6757	0.820
$[\Omega^{22}_{trn,max}(5x4)]$	4, 8, 9, 12, 18	0.700	0.705	0.845	0.780	0.880	0.920
$[\Omega^{22}_{trn,max}(5x4)]$	1, 8, 9, 12, 19	0.675	0.575	0.885	0.885	0.865	0.885
$[\Omega(20)]$		0.755	0.755	0.920	0.930	0.810	0.845
Normalized Performance[a]							
$[\Omega^{22}_{trn,max}(5x4)]$		1.24	1.29	0.893	0.988	1.34	1.26
$[\Omega^{22}_{trn,max}(5x4)]$		1.20	1.24	0.774	0.953	1.21	1.29
$[\Omega^{22}_{trn,max}(5x4)]$		0.745	0.549	1.11	1.10	1.27	1.23
$[\Omega^{22}_{trn,max}(5x4)]$		0.882	0.549	1.01	1.06	0.565	0.928
$[\Omega^{22}_{trn,max}(5x4)]$		0.784	0.804	0.821	0.651	1.23	1.22
$[\Omega^{22}_{trn,max}(5x4)]$		0.686	0.294	0.917	0.895	1.18	1.12

[a]Prediction abilities for each task are normalized, with 1.00 representing a task being applied to all 20 sensors. Normalization was accomplished from the formula $value_{norm} = \frac{value_{raw}-0.5}{norm-0.5}$, where *norm* represents the number by which $value_{raw}$ is to be normalized

arrays was then evaluated for each task in the range 22< J <27 of the Series II data, and the test set classification performance of these arrays was compared to the test set classification performance $\mathcal{P}_{tst}\{[\Omega(20)] \Rightarrow J\}$, 22< J <27, produced by the first set of 20 compositionally different detectors in the Set B array. Table 5 presents the results of this comparison for the absolute and normalized classification performance of these 20-detector arrays, with the normalization performed using the approach of eq. 2.4 to account for inherent differences in difficulty between tasks as well as to account for the two-class character of the discrimination tasks being evaluated.

Comparison of Tables 4 and 5 indicates that the mean absolute classification performance for a given task increased by ≈ 11% when three additional copies of each detector were included in the array. This increase can be attributed to averaging of noise through use of multiple copies of a given detector type, in accord with expectations and with recent observations indicating that the noise of polymer/carbon-black detectors decreases as the inverse square root of the detector area, for a constant detector film thickness [31]. Furthermore, the benefit of dimensionality reduction was clearly evident in that the $[\Omega^{J}_{trn,max}(5x4)]$ 20-detector arrays always yielded better classification performance than $[\Omega(20)]$ in the task J for which the $[\Omega^{J}_{trn,max}(5)]$ detector arrays were selected as providing the best test set classification performance for five-detector arrays (Table 2.5). Similarly, arrays that were identified as producing the optimal Fisher RF value on training set data for a specific task yielded excellent test set classification performance relative to $[\Omega(20)]$ in a duplicate

trial of that same task. However, these 20-detector $[\Omega^{J}_{trn,max}(5x4)]$ arrays generally yielded inferior test set classification performance relative to the set of 20 compositionally different Set B detectors when the specific 20-detector $[\Omega^{J}_{trn,max}(5x4)]$ arrays were used for other tasks in the Series II data run. For example, the set of 20 compositionally different Set B detectors yielded ≈ 13% better test set classification performance than did any of the $[\Omega^{J}_{trn,max}(5x4)]$, 24< J <27 detector sets for either of the 1-propanol/2-propanol mixture classification tasks (J=22,23).

Similarly, the detector sets $[\Omega^{J}_{trn,max}(5x4)]$ that yielded optimal training set Fisher RF values for the heptane/hexane mixture classification tasks, J=24,25 in test set classification performance $\mathcal{P}_{tst}\{[\Omega^{J}_{trn,max}(5x4)] \Rightarrow K\}$ when J, but $[\Omega(20)]$ yielded better test set classification performance than $[\Omega^{J}_{trn,max}(5x4)]$, J=24,25, for either of the 1-propanol/ 2-propanol mixture classification tasks J=22,23 Interestingly, any $[\Omega^{J}_{trn,max}(5x4)]$, J=22-27 array except $[\Omega^{25}_{trn,max}(5x4)]$ yielded a better test set classification performance for the m-xylene/p-xylene mixture classification tasks, J=26,27, than was obtained using the full 20 compositionally different Set B detector array, $[\Omega(20)]$. The relatively low test set classification performance obtained when using the entire 20-detector Set B array on tasks 26 and 27 suggests that the m-xylene/p-xylene separation is impeded by some detectors that have particularly low signal/noise ratios for this specific classification task. For example, the $[\Omega^{25}_{trn,max}(5x4)]$ array did not outperform the full 20-detector compositionally different Set B array, $[\Omega(20)]$, on tasks 26 and 27, and $[\Omega^{25}_{trn,max}(5x4)]$ did not contain detector 8, which was contained in all the other $[\Omega^{J}_{trn,max}(5x4)]$ arrays (i.e. J=22-24,26,27). Consistently, tasks 26 and 27 depended heavily on the use of detector 8 in Set B to achieve high test set classification performance, so the absence of detector 8 from the set $[\Omega^{25}_{trn,max}(5x4)]$ hindered the classification performance of $[\Omega^{25}_{trn,max}(5x4)]$ when that array was used in a very different mixture classification task than the one for which it was identified to yield optimum classification performance. Again, however, the results of Table 2.5 demonstrate that the use of compositionally different detectors improves the overall array classification performance for at least some tasks relative to the classification performance that is obtained upon addition of detectors having nominally identical compositions to those present in the original (best-performing at a given size for one specific task) detector array.

2.4.6 Discrimination Performance Between Benzene and Toluene and Between 1-Propanol and 2-Propanol as a Function of Analyte Concentration

The vapor classification tasks discussed above involved differentiation, at relatively high signal/noise ratios, between analytes that are so chemically similar that the signals produced by the detector arrays of interest did not result in well-separated clusters for the various analytes in ddimensional detector response space. A conceptually different challenge for an array of semiselective vapor detectors involves classification between analytes at sufficiently small vapor phase

concentrations that the low corresponding signal/noise responses for many of the detectors in the array will reduce the separation between data clusters that are otherwise well-separated at high analyte concentrations. To probe the effects of array size and array composition on such tasks, detector response data were collected for conducting polymer composites that were exposed to benzene, toluene, 1-propanol, and 2-propanol, respectively, each at both high and low vapor phase concentrations (tasks 28-31, Table 2.2).

The five-detector arrays that yielded the maximum Fisher RF value on training set data,$[\Omega^{J}_{trn,max}(5)]$, for each classi- fication task in the range J=28-31 were again selected from the collection of 20 compositionally different Set B detectors, $[\Omega(20)]$. Table 6 indicates the detectors in these $[\Omega^{J}_{trn,max}(5)]$ arrays and presents the test set classification performance of these arrays on tasks 28-31. Comparison of Tables 2.3 a-d and 6 clearly shows that different detectors formed the best-performing five-detector arrays for different analyte classification tasks. Additionally, this comparison shows that different collections of detectors formed the $[\Omega^{J}_{trn,max}(5)]$ arrays for classification of analytes at high concentration relative to the $[\Omega^{J}_{trn,max}(5)]$ arrays that were identified as producing the best classification performance for analytes at low vapor phase concentrations.

The compositional differences between these optimally performing five-detector $[\Omega^{J}_{trn,max}(5)]$ arrays were investigated in more detail for tasks 28-31 of the Series II data run. To avoid any bias that might arise from incrementally small differences in classification performance between the bestperforming detector set for a task, $[\Omega^{J}_{trn,max}(5)]$, and other nearly optimal five-detector sets for the same classification task, detectors were ranked based on the frequency with which a detector was contained in the 20 five-detector arrays that produced the 20 best Fisher RF values on the training set data for the classification task of interest. Table 6 summarizes the results of such an analysis for tasks 28-31 of the Series II data.

For the low concentration, 1-propanol versus 2-propanol analyte classification task (task 28), the detectors that were most commonly contained in the 20 best-performing arrays (appearing in more than 10 out of the 20 array that yielded the best training set Fisher RF values) were detectors 16 (17), 18 (20), 19 (20), and 20 (16), with the numbers in parentheses indicating the number of the 20 total bestperforming arrays for the task under consideration which contained that specific detector. Three of these detectors, 18-20, were among the four detectors that exhibited the largest S/N ratios for 1-proponol at $P/P°$=0.0075, having S/N values of 186, 63.2, and 30.0, respectively. This makes sense in that a priority at low concentration conditions is to robustly detect the signal relative to the baseline detector noise, and robust detection will in general correlate with good analyte discrimination under such conditions. Detector 16 only exhibited a S/N value of 9.83 for 1-proponol at $P/P°$=0.0075, and its frequent presence in the best-performing five-detector arrays for this classification task can not be explained solely on the basis of S/N ratios and therefore reflects important analyte discrimination power for the task of concern. In contrast, the detectors

Table 2.6: Detectors represented in best-performing five-detector arrays for high concentration and low concentration mixture classification tasks.

(a)

Task	Sens	S/N[a]	#Sel[b]	Sens	S/N	#Sel	Sens	S/N	#Sel	Sens	S/N	#Sel	Sens	S/N	#Sel
28	15	16.9	6	16	9.83	17	18	186	20	19	63.2	20	20	30.0	16
29	1	131	20	3	50.9	20	5	41.6	20	16	24.6	16	18	123	004
30	8	66.2	16	10	7.80	5	15	94.1	20	18	187	20	19	33.6	20
31	2	309	8	5	94.6	14	8	385	20	15	570	12	19	78.9	17

(b)

Designation	Best Sets of Five					Performance
$[\Omega^{28}_{trn,max}(5)]$	1	16	18	19	20	0.920
$[\Omega^{29}_{trn,max}(5)]$	1	3	5	10	16	1.00
$[\Omega^{30}_{trn,max}(5)]$	7	8	15	18	19	0.955
$[\Omega^{31}_{trn,max}(5)]$	2	3	5	8	19	1.00

[a]Signal to Noise, calculated as three times thebaseline standard deviation

[b]Number of times sensor was chosen among the best 20 5-sensor sets for a particular task

most frequently contained in the 20 five-detector arrays that yielded the best Fisher RF training set values for the high concentration 1-propanol versus 2-propanol analyte classi- fication task (task 29) were detectors 1 (20), 3 (20), 5 (20), and 16 (16). Only detector 1 had a relatively high S/N value (131) for 1-propanol at $P/P°$=0.075, with the other detectors exhibited S/N ratios for 1-propanol at $P/P°$=0.075 of 50.9, 41.6, and 24.6, respectively. Although these detectors were not among the most sensitive for the task at hand, they were selected among the 20 best-performing five-detector arrays a total of 76 times, which is the maximum frequency with which four detectors could be selected among 20 unique sets of five detectors. Because this detector selection preference persisted despite the presence of detectors that possessed higher S/N values for 1-propanol at $P/P°$=0.075, such as detectors 8, 18, and 20 (S/N between 309 and 1240), the selected detectors are clearly providing more classification information between the two analytes in the test set under conditions when discrimination is more important than robust signal detection. Thus, different five-detector combinations provided optimal analyte classification performance for analytes at high vapor phase concentrations relative to classification of these same analytes at low vapor phase concentrations, due to the different relative importance of signal/noise ratios and detector response differences that are crucial for optimizing analyte classification under the two different task conditions.

Similar trends were found in the low concentration versus high concentration analyte classification tasks for benzene versus toluene (tasks 30, 31). At low vapor phase analyte concentration (task 30), four detectors (8 (16), 15 (20), 18 (20) and 19 (20)) were selected a total of 76 times (the maximum) in the 20 five-detector arrays that yielded the best training set Fisher RF values for this classification task. These detectors had S/N values for benzene at $P/P°$=0.0065 of 66.2, 94.1, 187, and 33.6, respectively, which were the four highest S/N ratios for benzene at $P/P°$=0.0065 produced by

any of the 20 detectors in the Set B detector array. None of the other 16 detectors in the Set B array were represented at a significant frequency in the 20 best-performing five-detector arrays for this analyte classification task. The consistent selection of the four detectors with the highest S/N values again indicates the relative importance of robust analyte detection at low analyte concentration in order to perform analyte classification.

In contrast, at high benzene and toluene vapor phase concentration, different detectors appeared with high frequency in the best-performing five-detector arrays (task 31). Four detectors (5 (14), 8 (20), 15 (12), and 19 (17)) were contained at least 10 times among the 20 best-performing five-detector arrays for this task. Three of these detectors (8, 15, 19) were also amongst the most frequently selected detectors for the low concentration benzene versus toluene classification task, but detector 15 was not represented as often in the five-detector arrays that performed best in the low concentration benzene versus toluene classification task. Furthermore, detector 18 was not among the most commonly chosen detectors for task 31, being represented only three times among the best 20 five-detector arrays for this classification task. This change in selection frequency suggests that detector 18 completely lost its advantage at higher concentrations compared to other available detectors in the Set B array. Detector 18 and, to an extent, detector 15, were replaced with four detectors that had lower S/N values than those of detectors that were not frequently selected at low benzene or toluene concentration: detectors 5 (14), 2 (8), 4 (8), and 10 (7). The differences in selection preferences among these last four tasks again indicate that that the relative importance of sensitivity versus selectivity varies with analyte concentration. These results therefore clearly indicate that different five-detector arrays will provide the best classification performance for different tasks, even when the differences in tasks involve the same analytes but different analyte concentrations. Again having more compositionally different detectors in the array therefore allows selection of the detectors that will perform best for the task at hand and will yield improved performance for at least some classification tasks relative to a five-detector array that has been optimized for some other analyte classification purpose.

2.5 Discussion

In theory, only two broadly responsive and partially uncorrelated detectors should be required to identify any single-component analyte, at any concentration. These two detectors would provide two descriptors that would allow solution for the two unknowns, analyte identity and analyte concentration, from the vapor detection system. An array having more than two partially uncorrelated detectors therefore constitutes, in principle, an overdetermined system for "universal" classification and quantification of single-component analyte vapors.

A similar analysis applies to differentiation between any two different, possibly multi-component,

analytes. In principle, only one detector is needed for such a task, provided that the detector response can be determined with sufficient accuracy and precision to allow differentiation between the presence of the two different analytes when exposed to the detector.

However, practical limitations on the vapor classification performance of actual semi-selective detector arrays arise from the fact that the partition coefficients of two singlecomponent analytes can not be measured with arbitrarily high precision from the response of a single sorption-based detector. The exact number of different detectors required to achieve a certain level of analyte classification performance in any practical system will therefore depend upon the details of the response properties of the detectors and on the chemical diversity of the analytes in the test set. For a set of chemically very different, single-component analytes being probed by a set of chemically diverse detectors, the lack of ultrahigh precision on a single, or on two, detectors can generally be compensated through obtaining data from a few additional (typically four to five total) detectors, provided that the responses of the additional detectors are at least partially uncorrelated with those of the first two detectors.

However, when the task (or the background clutter to be discriminated against) is variable, and/or when the analytes are not very different chemically, the situation can be somewhat different. In this mode of application of detector arrays, the analysis might for example be a "forced choice" situation determining which of two different analytes were present in a sample, having the knowledge that the samples of interest will only consist of one or the other of the pair of specified analytes. Arrays of four to five detectors may well suffice to produce good classification performance between a particular pair of analytes, but four to five different detectors will in general produce the best classification performance for different tasks. For example, detectors containing polar polymers will generally display higher partition coefficients and therefore will exhibit better signal/ noise ratios towards polar analytes than will detectors formed using very nonpolar polymers. Thus, for the task of classifying several polar analytes, arrays that largely consist of the polar detectors will in general outperform arrays that largely consist of nonpolar detectors, because the nonpolar polymers will hardly respond to the polar analytes and will provide low signal/noise ratios in the response descriptors to be used for this analyte resolution task. Therefore, if the task is known fully in advance, i.e. if the analyte pairs and their concentrations are defined precisely, and if the background levels and interferent identities and levels are fully known, a nearly optimal subset of detectors can in principle be selected from a broader set of detectors. However, when the task is variable or when the background is not well defined, having more detectors available of roughly equal inherent signal/noise characteristics at a given response level allows flexibility in the choice of down-selected detector sets to achieve optimal vapor classification performance for different tasks. In this approach, a hierarchial classification approach would be used, in which the full detector array response would be used to determine which family of analytes were present and the responses

produced by the appropriate subset of detectors would be used to further discriminate between analytes within a family. As demonstrated herein, it is not necessary to down-select such subsets of detectors in advance of data analysis. Instead, obtaining data for a chemically diverse set of detectors and importing the full data set into the Fisher linear discriminant algorithm produced performance that was as good, or nearly as good, as the optimal detector subset of any size for any of the tasks evaluated. Additionally such a procedure yielded classification performance that for at least some tasks was superior to that of an individual down-selected detector subset when the subset was subsequently used for tasks for which it was not originally evaluated.

A related issue is that capturing the most variance between a chemically diverse test set of analytes is not in all instances the critical factor for achieving analyte classification. Rather, the variance between the classes of interest is the key quantity. Consider down-selecting a set of ”optimum” detectors for best classifying a training set of presentations comprised of achiral analytes. Further consider the situation in which the detectors of interest are comprised of a set of achiral polymer films, except for one detector that is formed using an enantiomerically pure chiral polymer. Additionally, consider the situation in which the chiral detector is significantly noisier at a given signal response level than the achiral detectors. Clearly, the noisy chiral detector would be eliminated in a search for the down-selected detector subset that produces the best classification performance for the test set of achiral analytes. However, if the test set is changed so that it includes a pair of enantiomers, inclusion of the chiral detector is absolutely necessary, regardless of its noise level, to obtain any resolution or classification information in the chiral analyte resolution portion of the task. If this chiral detector replaced one detector in the original array, then the original array would by definition perform more poorly in classification of the achiral analytes than it did previously. This reduction in performance probably would not be obtained if instead the chiral detector had been added to the original array and the Fisher discriminant algorithm had been used to form the optimal decision boundary for each task of concern using the responses produced by the complete set of detectors in the array.

Similarly, if the analyte test set were switched such that it contained only enantiomerically pure chiral analytes, the optimal detector set would likely contain only detectors that incorporated enantiomerically pure chiral polymers, regard- less of whether these detectors were in the down-selected set that yielded the best classification performance for tasks involving the achiral analytes. Thus, the variance between classes being discriminated, as opposed to the overall variance between all possible chemical analytes representing all possible diverse chemical features, is the key factor in determining the optimization of an array for a specific task. This between-class variance necessarily depends on the nature of the analyte test set and on the differences between the various responses that this test set produces on the detectors. Having more detectors available clearly allows selection of the most appropriate detector set for the task at hand.

The nature of the classification task also affects which subsets of detectors will be most useful for that task. At low analyte concentration, detectors that exhibit the best signal/ noise ratios will generally be favored because such detector responses will produce the best characterization of the individual analyte data points that are used to form the data clusters which dictate the position of the classification decision boundary. However, a single set of low-noise detectors is not sufficient to provide optimal classification performance between a chemically diverse group of test analyte pairs, even for a group of tasks that consists solely of differentiation between various analytes at low vapor phase concentrations. Detectors with low S/N values will indeed be generally included in optimally performing arrays of a given size for classification between pairs of analytes at low vapor phase concentration. However, the signal of a given detector will vary significantly for different analyte classes, causing its S/N ratio to change significantly for different discrimination tasks. Hence, while a given detector might have a relatively high S/N value in response to low concentrations of hydrophobic analytes, it will likely have a much lower S/N value in response to low concentrations of hydrophilic analytes. Different collections of detectors were thus found to comprise optimally performing five-detector arrays for classification of low concentrations of 1-propanol versus 2-propanol relative to the detectors that formed the optimally performing five-detector arrays for classification of low concentrations of benzene versus toluene. In contrast, separation between two chemically similar analytes at high concentration (for example, enantiomers or structural isomers) favors the use of detectors that probe the perhaps subtle, but critical, chemical differences between the analytes of interest. Hence having a diverse collection of compositionally different detectors in the array provides access to the best set of detector responses for the task at hand without having to physically replace and redesign the array for each task of interest.

2.6 Conclusions

For a broad set of chemically very different test solvents, in principle only two semi-selective detectors are needed to provide robust information on the identity and concentration of any pure analyte in the test set. For polymer/carbon-black composite chemiresistive vapor detectors, excellent classification performance was observed for arrays as small as three to four detectors for pure analyte vapors at concentrations high enough to produce high detector signal/noise ratios for the analytes of interest. However, when the signal strength was lowered, or when the analytes were chemically very similar, more detectors were required to achieve optimum classification performance for all tasks investigated. Classification performance in general either increased or did not decrease significantly as the number of chemically different detectors increased. Furthermore, different subsets of the detector array produced the best classification performance at a fixed array size for different analyte classification tasks. Hence, the full compositionally different detector array

always yielded better classification performance than any smaller size array for at least some vapor classification tasks. Reduction in dimensionality was observed to be advantageous when the task was identified in advance, because multiple copies of detectors that had been identified as yielding the best training set classification performance for a given task at a fixed array size yielded better classification performance than the same total number of compositionally different detectors when both arrays were used for that specific classification task. However, the full array of compositionally different detectors yielded better test set classification performance than did any fixed array containing multiple copies of fewer compositionally different detectors for at least some other task of interest in a broadly construed set of analyte classification tasks. Subsets of detectors were identified that yielded robust discrimination between D_2O and H_2O, between compositionally similar mixtures of 1-propanol and 2-propanol, and between compositionally similar mixtures of n-hexane and n-heptane, attesting to the excellent analyte discrimination power that can be obtained at least in certain tasks through use of an array of semi-selective chemiresistive vapor detectors even when no single detector provides the needed chemical resolution to differentiate between the analytes of interest.

Bibliography

[1] Gardner, J.; Bartlett, P. *Electronic Noses: Principles and Applications;* Oxford Science Publications: Oxford, 1999.

[2] Albert, K.; Lewis, N.; Schauer, C.; Sotzing, G.; Stitzel, S.; Vaid, T.; Walt, D. *Chemical Reviews* **2000,** *100,* 2595–2626.

[3] Patel, S.; Jenkins, M.; Hughes, R.; Yelton, W.; Ricco, A. *Analytical Chemistry* **2000,** *72,* 1532–1542.

[4] Park, J.; Groves, W.; Zellers, E. *Analytical Chemistry* **1999,** *71,* 3877–3886.

[5] Doleman, B.; Lonergan, M.; Severin, E.; Vaid, T.; Lewis, N. *Analytical Chemistry* **1998,** *70,* 4177–4190.

[6] Zellers, E.; Batterman, S.; Han, M.; Patrash, S. *Analytical Chemistry* **1995,** *67,* 1092–1106.

[7] Zellers, E.; Park, J.; Hsu, T.; Groves, W. *Analytical Chemistry* **1998,** *70,* 4191–4201.

[8] Rose-Pehrsson, S.; Grate, J.; Ballantine, D.; Jurs, P. *Analytical Chemistry* **1988,** *60,* 2801–2811.

[9] Grate, J.; Abraham, M. *Sensors and Actuators B - Chemical* **1991,** *3,* 85–111.

[10] Grate, J.; Martin, S.; White, R. *Analytical Chemistry* **1993,** *65,* 940A–948A.

[11] Hierlemann, A.; Weimar, U.; Kraus, G.; Schweizerberberich, M.; Gopel, W. *Sensors and Actuators B - Chemical* **1995,** *26,* 126–134.

[12] Nakamoto, T.; Fukunishi, K.; Moriizumi, T. *Sensors and Actuators B - Chemical* **1990,** *1,* 473–476.

[13] Hatfield, J.; Neaves, P.; Hicks, P.; Persaud, K.; Travers, P. *Sensors and Actuators B - Chemical* **1994,** *18,* 221–228.

[14] Hodgins, D. *Sensors and Actuators B - Chemical* **1995,** *27,* 255–258.

[15] Dickinson, T.; Michael, K.; Kauer, J.; Walt, D. *Analytical Chemistry* **2000,** *71,* 2192-2198.

[16] Albert, K.; Walt, D.; Gill, D.; Pearce, T. *Analytical Chemistry* **2001,** *73,* 2501–2508.

[17] Ronot, C.; Archenault, M.; Gagnaire, H.; Goure, J.; Jaffrezicrenault, N.; Pichery, T. *Sensors and Actuators B - Chemical* **1993,** *11,* 375–381.

[18] Sutter, J.; Jurs, P. *Analytical Chemistry* **1997,** *69,* 856–862.

[19] Lonergan, M.; Severin, E.; Doleman, B.; Beaber, S.; Grubb, R.; Lewis, N. *Chemistry of Materials* **1996,** *8,* 2298–2312.

[20] Doleman, B.; Sanner, R.; Severin, E.; Grubbs, R.; Lewis, N. *Analytical Chemistry* **1998,** *70,* 2560–2564.

[21] Severin, E.; Lewis, N. *Analytical Chemistry* **2000,** *72,* 2008–2015.

[22] Severin, E.; Sanner, R.; Doleman, B.; Lewis, N. *Analytical Chemistry* **1998,** *70,* 1440–1443.

[23] Vaid, T.; Burl, M.; Lewis, N. *Analytical Chemistry* **2001,** *73,* 321–331.

[24] Duda, R.; Hart, P. *Pattern Classification and Scene Analysis;* John Wiley and Sons: New York, 1984.

[25] Osbourn, G.; Martinez, R. *Pattern Recognition* **1995,** *28,* 1793–1806.

[26] Osbourn, G.; Bartholomew, J.; Ricco, A.; Frye, G. *Accounts of Chemical Research* **1998,** *31,* 297–305.

[27] Polikar, R.; Shinar, R.; Udpa, L.; Porter, M. *Sensors and Actuators B - Chemical* **2001,** *80,* 243–254.

[28] Jurs, P.; Bakken, G.; McClelland, H. *Chemical Reviews* **2000,** *100,* 2649–2678.

[29] Gardner, J.; Bartlett, P. *Sensors and Actuators B - Chemical* **1996,** *33,* 60–67.

[30] Severin, E.; Doleman, B.; Lewis, N. *Analytical Chemistry* **2000,** *72,* 658–668.

[31] Briglin, S.; Freund, M.; Tokumaru, P.; Lewis, N. *Sensors and Actuators B - Chemical* **2002,** *82,* 54–74.

[32] Skoog, D.; Holler, F.; Nieman, T. *Principles of Instrumental Analysis, 5th Ed.;* Harcourt-Brace: Philadelphia, 1998.

Chapter 3

Development and Characterization of Polymer/Carbon-black Chemical Vapor Sensors Utilizing Percolative Conduction Characteristics

3.1 Abstract

Chemical vapor sensor chemiresistors have been developed based on very low fractions of carbon-black (1-12% w/w) that allow for very high responses, with $\Delta R_{eq}/R_b$ values over 100 in many cases. The responses of a small array of low carbon-black sensors and a similar array of higher carbon-black analogs are both exposed to 16 different analytes, and compared on the basis of response magnitude, sensitivity, correlation, and classification.

Low carbon-black sensors are found to be highly nonlinear with respect to concentration presenting both challenges in terms of calibration and benefits in terms of additional modes of use impossible with standard, high carbon-black sensors. Low carbon-black sensors are typically more susceptible to drift and show somewhat less reproducibility, but exhibit both greater responses and predominantly higher sensitivity.

3.2 Introduction

In the past 20 years, arrays of broadly cross-reactive sensors have received significant attention for their possible use in detection and classification of analyte vapors. These systems can be based on many signal transduction modalities, including polymer-coated quartz-crystal microbalances (QCM) or surface-acoustic wave (SAW) devices,[1–3] glass beads or optical fibers coated with dye impregnated polymers,[4–8] conducting polymer[9–11] or polymer composite[12–14] chemically sensitive resistors, polymer-coated micromachined cantilevers,[15] polymer-based capacitors and FETs,[16,17]

and metal oxide chemiresistors.[18–21]

Work in our laboratory has focused on the development of conducting composite films consisting of ordinary insulating polymers and conductive carbon-black. These sensors pass current through continuous pathways of the conductive carbon-black that traverse the gap between a pair of electrodes which is between 10 μm and 1 mm wide. When the sensors encounter chemical vapors they swell, necessarily breaking some fraction of the continuous carbon-black pathways increasing the bulk dc resistance of the composite.[22] We typically report the resistance change in its fractional form, as $\Delta R_{eq}/R_b$, because it is more reproducible between sensors containing the same insulating polymer component but of different baseline resistances.[13]

Our sensors are designed to indirectly measure the mass uptake that occurs when polymer films are exposed to chemical vapors. In this way, they have been designed to essentially mimic QCM sensors, which directly measure mass uptake upon analyte sorption. However, the behavior of our sensors is dependent on the fraction of conductive carbon-black they contain. Films with small fractions of carbon-black have been shown to exhibit a power law relationship between resistance and degree of analyte uptake; this relationship approximates linearity at higher fractions of conductor.[13]

At the time polymer/carbon-black composite sensors were designed, emphasis was placed on their maintaining a linear relationship between resistance and vapor concentration. This necessitated the use of high fractions of carbon-black in films (20% w/w), ultimately resulting in sensors that exhibited highly linear responses as a function of concentration over a wide range of concentrations.[13,23] Development of sensors exhibiting linear responses facilitated normalization techniques that could easily extract a concentration-independent signal S from a raw signal S that corresponds to a $\Delta R_{eq}/R_b$ sensor response vector (eq 3.1).

$$S'_{i,j} = \frac{S_{i,j}}{\sum_j^d S_{i,j}} \tag{3.1}$$

However, the high degree of linearity came at a price because the responses were thus restricted, they were necessarily kept out of the percolation regime, which would have resulted in dramatically higher responses.[22] The result was that a typical sensor response yielded a resistance change upon typical analyte exposures of only a few percent at best. Only under exceptional circumstances would a sensor response pass 30%.[23–25] Given the extremely low noise of these sensors of approximately 1 part in 90,000[26] and their high degree of repeatability,[27] the relatively small responses on the order of a few percent can still be made with a high degree of sensitivity.

To achieve resolution of 1 part in 90,000, a minimum of 17 bits are required for full-scale measurements. This is not possible for all desired applications of our sensors, in which less expensive, smaller, or low-power electronics might be preferred. In cases where lower resolution

measurement hardware is used, the low noise of the high carbon-black sensors is meaningless, as it is dwarfed by the digital noise of the hardware. In these cases, the greater signals provided by sensors fabricated with lower fractions of carbon-black would be a strongly preferred, and in many cases necessary to measure a discernable signal.

Varying the carbon-black fraction in a sensor also affords the opportunity to generate a greater amount of sensor diversity without actually introducing new components (such as more types of polymers). This approach has proven useful when varying the amount of plasticizer in a polymer/plasticizer/conductor system, where a single polymer and plasticizer can be combined in varying ratios to yield a sensor array.[28] By including different sensors that employ the same sensing element but different fractions of carbon-black, sensors can be developed that might be significantly less correlated than nominally identical sensors would be, which would add some degree of useful information to the system.

Toward these goals we report here the development and characterization of a mixed array of high- and low-carbon-black sensing films. We seek to determine whether and in what circumstances using arrays containing lower carbon-black fractions is beneficial. Overall, for the inclusion of nominally identical sensors differing only in carbon-black content to prove useful, they must be more selective, more sensitive, more reproducible, and/or generate responses that are uncorrelated with their high carbon-black analogs. Here we compare sensors derived from high and low carbon-black fractions with regard to these considerations. We also detail methods for determining from a single analyte exposure whether a sensor has become "percolative," thereby leaving the linear concentration/response regime.

3.3 Experimental

3.3.1 Sensor Fabrication

Four polymers were used to generate the sensors used in this work: poly(ethylene oxide) (PEO), MW=100,000; poly(ethylene-co-propylene) (PEP), 40% propylene; poly(ethylene-co-vinyl acetate) (PEVA), 40% vinyl acetate; and poly(vinyl stearate) (PVS). Black Pearls 2000 carbon-black (Cabot Co.) was used as the conductive element for all sensors. Polymer and some fraction of carbon-black were mixed by co-dissolving them in a compatible solvent and sonicating for 30 min to break up the agglomerated carbon-black particles, which produces nanoparticles with radii of approximately 12 nm with a surface area of 1500 m^2/g.[29] The solutions were then deposited across leads of interdigitated electrodes with gaps of 10 μm and total interfacial contact distance of 2 cm using a hobby airbrush as described previously.[13,26,27]

Solutions were sprayed on sensor electrodes to generate films with resistances no higher than

single-digit MΩ, and lower where possible. For each polymer, two sensor types were created: one with a low fraction of carbon-black, and one with 40% (w/w) carbon-black, which has been shown previously to ensure linear response vs. concentration characteristics over a broad concentration range.[13] The low carbon-black fractions used varied by polymer type, with the fraction used representing the lowest fraction that ultimately produced a viable sensor for that polymer type. The fractions of carbon-black used were 2%, 12%, 7%, and 1% for PEO, PEP, PEVA, and PVS, respectively.

Additionally, two PVS films sensor films were generated from the same 1% carbon-black feedstock. One film was applied to an interdigitated electrode as described above, and one was applied to a 10 MHz resonant quartz crystal microbalance. Changes in resonant frequency of each coated crystal can be related to changes in the mass of the film through the Sauerbrey equation (eq 3.2):

$$\Delta f = \frac{-\Delta m \times f_0^2}{\rho \times v \times A} \tag{3.2}$$

Here, f_0 is the resonant frequency of the crystal, ρ is the density of quartz, v is the speed of sound in quartz, and A is the area of the crystal. Δf and Δm are changes in the resonant frequency and mass, respectively.

3.3.2 Data Collection

For a first set of experiments, a QCM crystal coated with a PVS/1% carbon-black sensing film was exposed to THF concentrations between 0 and P/P°=0, 0.067. The concentration vs. time profile was shaped as a sawtooth, with a single cycle (from 0 to P/P°=0, 0.067) requiring 1000 s. Frequency was measured using a Hewlett-Packard 53181A frequency counter, and resistance with a Hewlett-Packard 34970A multimeter. Frequency was converted to mass as described above. For a second set of experiments, the 2 copies (high and low carbon-black) of 5 sensor types were placed in an airtight stainless steel flow chamber which was connected to computer controlled vapor delivery apparatus, as described previously.[30] Resistances were measured by a Keithley 2002 multimeter and 7001 multiplexer. 16 different analytes were used to test the sensors, as shown in Table 3.1. These analytes were delivered at fractional saturation concentrations of P/P°= 0.005, 0.01, 0.02, 0.03, 0.04, 0.06, 0.10, 0.12, 0.14, and 0.16 each. Within a given run, only 8 analytes were presented, but each combination of the 8 analytes and 10 concentrations was repeated 10 times within a run, with the exposure order of the analytes randomized to minimize the effect of sensor history effects upon the results. Each analyte exposure consisted of 70 seconds of clean laboratory air, 80 seconds of analyte, followed by another 60 seconds of laboratory air during which time the sensors are allowed to rest before the next exposure. Of the 8 analytes used in each experiment, four of them iso-octane, toluene, THF, and chloroform were used in each run, accompanied in all cases by four

Table 3.1: Summary of experiment runs.

Run	Since Last Run	Analyte 1	Analyte 2	Analyte 3	Analyte 4
1	N/A	isobutyl acetate	1-chlorobutane	ethanol	water
2	2 weeks	isobutyl acetate	1-chlorobutane	ethanol	water
3	1 day	isobutyl acetate	1-chlorobutane	ethanol	water
4	3 months	isobutyl acetate	1-chlorobutane	ethanol	water
5	1 day	1,1,1-TCE	pyridine	decane	n-octanol
6	1 day	dichloromethane	isopropyl benzene	methyl acetate	methanol

Note: All experiments also used iso-octane, toluene, THF, and chloroform.

other analytes. The first three experiments all used the same analyte set, and also the same order of presentation for the analytes. Two weeks passed between experiments 1 and 2, and nearly 3 months between 2 and 3. Experiment 4 was run a day after the end of experiment 3, and experiment 5 one day after the end of experiment 4. Table 3.1 summarizes these experiment runs.

3.3.3 Data Analysis

For each response signal derived from a single analyte exposure to a given sensor, the data were first baseline corrected and then transduced to extract a single characteristic value. Baseline correction was accomplished by fitting a trendline to the analyte-free pre-exposure period, with the trend extracted from the entire signal. From the baseline corrected data were then extracted the equilibrium resistance change (ΔR_{eq}) and the baseline resistance (R_b). The characteristic chosen is $\Delta R_{eq}/R_b$, which has been shown to be more consistent than ΔR_{eq} between sensors such as those used in this work that are derived from a single polymeric component but differing carbon-black fractions.

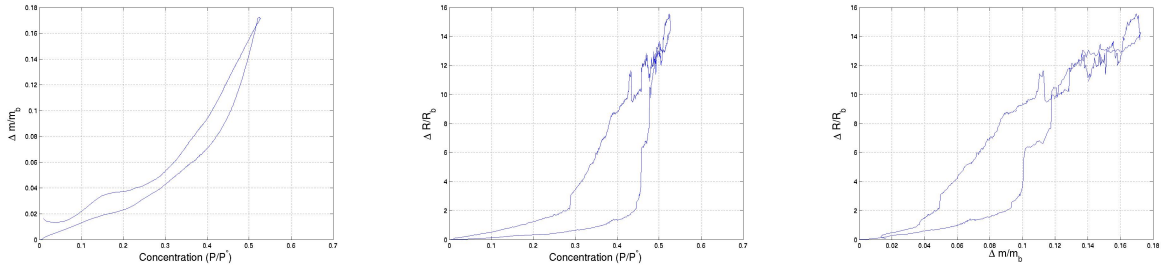
This work employs principal components analysis (PCA) and Fishers Linear Discriminant (FLD) to estimate how successfully our sensors distinguish among different analytes (Duda). PCA simply rotates the data to ensure that the first few dimensions contain as much of the variance as possible, while FLD is a supervised algorithm that attempts to maximize separation between clusters of pre-identified clusters of points. These algorithms are implemented here as in previous work.[27,31,32]

3.4 Results and Discussion

3.4.1 Determination of the Relationship Among Mass, Resistance, and Concentration

Figure 3.1 shows response vs. concentration profiles for a PVS/1% carbon-black composite sensor: a) response as $\Delta m_{eq}/m_b$, b) response as $\Delta R_{eq}/R_b$, as well as c) $\Delta m_{eq}/m_b$ vs. $\Delta R_{eq}/R_b$. Mass changes are linear with concentration until approximately P/P°=0.20-0.30, after which the curve gradually

Figure 3.1: Response vs. concentration profiles for a poly(vinyl stearate) sensor (1% carbon-black) exposed to tetrahydrofuran: a) $\Delta m_{eq}/m_b$ vs. concentration, and b) $\Delta R_{eq}/R_b$ vs. concentration. Data were recorded from a sawtooth concentration vs. time profile. Figure c) displays $\Delta R_{eq}/R_b$ vs. $\Delta m_{eq}/m_b$.



(a) (b) (c)

increases past linear. This implies that PVS/THF partition coefficient is consistent at moderate concentrations, increasing somewhat at high THF loadings. Changes in resistance are linear with changes in mass until approximately $P/P°$=0.10, at which resistance dramatically increases with respect to mass. This result clearly shows that what effects may be taking place, they are not explainable by polymer/vapor sorption phenomena, which would also be manifest in the mass plot. Thus, the seemingly exponential (or at least highly supra-linear) responses derived from the low carbon-black sensors are not due to changes in polymer morphology, but rather to percolative behavior that is manifest at low fractions of carbon-black and relatively high analyte concentrations.

3.4.2 Spatial Analysis of Sensor Responses

Figures 3.2 a-h show raw sensor responses ($\Delta R_{eq}/R_b$) vs. concentration ($P/P°$) for iso-octane, THF, and chloroform. As shown in the previous section, the low carbon-black sensors show significant nonlinear characteristics with respect to concentration, while the high carbon-black sensors do not. Figure 3.3 displays first principal component derived from the a) high carbon-black sensors and b) low carbon-black sensor sets, plotted vs. analyte concentration. As expected, the first principal component that which contains the bulk of the variance from the data is monotonic, and roughly linear for the high carbon-black sensors and roughly exponential for the low carbon-black sensors. Characteristically, further principal components of each grow noisier.

For the data collected in this work, the first principal component derived from high carbon-black sensors serves as an effective means of removing concentration effects, as it remains linear with concentration even if some of the high carbon-black sensors in the array do not. Decomposing the data into principal components also allows for recognition of percolative effects earlier than

Figure 3.2: Raw $\Delta R_{eq}/R_b$ Responses of High and Low Carbon-black Sensors to Iso-octane (○), Toluene (□), and Tetrahydrofuran (×)

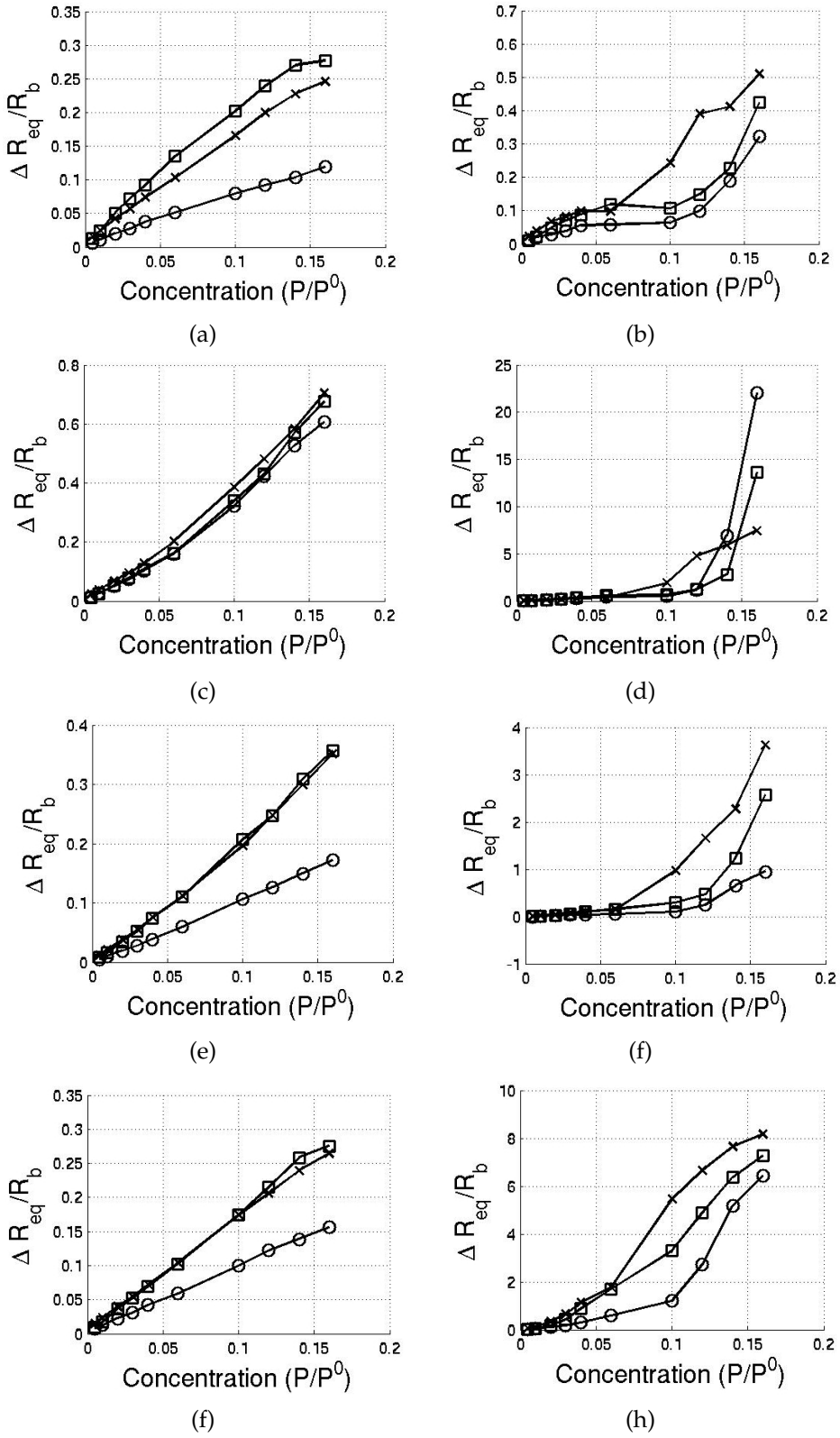
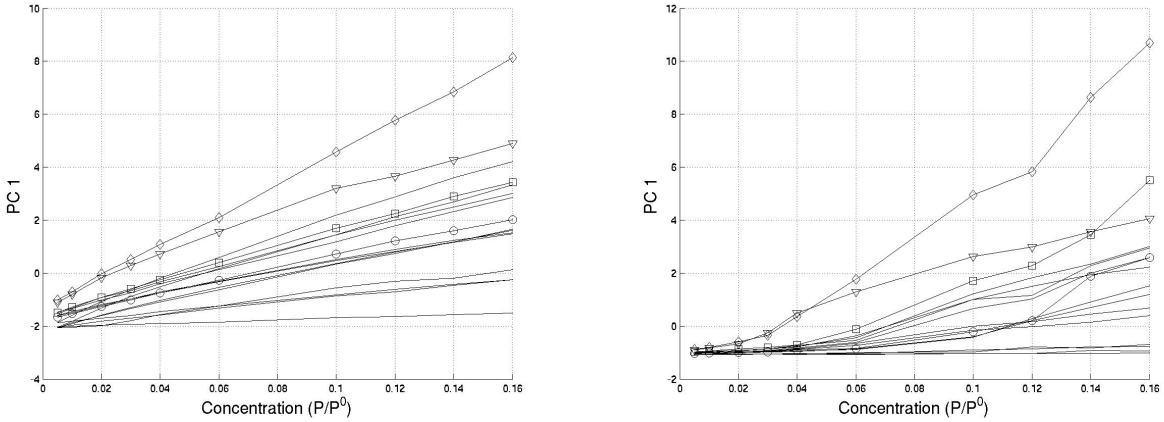


Figure 3.3: PCA Analysis of Sensor Response Data Collected From 16 Analytes in Table 3.1. Marked are four selected analytes: isopropyl benzene (○), water (□), chloroform (◇), and methyl acetate (▽).



(a) (b)

may be possible on a simple sensor by sensor basis for example, 1-chlorobutane and isopropyl benzene are easily recognized as displaying significant nonlinear behavior in the fourth and least significant principal component at $P/P°$=0.06, while the first principal component does not display such behavior until $P/P°$=0.10.

3.4.3 Determination of Signal-to-Noise Characteristics

The goal of using percolative, low carbon-black sensors is primarily to generate sensors capable of generating stronger responses than the higher carbon-black sensors typically used. Enhanced responses, however, can be undone if the noise inherent in such sensors is also significantly higher. To this end, we compared signal to noise ratios (S/N) for high and low carbon-black sensors as a function of concentration. Signal was taken from ΔR_{eq} values, and noise determined from three times the de-trended baseline standard deviation, in accordance with standard practices.[33] S/N values were derived from each exposure to all analytes of the first data collection period.

As would be expected, the high carbon-black sensors show a roughly linear dependence of S/N vs. concentration, while the low carbon-black sensors show a generally exponential dependence. At low concentrations, the high carbon-black sensors show higher average S/N, while the low carbon-black sensors show higher S/N at higher concentrations. The curves cross near $P/P°$=0.06, roughly in the region at which the low carbon-black sensors tend to begin displaying percolative behavior.

3.4.4 Analysis of Drift, Scatter, and Analyte History upon Sensor Responses

Prior work in our group showed that, for well broken-in sensors containing relatively high fractions of carbon-black, sensor drift rarely causes irrevocable problems resulting in analyte misidentification.[27] What drift does occur is largely due to changes in the environment of the sensors or due to long periods without using the sensors. Additionally, so long as the sensors are not exposed to unduly high concentrations of analyte and are allowed to return to baseline before the next analyte exposure, analyte history plays little role upon sensor responses. For this work, however, we used higher concentrations of analyte and using very low fractions of carbon-black, and expected that the sources and character of the variance in the system might change.

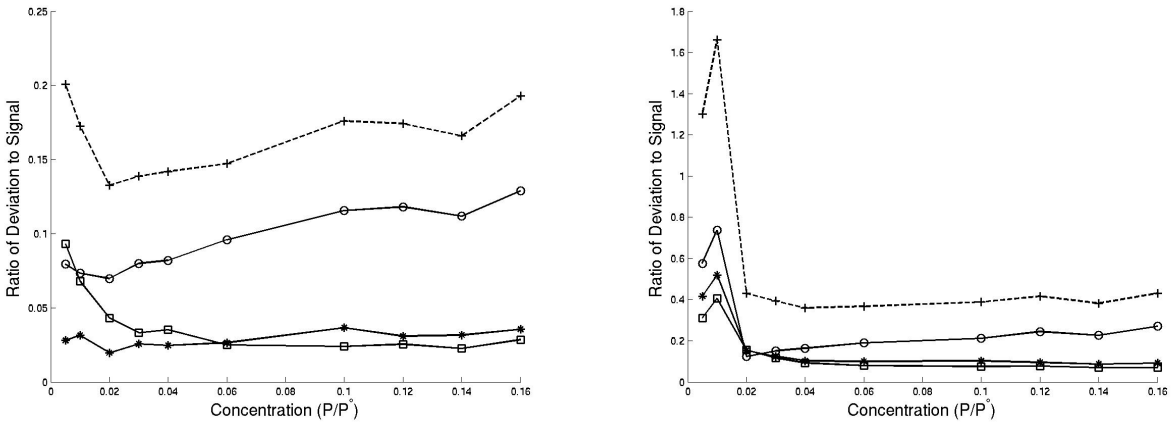
To determine the source of variation in groups of data derived from exposures of a given analyte at a fixed concentration to a certain sensor, we sough to establish the source of this variation namely, whether it derived from history or drift effects (which are correctable) or from scatter (which is not). We considered data from the first four runs (Table 3.1), which were comprised of the same analytes exposed in the same order. Thus, even in the presence of history effects, the data acquired from each run should (in principle) yield the same responses. As such, any disparities between the four runs will be due only to drift and scatter. For this work, it is assumed that little drift occurs within a run (consistent with prior work[27]), so drift is determined to be the difference in the mean response change of the responses to an analyte between runs. With the effects due to analyte history and drift thus determined, any further variance is assumed to arise from scatter or white noise.

Figure 3.4 a) and b) display the contributions of drift, analyte history, and scatter for the group of high carbon-black and low carbon-black sensors, respectively, between the first two runs (separated by two weeks). For both the high and low carbon-black sensor sets, drift makes up the primary component of variance, with analyte history and scatter approximately equal. For the low carbon-black sensors, the scatter drops to approximately 8% of the sensor responses, on average, at higher concentrations, with a greater contribution at lower concentrations. For the high carbon-black sensors, scatter remains approximately 3-4% of the total sensor response. After further months of use including responses to high concentrations of various analytes the sensors showed more significant drift, although the scatter and history effects remained consistent.

3.4.5 Comparison of Optimal Effectiveness of Percolative and Non-percolative Sensors

Though it is clear that low carbon-black sensors generally show less reproducibility than sensors with greater conductor fractions, this knowledge must be weighed against the extra information that percolative sensors can contribute. Table 3.2 displays correlation matrices for data derived from the first run of the PEO, PEP, PEVA, and PVS sensors (high and low carbon-black) to iso-

Figure 3.4: Separation of Total Response Variation (+) into Response Drift (∘), Response History Effects (□), and Scatter (∗) for the high carbon-black (a) and low carbon-black (b) sensor arrays.



(a) (b)

octane, toluene, THF, iso-butyl acetate, 1-chlorobutane, chloroform, ethanol, and water. Table 3.2a displays this correlation matrix derived from the lowest concentrations ($P/P°$¡0.06) of each analyte (averaged over the 8 analytes used), and Table 3.2b data derived from the highest concentrations ($P/P°$¿0.10). Two results are clear: first, the high carbon-black sensors tend to correlate more highly. Second, this effect is most significant at the higher concentrations studied, where one would expect percolation behavior to be manifest. Note that even the highest concentrations studied in this work are not particularly high, and one would expect the low carbon-black sensors to generate even less correlated data at yet higher concentrations. Examination of classification performance of the arrays with and without low carbon-black sensors also suggests that the low carbon-black sensors can be of significant benefit. Fishers Linear Discriminant was used to distinguish among the eight different analytes used in the first data collection run, and was implemented on a pairwise basis considering all possible tasks of distinguishing between two analytes at the same concentration, and repeated for each of the concentrations at which data was collected. For each of these binary separation tasks, a resolution factor (*RF*) derived from the data.[34] Results of this examination show that using a sensor array comprised of only low carbon-black detectors yields greater *RF* values than an array of only high carbon-black detector array for 31.7% of the FLD separation tasks for those separation tasks that involve analytes with concentrations at least $P/P°$=0.06. Additionally, the low carbon-black array yielded *RF* values at least 50% higher than the carbon-black sensor array for 14.8% of the tasks, and 100% higher for 6.2% of the separation tasks. The respective ratios for separation tasks below $P/P°$=0.06 are 3.0%, 0.67% and 0.17%, respectively. Not coincidently, $P/P°$=0.06 is approximately where many of the analytes investigated in this study begin to exhibit nonlinear behavior (Figure 3.2). As such, for distinguishing among analytes that exhibit nonlinear,

Table 3.2: Correlation matrix between sensors for analytes from the 8 analytes of the first data collection run.

(a) Concentrations $P/P^{\circ} < 0.06$

	PEO(2%)	PEP(12%)	PEVA(7%)	PVS(1%)	PEO(40%)	PEP(40%)	PEVA(40%)	PVS(40%)
PEO(2%)	1	0.731	0.820	0.876	0.959	0.857	0.960	0.949
PEP(12%)		1	0.505	0.619	0.732	0.948	0.732	0.854
PEVA(7%)			1	0.845	0.843	0.644	0.845	0.769
PVS(1%)				1	0.906	0.774	0.942	0.870
PEO(40%)					1	0.862	0.983	0.968
PEP(40%)						1	0.874	0.957
PEVA(40%)							1	0.963
PVS(40%)								1

(b) Concentrations $P/P^{\circ} > 0.10$

	PEO(2%)	PEP(12%)	PEVA(7%)	PVS(1%)	PEO(40%)	PEP(40%)	PEVA(40%)	PVS(40%)
PEO(2%)	1	0.452	0.924	0.819	0.930	0.650	0.960	0.834
PEP(12%)		1	0.368	0.626	0.412	0.729	0.454	0.531
PEVA(7%)			1	0.713	0.880	0.533	0.916	0.741
PVS(1%)				1	0.883	0.923	0.893	0.956
PEO(40%)					1	0.748	0.984	0.945
PEP(40%)						1	0.753	0.905
PEVA(40%)							1	0.929
PVS(40%)								1

percolative sensor responses, using low carbon-black sensors is beneficial. In other cases, the lessened reproducibility of the low carbon-black sensors typically causes arrays utilizing them to be less effective.

The previous FLD analysis assumes that reasonable calibration curves are available for all analytes to allow identification instead of simply resolution. Assuming such robust calibration exists, then classification performance would follow FLD *RF* values. In practice, all calibration curves are necessarily coarse, derived from data collected at a discrete set of concentrations. For high carbon-black sensors, this is not a problem, as their linear nature allows a one-point calibration to be in practice continuous. However, it is a much greater problem for low carbon black sensors, which do not follow a strict parametric form, requiring a full calibration curve be derived. As such, the ability to classify an unknown anlayte correctly depends on both its differentiability with respect to potential interferents as well as the coarseness of the calibration curve. To test the effect of this limitation, we considered a nearest-neighbor[34] approach, considering each of the 80 analyte/concentration combinations as a separate analyte. For each analyte exposure, its distance to each of the other 79 analyte/concentration clusters was measured, specifically excluding the cluster derived from that exposure and replicates. If the nearest cluster to a given analyte exposure was a different concentration of the same analyte, effectively treating the other 9 clusters of that analyte as a 9-point calibration curve, the classification was considered successful.

Figure 3.5 shows the results of this analysis as a function of concentration, for classification using low carbon-black sensors (○) and only high carbon-black sensors (□). Here, the low carbon-black sensors are revealed to be at a significant disadvantage, while they performed very competitively in many cases when considering only resolution. The reduced performance of low carbon-black sensors in terms of classification ability compared to resolution is due in large part to the nonlinearity of the low carbon-black sensors the strong responses that make them so desirable also spreads consecutive calibration points much farther apart compared to high carbon-black sensors. The magnitude of this conclusion, however, is peculiar to this study, namely upon the coarseness of the calibration and spacing of the calibration concentrations used.

3.4.6 Use of Percolative Sensors as "Binary" Vapor Sensors

Despite the advantages that high carbon-black sensors possess regarding noise, this is only relevant when using an analog-to-digital converter that possesses the resolution to measure sensor responses at their full precision. If this is not the case, and digital noise is significant compared to the noise of the sensors themselves, then any advantage derived from the low noise of high carbon-black sensors is lessened. If the digital noise of the electronics is actually higher than the noise of the low carbon-black sensors, then such advantages are completely irrelevant. In such cases, the sensor that delivers the greatest signal will be preferred. This is nearly always the low carbon-black sensor.

Figure 3.5: Classification efficiency of low carbon-black (∘) and high carbon-black (□) sensor arrays using a coarse calibration curve

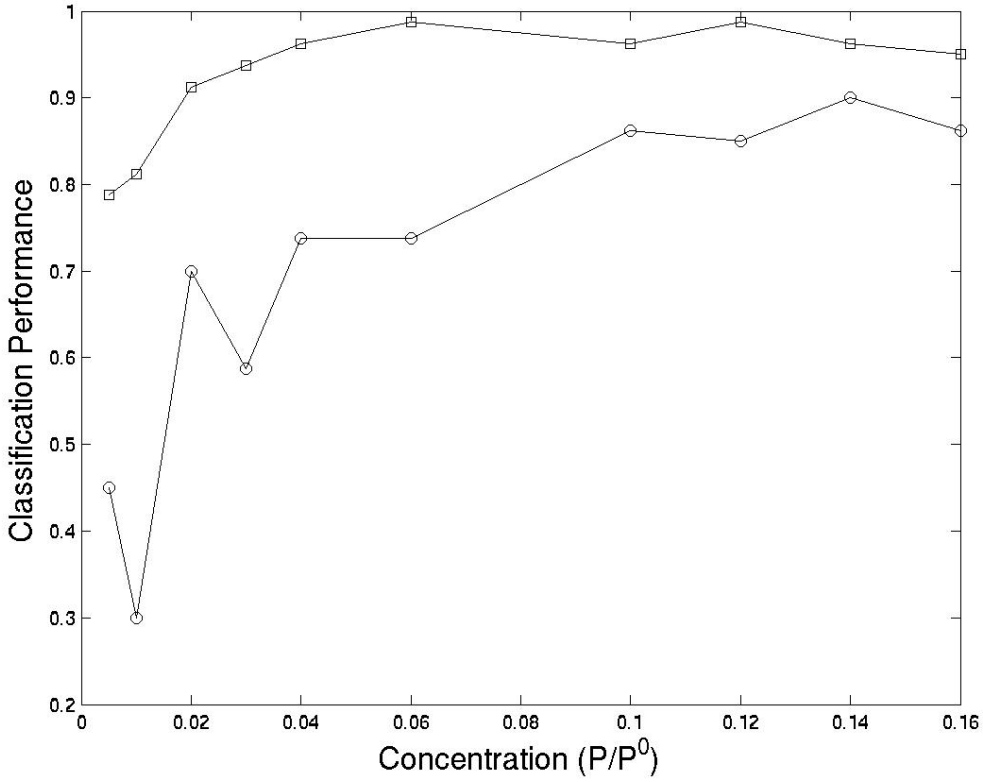


Table 3.3: Correlation matrix between sensors for analytes from the 8 analytes of the first data collection run.

	I-octane	Toluene	THF	Bu-Acetate	Cl-Butane	$CHCl_3$	Ethanol	Water
PEO	0.131	0.142	0.090	0.143	0.104	0.062	N/A	N/A
PEP	0.116	0.122	0.093	0.148	0.101	0.064	N/A	N/A
PEVA	0.138	0.137	0.102	0.154	0.108	0.050	N/A	N/A
PVS	0.040	0.030	0.017	0.038	0.022	0.011	0.150	N/A

N/A *denotes an analyte that did not reach the percolation threshold over the concentration range tested.*

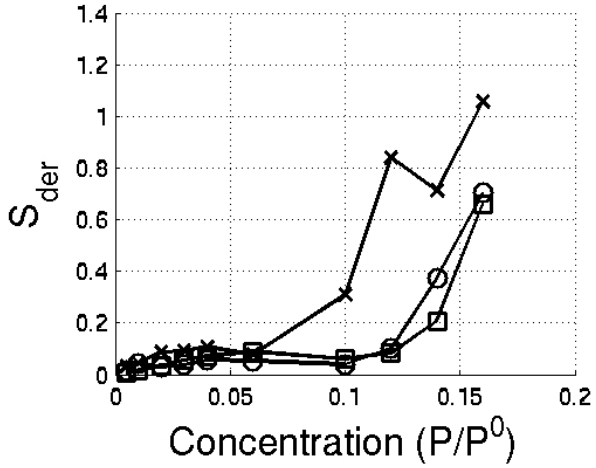
Considering the most extreme case, one might seek a sensor that can yield meaningful results when using even a 1-bit conversion, assuming only that we have the sensor in a circuit that allows it to have a variable "turn-on" resistance. In this case high carbon-black sensors would be generally useless, as their very linearity would provide greatly reduced signal and no convenient point (such as a percolation threshold) at which to designate the "turn-on" point. For low carbon-black sensors, the enhanced response and nature of the percolation threshold could allow low carbon-black sensors to be treated as either "on" or "off."

Though the percolation threshold may not be perfectly defined in all cases, it is reasonable to presume that the point at which the responses of the high and low carbon-black sensors diverge represents the percolation threshold. Given only one exposure at a single concentration, however, one must guess what degree of difference is necessary between the sensors of the pair actually represents percolation behavior. To derive a useful metric from the relationship between the high carbon-black and low carbon-black sensors, we have applied the following relationship to determine a derived signal (S_{der}) from the signals S (here, $\Delta R_{eq}/R_b$ values) of a high/low carbon-black sensor pair (eq 3.3):

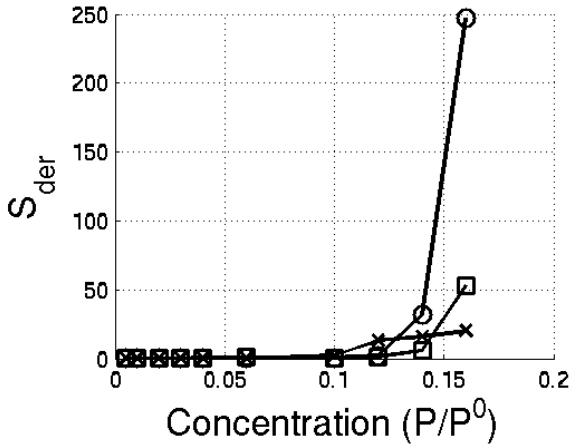
$$S'_{der} = \frac{S_{low}^2}{S_{high}} \tag{3.3}$$

Figures 3.6 a-d show S_{der} for PEO and PEP, PEVA, and PVS, respectively. For PEO, S_{der} values of less than 0.25 lie below the percolation threshold, and higher values above it. For PEP, PEVA, and PVS this threshold is sharper, and the threshold concentration may be set at 1.5, 2.0, and 1.0, respectively. Also, for a given analyte presented to a given sensor pair, the P/P° concentration required to generate an above-threshold S_{der} value varies and can be taken as characteristic of the particular analyte/polymer combination. Table 3.3 shows the S_{der} threshold for each sensor pair, and the P/P° concentration (estimated by interpolation) for each analyte required to exceed it. Another potential benefit of considering the highly nonlinear percolative sensors from a binary on/off standpoint is as a mimic of the mammilian olfaction system, which (at the lowest level) consists of many different receptor neurons, each of which fires when sufficiently stimulated by

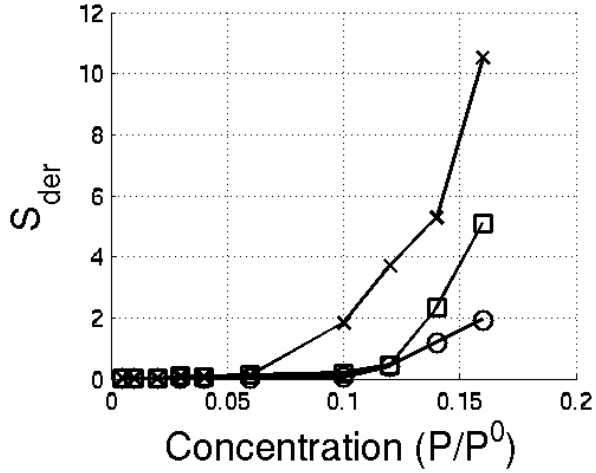
Figure 3.6: $\frac{S_{low}^2}{S_{high}}$ vs. analyte concentration for iso-octane, toluene, and tetrahydroduran



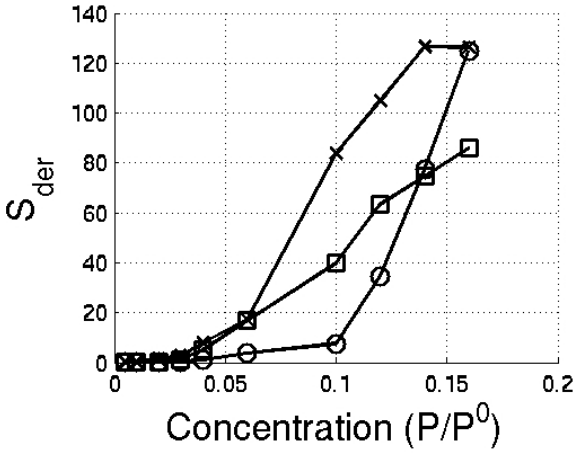
(a)



(b)

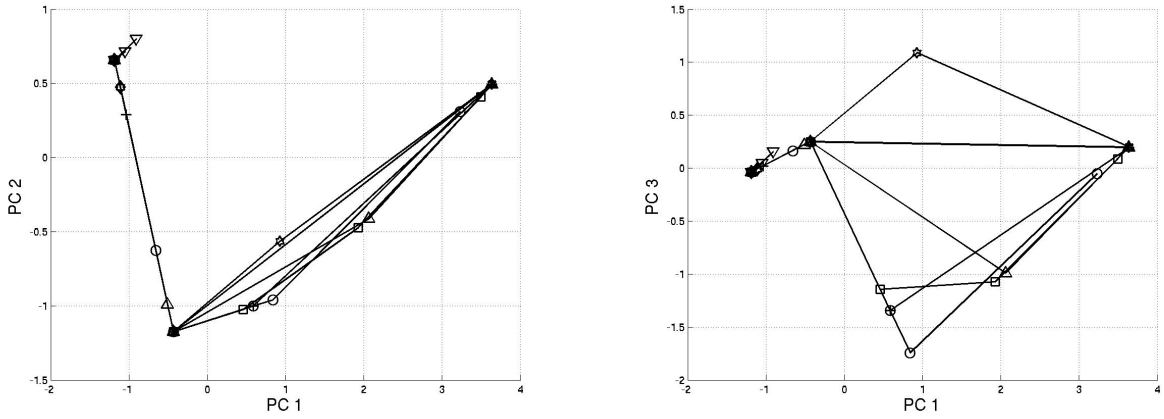


(a)



(b)

Figure 3.7: Principal Components Analysis of S_{der} Data Converted to Binary Format. Analytes are iso-octane (○), toluene (□), tetrahydrofuran (△), butyl acetate (×), chlorobutane (+), chloroform (★), ethanol (◇), water (▽)



(a) (b)

analyte vapor.[35] Taken together, a large array of percolative sensors would then generate, in effect, a bit vector for any analyte exposure, similar to how the response set generated by olfactory receptors. Figure 3.7 shows response data of the 8 primary analytes investigated in this study, first converted into binary form using the thresholds in Table 3.3 and then decomposed into its principal components. Figure 3.7 a) shows principal components 1 and 2, and b) shows principal components 1 and 3. Principal components 1 and 2 show an arc through the decomposed binary sensor space that is traced by all of the analytes; components 1 and 3 show the analytes sharply diverging near $P/P°$=0.06 at which point the sensors begin to show strong percolative responses to some analytes and then reconverge at $P/P°$=0.16, by which point many of the responses were percolative. These characteristics are somewhat biomimetic. Different regions of the brain have been shown to be responsible for determining odor intensity (amygdala) and identity (portions of the orbitofrontal cortex).[36] PCA acts as an analog to the processes that transform raw responses into different pieces of information that are ultimately represented as identity or concentration. Other work indicates that higher concentrations of a given analyte stimulate more regions of the glomerular layer, in addition to more olfactory receptors, which is also analogous to the sensor responses in this work.[37] The reconvergence effect displayed in PCA would indicate that chemicals should smell similar at very high concentrations. In the mammilian olfaction system, this has not found to be the case, in that the ability to distinguish chemicals increases with concentration.[38] This research does show, however, that chemicals are perceived differently at different concentrations, and suggests that many chemicals possess some common characteristics at high concentrations, namely, that they are perceived as "chemical" or "varnish" or similar unpleasant characteristics.

The authors do not speculate as to why this might be, but note the lack of trigeminal stimulation, which suggests that what effects exist are olfactory in nature.

3.5 Conclusions

Using low carbon-black sensors presents a variety of advantages when used in conjunction with high carbon-black sensors. They generate stronger responses, often better resolve analytes, and are generally more sensitive above low concentrations of approximately P/P°=0.06. These advantages are tempered, however, by the more significant degree of drift, loss of response vs. concentration linearity, and lessened reproducibility of the responses of such sensors. Even including such considerations, sensor arrays that contain fractions of carbon-black typically perform at least as well as high carbon-black sensor arrays, and often much better. Arrays of low carbon-black sensors also offer unique ways to mimic biological olfaction phenomena such as response saturation.

While low carbon-black sensors often present greater signal to noise ratios as compared to high carbon-black analogs given ideal signal conversion electrionics, less capable electronics may be unusable with high carbon-black sensors, as digital noise may overwhelm the responses of the sensors in such cases. In these situations, high carbon-black sensors will not function and only lower carbon-black sensors may be used at all to any degree of success.

Given the additional information they add and new use modes they present, low carbon-black sensors show promise for use in arrays also containing high carbon-black sensors. Their limitations will prevent them from actually replacing high carbon-black sensors in sensor arrays, but their benefits allow them to complement the function of high carbon-black sensors well.

Bibliography

[1] Ballantine, D.; Rose, S.; Grate, J.; Wohltjen, H. *Analytical Chemistry* **1986,** *58,* 3058–3066.

[2] Rose-Pehrsson, S.; Grate, J.; Ballantine, D.; Jurs, P. *Analytical Chemistry* **1988,** *60,* 2801–2811.

[3] Patrash, S.; Zellers, E. *Analytical Chemistry* **1993,** *65,* 2055–2066.

[4] Ronot, C.; Archenault, M.; Gagnaire, H.; Goure, J.; Jaffrezicrenault, N.; Pichery, T. *Sensors and Actuators B - Chemical* **1993,** *11,* 375–381.

[5] Rakow, N.; Suslick, K. *Nature* **2000,** *406,* 710–713.

[6] White, J.; Kauer, J.; Dickinson, T.; Walt, D. *Analytical Chemistry* **1996,** *68,* 2191–2202.

[7] Dickinson, T.; White, J.; Kauer, J.; Walt, D. *Nature* **1996,** *382,* 697–700.

[8] Dickinson, T.; Michael, K.; Kauer, J.; Walt, D. *Analytical Chemistry* **1999,** *71,* 2192–2198.

[9] Bartlett, P.; Archer, P.; Ling-Chung, S. *Sensors and Actuators* **1989,** *19,* 125–140.

[10] Gardner, J.; Pike, A.; Derooij, N.; Koudelkahep, M.; Clerc, P.; Hierlemann, A.; Gopel, W. *Sensors and Actuators B - Chemical* **1995,** *26,* 135–139.

[11] Shurmer, H.; Gardner, J.; P., C. *Sensors and Actuators B - Chemical* **1990,** *1,* 256–260.

[12] Freund, M.; Lewis, N. *Proceedings of the National Academy of Sciences of the United States of America* **1995,** *92,* 2652–2656.

[13] Lonergan, M.; Severin, E.; Doleman, B.; Beaber, S.; Grubb, R.; Lewis, N. *Chemistry of Materials* **1996,** *8,* 2298–2312.

[14] Severin, E.; Doleman, B.; Lewis, N. *Analytical Chemistry* **2000,** *72,* 658–668.

[15] Lang, H.; Baller, M.; Berger, R.; Gerber, C.; Gimzewski, J.; Battiston, F.; Fornaro, P.; Ramseyer, J.; Meyer, E.; Guntherodt, H. *Analytica Chimica Acta* **1999,** *393,* 59–65.

[16] Torsi, L.; Dodabalapur, L.; Sabbatini, L.; Zambonin, G. *Sensors and Actuators B - Chemical* **2000,** *67,* 312-316.

[17] Cornila, C.; Hierlemann, A.; Lenggenhager, R.; Malcovati, P.; Baltes, H.; Noetzel, G.; Weimar, U.; Gopel, W. *Sensors and Actuators B - Chemical* **1995,** *25,* 357–361.

[18] Gardner, J.; Shurmer, H.; Corcoran, P. *Sensors and Actuators B - Chemical* **1991,** *4,* 117–121.

[19] Corcoran, P.; Shurmer, H.; Gardner, J. *Sensors and Actuators B - Chemical* **1993,** *15,* 32–37.

[20] Watson, J. *Sensors and Actuators* **1984,** *5,* 29–42.

[21] Yamazoe, N. *Sensors and Actuators B - Chemical* **1991,** *5,* 7–19.

[22] Blythe, A. *Electrical Properties of Polymers;* Cambridge University Press: Cambridge, UK, 1979.

[23] Severin, E.; Lewis, N. *Analytical Chemistry* **2000,** *72,* 2008–2015.

[24] Sotzing, G.; Phend, J.; Grubbs, R.; Lewis, N. *Chemistry of Materials* **2000,** *12,* 595–595.

[25] Tillman, E.; Koscho, M.; Grubbs, R.; Lewis, N. *Analytical Chemistry* **2003,** *75,* 1748–1753.

[26] Briglin, S.; Freund, M.; Tokumaru, P.; Lewis, N. *Sensors and Actuators B - Chemical* **2002,** *82,* 54–74.

[27] Sisk, B.; Lewis, N. *Sensors and Actuators B - Chemical* **2004,** *104,* 249-268.

[28] Koscho, M.; Grubbs, R.; Lewis, N. *Analytical Chemistry* **2002,** *74,* 1307–1315.

[29] "Cabot Carbon Blacks for Specialty Applications", North American Technical Report S-136, Cabot Corporation Special Blacks Division, Billerica, MA, 2004.

[30] Doleman, B.; Sanner, R.; Severin, E.; Lewis, N. *Analytical Chemistry* **1998,** *70,* 2560–2564.

[31] Vaid, T.; Burl, M.; Lewis, N. *Analytical Chemistry* **2001,** *73,* 321–331.

[32] Burl, M.; Sisk, B.; Vaid, T.; Lewis, N. *Sensors and Actuators B - Chemical* **2002,** *87,* 130–149.

[33] Skoog, D.; Holler, F.; Nieman, T. *Principles of Instrumental Analysis, 5th Ed.;* Harcourt-Brace: Philadelphia, 1998.

[34] Duda, R.; Hart, P. *Pattern Classification and Scene Analysis;* John Wiley and Sons: New York, 1984.

[35] Buck, L.; Axel, R. *Cell 65,* 175-187.

[36] Anderson, A.; Christoff, K.; Stappen, I.; Panitz, D.; Ghahremani, D.; Glover, G.; Gabrieli, J.; Sobel, N. *Nature Neuroscience* **2003,** *6,* 196–202.

[37] Johnson, B.; Leon, M. *The Journal of Comparative Neurology* **2000,** *422,* 496–509.

[38] Laing, D.; Legha, P.; Links, A.; Hutchinson, I. *Chemical Senses* **2003,** *28,* 57–69.

Chapter 4

Estimation of Chemical and Physical Characteristics of Analyte Vapors Through Analysis of the Response Data of Arrays of Polymer/Carbon-Black Composite Vapor Detectors

4.1 Abstract

Analysis of the signals produced by a collection of organic polymer/carbon-black composite vapor detectors has been performed to assess the ability to estimate various chemical and physical properties of analyte vapors based on information contained in the response patterns of the detector array. A diverse array of composite chemiresistive vapor detectors was exposed to a series of 75 test analytes that had been selected from among five different chemical classes: alcohols, halogenated hydrocarbons, aromatics, unsubstituted hydrocarbons, and esters. The algorithmic task of interest was to use the resulting array of response data to assign one of the five chemical class labels to a test analyte, despite having left that analyte out of the model used to generate the class labels. The *k*-nearest neighbor algorithm was employed for this task using either Euclidean or Mahalanobis distance calculations with raw data, or Euclidean distances with data preprocessed using Fisher's Linear Discriminant algorithm. Each data cluster that was produced by replicate exposures to an individual analyte was well resolved from all of the other 74 analyte clusters. Furthermore the analyte response clusters could be robustly grouped into supersets such that each of the five individual chemical classes was well-separated from every other class of analytes in principal component space. Up to 85% percent of the test analyte exposures were correctly assigned to their chemical

classes. The detector array response data also was found to contain semi-quantitative information regarding physicochemical properties of the members of the test analyte series, such as the degree of unsaturation of the carbon chain, the number of halogen atoms, and type of aromatic ring in the test analytes. Using multiple linear regression, quantitative information related to analyte sorption, such as water/octanol partition coefficients, dielectric constants, and molar volumes were also predicted. The performance in these types of tasks is relevant for applications of a semi-selective array of vapor detectors in situations when no prior knowledge of the analyte identity is available and when there is no assurance that the test analyte will have been contained in the training set database produced by a compiling a library of responses from the detector array.

4.2 Introduction

Arrays of broadly cross-reactive sorption-based detectors have received much recent attention. Typically the sorption detectors are either conductive polymers[1–3] or conducting polymer composites[4–7] polymers that have been impregnated with dyes whose absorption or luminescence signals are sensitive to their environments,[8–12] polymer films that have been coated onto surface or bulk resonating crystals,[13–15] or polymers that have been coated onto the ends of micromachined cantilevers.[16] In any of these architectures, an analyte elicits a response from many detectors, and in turn each detector responds to many analytes. Pattern recognition algorithms are then used to classify, and in some cases quantify, the analyte of interest.[17,18] Arrays of 5-20 different polymeric sorption detectors have been shown to provide excellent analyte classification and quantification characteristics in a variety of laboratory-based situations .[19–22]

Detectors of particular interest in our laboratory are composites that consist of regions of an electrical conductor and regions of an insulating organic polymer.[5,23] Swelling of the polymer by sorption of an analyte induces a reversible, characteristic change in the dc electrical resistance of the detector film. Arrays of such detectors have been shown to provide excellent pairwise resolution between both closely related and diverse analytes, easily resolving between pairs of homologous alkanes, homologous alcohols, H_2O vs. D_2O, or very similar binary analyte mixtures.[6,20,22,24] Thus, over the timescale of these laboratory experiments, once the detector system has been trained towards these particular odorants, it can readily identify, with a high probability of correct identification and a low rate of false positives, the identity of one of these vapors presented in a subsequent test exposure to the array.

In this work, we have focused on a different question than matching a response pattern to one of the patterns that is already known to be contained in the stored response database for the array of interest. In the present work, we assume that the analyte response information is not in the database, and wish to evaluate what can be deduced about the test analyte through analysis of

its array response signals. In many instances, for example, it would be sufficient to be able to classify the general characteristics of an unknown analyte in terms of a chemical classification as an aromatic, aliphatic, chlorinated hydrocarbon, alcohol, ester, or other designated chemical class grouping based on the presence of certain functional groups. Additionally, within such descriptions of analyte classes, it would be of interest to obtain an estimate of selected physicochemical properties of an analyte, such as the value of an analytes dipole moment, vapor pressure, and/or its molecular volume. Some of this information, such as functional group analysis, is routinely available through analysis of the infrared spectrum of an organic vapor. Other information however, such as molecular volume or substrate binding affinity,[25] is more likely to be probed directly by a sorption-based detector than by the amplitude and position of a molecular electromagnetic absorption or emission signal.

As shown in prior work,[1,4,6,7] on a sufficiently compositionally diverse array of vapor detectors, each single-component pure analyte will yield its own characteristic response cluster in odor space. The polymer/carbon-black composite detectors generally exhibit a response that is linear with analyte concentration, so that the response cluster for a given analyte is maintained for normalized data over a wide range of analyte concentrations.[6] Each response cluster can furthermore typically be differentiated with relatively high resolution from the response clusters produced by exposure of the detector array to all of the other analytes in the training set. The question of interest is whether a decision surface can be drawn in the resulting n-dimensional odor space (where n equals the number of detectors used) such that if a test analyte exposure falls inside of the decision surface, the test analyte can also be correctly identified with high probability as being a member of the same chemical class as the training set of analytes that is contained inside the decision surface. An example for a hypothetical two-dimensional odor space is shown in Figure 4.1, in which analyte X is correctly assigned to one of three possible analyte clusters, whereas analyte Y is not successfully assigned to a class. Additional class boundaries can in principle be formulated to describe other physicochemical properties of the analytes of interest, such as hydrocarbon unsaturation or aromatic ring type, and regression, and regression techniques can allow prediction of continuous variables based upon sensor responses, as shown in prior work.[25]

To evaluate these possibilities, arrays of chemically sensitive resistors based on composites of organic polymer and carbon black were exposed to a series of single-component organic vapors. The vapors were members of one of five distinct chemical classes, as indicated in Table 4.1. In a "leave-one-out" (LOO) approach,[17] subsets of these analytes formed the models and databases that were used in conjunction with data analysis algorithms to extract physicochemical information on the test analyte. Principal components analysis (PCA) and *k*-nearest neighbor (*k*-NN) analysis using both Euclidean and Mahalanobis distance calculations were evaluated for their ability to assign correctly the chemical class, degree of unsaturation, number of halogen atoms, and nature of the aromatic

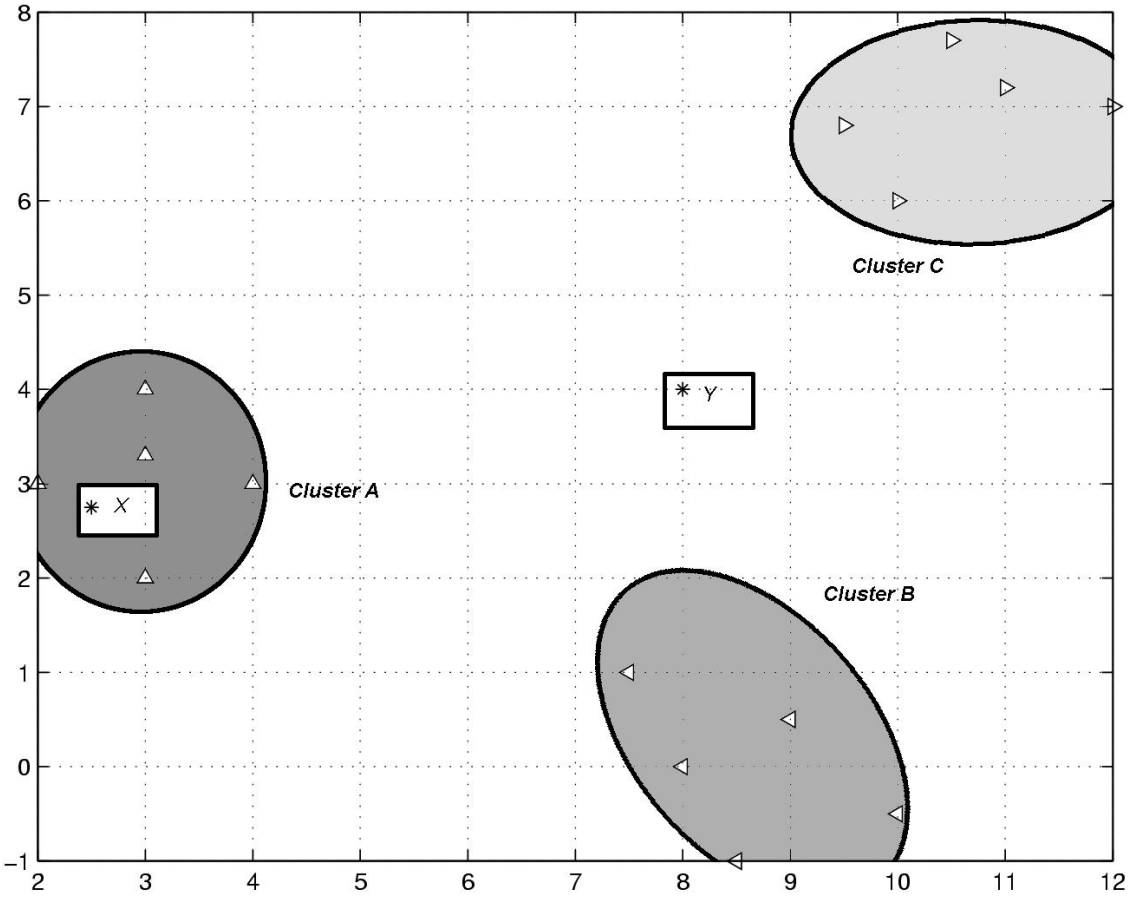


Figure 4.1: Description of clustering as implemented in this study. Three distinct analyte clusters (A, B, and C), consisting of five members each, are shown in a simulated principal components analysis plot. Each of the three clusters is completely separated from the others. Also included are two unknown test analytes, X, and Y. X is completely contained by the boundary that defines class A, and each of the three points nearest to X is a member of class A. Therefore, X is assumed to belong to class A. Y is well outside the boundaries of all three classes, and the analytes nearest to it do not correspond to the same analyte class. Therefore, Y is likely not a member of any of classes A, B, or C.

ring. Multiple linear regression was used to predict the water/octanol partition coefficient (K_{ow}), dielectric constant (ϵ),van der Waals size parameters, and Hildebrand[26] and Hansen[27] solubility parameters, for each of the 75 test analyte vapors. K_{ow} represents the partitioning of a liquid analyte into a water/octanol bilayer. Hildebrand solubility parameters (δ) are defined in eq. 4.1, Hansen in eq. 4.2.

$$\delta_t = (\frac{\Delta H_{vap} - RT}{V_m})^{1/2} \tag{4.1}$$

$$\delta_t^2 = \delta_e^2 + \delta_d^2 + \delta_h^2 \tag{4.2}$$

Here, δ_t represents the total solubility parameter, and δ_e, δ_d, and δ_h its electrostatic, dipole, and hydrogen bonding terms, respectively. ΔH_{vap} is the enthalpy of vaporization for an analyte, R the ideal gas constant, T the temperature, and V_m the liquid molar volume.

4.3 Experimental

The detector array consisted of 2 copies each of 20 compositionally distinct polymer/carbon-black composite chemically sensitive resistors (Table 4.2), for a total of 40 detectors. Detector films were cast from mixtures of 80% polymer and 20% by weight of carbon-black (Black Pearls 2000, Cabot Inc), as described previously.[5] The detector films were deposited between two Au leads that had been evaporated onto a glass slide, and the array was housed in a stainless steel assembly that was connected by teflon tubing to a computer-controlled, calibrated vapor generation and delivery system.

The set of 75 pure single-component analyte vapors was formed from approximately 15 members from each of five distinct analyte classes: alcohols, halides, aromatics, hydrocarbons, and esters (Table 4.1). Due to a limited number of solvent bubblers available in the experimental apparatus, data collection was divided into many runs, with each run consisting of exposures to 8 out of the 75 analytes. The first run consisted of 10 exposures of the detector array to each of analytes 1-8 from Table 4.1, with the analytes presented in random order. The next run was made up of 2 randomly selected analytes from run 1, as well as 6 new analytes, numbers 9-14 from Table 4.1. The selection process was then repeated again, and a new set of 8 analytes was created from a pair of analytes from the previous run and 6 new analytes. The selection process was continued until each analyte had been included at least once in a run. Then, another round of runs was performed with each run containing two randomly chosen analytes from 4 of the 5 analyte classes, and this process was repeated until each analyte had been included in a second run. In this way, each analyte was

Table 4.1: Analytes presented to the detector array.

Alcohols	*Halides*	*Aromatics*	*Hydrocarbons*	*Esters*
Methanol	1-Chloro-benzene	Benzene	Cyclooctane	Isopropyl acetate
Cyclopentanol	1-Bromobutane	Propyl benzene	n-Hexane	Butyl acetate
2-Butanol	Cyclohexyl-chloride	m-Xylene	n-Octane	Pentyl acetate
1-Pentanol	1,1,2-Trichloro-ethane	o-Xylene	n-Decane	Methyl acetate
2-Pentanol	1-Bromopentane	p-Xylene	3,3-Dimethyl 1-butene	Isobutyl acetate
3-Pentanol	3-Chloro 2-methyl propene	Isopropyl benzene	n-Heptane	trans-2-Hexenyl acetate
Isopropanol	1-Chloro-propane	Ethyl benzene	n-Nonane	Hexyl acetate
Ethanol	2-Chlorobutane	Toluene	Cyclopentane	Isopentyl acetate
1-Butanol	1-Fluorobenzene	1,2,4-Trimethyl benzene	2,2,4-Trimethyl-pentane	Ethyl propionate
2-Methyl-1-propanol	1-Iodopropane	2,6-Lutidine	Cyclohexane	Propyl acetate
3-Methyl-1-butanol	2-Bromo 2-methyl propane	2-Picoline	n-Pentane	sec-Butyl acetate
2-Methyl-2-butanol	1-Iodobutane	Pyridine	2,5-Dimethyl 2,4-hexadiene	Isopentyl propionate
2-Propen-1-ol	Chloroform	Anisole	2-Methyl-2-butene	Pentyl butyrate
1-Hexanol	Methylene chloride		7-Methyl 1,6-octadiene	Isopentyl benzoate
2-Methyl-3-buten-2-ol	1-Chlorobutane		1,7-Octadiene	Ethyl butyrate
			Cyclopentene	
			Cyclooctene	

Table 4.2: Polymers used to fabricate the polymer/carbon-black composite detector array.

1	poly(ethylene oxide)
2	Poly(ethylene oxide)-co-poly(amidoamine), diblock gen. 4[a]
3	Poly(ethylene-*co*-vinyl acetate) (45% vinyl acetate)
4	Poly(ethylene oxide)-*co*-poly(amidoamine), diblock gen. 1[b]
5	Poly(styrene-*b*-butadiene)
6	Kraton G[c]
7	poly(vinyl carbazole)
8	Kraton D[d]
9	poly(vinyl acetate)
10	poly(diphenoxyphosphazene)
11	polycaprolactone
12	polychloroprene
13	polysulfone
14	polyaniline-0.5-HDBSA[e]
15	poly(n-vinyl pyrrolidone)
16	bis(cyanoallyl polysiloxane)
17	poly(4-vinyl phenol)
18	poly(styrene-*co*-allyl alcohol)
19	poly(methyl octadecyl siloxane)
20	ethyl hydroxyethyl cellulose

[a]PEO-PAMAM diblock copolymer, with 5000 MW linear PEO and generation 4.0 PAMAM dendrimer (total MW=8420
[b]PEO-PAMAM diblock copolymer, with 5000 MW linear PEO and generation 1.0 PAMAM dendrimer (total MW=5230
[c]Commercial Polymer from Shell Corp.
[d]Commercial Polymer from Shell Corp.
[e]Polyaniline with a 0.5 fraction of all amine sites protonated by hexadecyl benzene sulfonic acid (HDBSA

presented to the detector array on at least two different occasions. Analytes were all presented at a fixed (0.020) fraction of their vapor pressure at room temperature, 21±1°C, to insure a constant vapor phase analyte activity throughout the runs.

Each analyte exposure consisted of a three-minute pre-exposure period to allow measurement of a stable baseline resistance, followed by a five-minute period of analyte flow during which the steady-state differential resistance change of the detector was recorded. A seven-minute post-exposure period allowed the detector resistances to return to baseline after each analyte exposure. Resistance data were recorded using a multiplexing Keithley multimeter and a data acquisition computer as described previously.[20]

Baseline correction of the data was performed by fitting a regression line to the first 10 points of the pre-exposure resistance readings, and correcting all subsequent data points by the difference in the value of the regression fit at the time of the measurement of that data point and at $t = 0$. A single descriptor, the relative differential resistance change, $\Delta R_{eq}/R_b$, was used in the analysis of the response of each detector to an analyte exposure. The resistance values upon analyte exposure were measured as the average over 10 data points after reaching equilibrium. ΔR_{eq} was measured as the difference in resistance between the equilibrium response states before (R_b) and during analyte

exposure. Each analyte exposure therefore produced a 40-dimensional vector, as follows:

$$X = \sum_{n=1}^{40} c_i x_i \tag{4.3}$$

Prior to quantitative data analysis, principal components analysis was used to visualize portions of the unnormalized, 40-dimensional detector array response data. The first two principal components contained 66% of the total variance of the 40-dimensional data while 76% of the total variance was contained in the first three principal components; thus, the data were visualized using the first three principal components as axes. The three-dimensional data clouds were rotated manually while the data were viewed along a fixed axis to assess the separation between clusters for various tasks of interest.

The *k*-nearest-neighbor (*k*-NN) approach was used to obtain a quantitative measure of analyte classification ability in different tasks. First, the mean response vectors for each of the 75 analytes were calculated by averaging the unnormalized array responses recorded during the replicate exposures to each analyte. A leave-one-out approach was then used, and a model data set was formed from the mean response vectors produced by exposures to 74 of the 75 total analytes. No response data from the analyte of interest was included in the construction of this model database of response vectors. For each individual exposure to the test analyte of interest, up to 7 of the nearest mean response vectors in the model database were then identified. The procedure was repeated for each of the 20 exposure data points for the analyte of interest. Finally, the entire process of model database construction, excluding data for the analyte of interest, and assessing distances to other mean analyte response vectors in the database, was repeated for each of the 75 analytes studied in this work.

Distance measurements in the *k*-NN analysis were made using both Euclidean distances and Mahalanobis distance methods applied to raw data, as well as Euclidean distances based upon data preprocessed using Fisher's Linear Discriminant (FLD).[17] The Euclidean values were calculated simply by determining the distances between two points in the 40 dimensional space, with no prior scaling or normalizing of the $\Delta R_{eq}/R_b$ response data values:

$$r^2 = (x - \mu)^T (x - \mu) \tag{4.4}$$

FLD is used to to preprocess the data by retaining information that successfully separates data clusters, while rejecting information that does not. The 20 exposures to a given analyte were treated as an individual data cluster, resulting in 75 total clusters from the 75 unique analytes. For this analysis, it was assumed that it was known which analyte exposures were mutually replicate, meaning that we did not seek to make this an analyte recognition problem but rather a class

recognition problem. As such, no validation scheme was used for FLD. Because more data clusters were used than available features, FLD returned a new 40-dimensional dataset. For this work, the least significant 5 dimensions - those which contain data least capable to separate the data clusters - were rejected, and the other 35 were used for regression.

Mahalanobis distances differ from Euclidean distances in that they are calculated on data for which each of the 40 individual descriptors is first autoscaled across the entire data set. When autoscaling, each of the 40 dimensions is mean-centered, then divided by the variance of that particular dimension:[17]

$$r^2 = (x - \mu)^T \Sigma^{-1} (x - \mu) \tag{4.5}$$

The Mahalanobis distance approach has the advantage in principle that noisier dimensions (detectors) do not unduly dominate the distance calculated between two data points in the 40-dimensional response space. However, it is possible for this data transformation to introduce artifacts in certain cases.

Each analyte exposure was then assigned a class identity using each of these three methods. The class of each test data point was assigned to the class that was represented by the majority of the nearest k mean response vectors, with k varying as 1, 3, 5 or 7. Each analyte was assigned to only be a member of a single class (Table 4.1). No class assignment was made to a data point in instances when the nearest neighbor data points selected did not produce a majority class consensus.

To predict quantitative analyte properties, data were first preprocessed using the FLD-preprocessed data. The FLD data was then fit to a multiple linear regression model used with a LOO cross-validation scheme. For each of the quantitative properties investigated, not all of the FLD features were used to build the model; rather, the N most significant FLD features were used, and the value N chosen was that which maximized the significance of fit for the given model.

Quantitative analyte properties to be predicted included properties likely to be related to analyte sorption into polymer matrices, such as water/octanol partition coefficients (K_{ow}), dielectric constants (ϵ), van der Waals volume and area, Hildebrand solubility parameters (δ_t), and the Hansen electrostatic (δ_e) and dipole (δ_d) components of the Hildebrand solubility parameter. The Hansen hydrogen bonding component (δ_h) was not analyzed because it was nonzero only for the alcohols. Values for K_{ow}, ϵ, and δ_t were taken from previous experimental results, and the van der Waals parameters were determined from experimental values.[28] If a value for an analyte/property combination was not available, that analyte was omitted from analysis for that property only. Hansen parameters were determined experimentally using the Cerius^2 and Jaguar software packages.

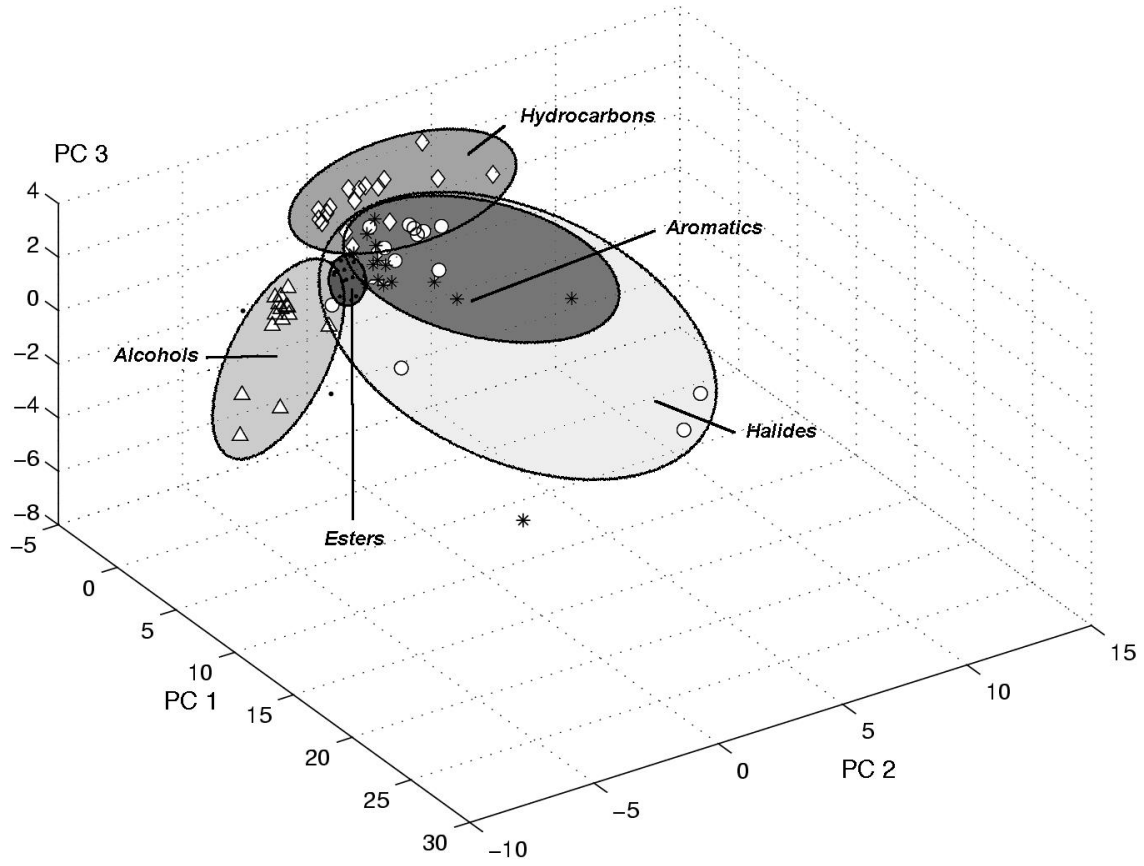


Figure 4.2: Principal components analysis plot of the mean response vector termini for each analyte used in the study, each of which corresponds to one of five analyte classes: alcohols, halides, aromatics, hydrocarbons, and esters. Approximate classification boundaries have been drawn around each of the five classes.

4.4 Results

4.4.1 Class Assignment

Figure 4.2 presents a plot of the analyte response data for all 75 analytes as projected onto the first three principal components of the 40-dimensional detector $\Delta R_{max}/R_b$ response space. For clarity, only the mean response vector termini, obtained from the average of multiple exposures to each analyte, are displayed. Each of the 75 different analytes produced a response cluster that was clearly separable from the response cluster of every other analyte investigated in this work. As is apparent from Figure 4.2, the mean vector response termini clustered into four well-defined, mutually separated regions. Three of these regions individually contained the mean vector response termini for the alcohols, hydrocarbons, and esters, respectively. In contrast, substantial overlap was present between the region that contained vector response termini produced by exposure of the detector array to halides and the regions containing either aromatic or ester organic vapors. The class

Table 4.3: Fractions of analyte exposures correctly classified using *k*-nearest neighbor analysis.

Neighbors		Mahalanobis (Raw)	Euclidean (Raw)	Euclidean (FLD)
1	Correct	0.777	0.789	0.951
1	Incorrect	0.223	0.211	0.049
1	Non-classified	N/A	N/A	N/A
3	Correct	0.786	0.771	0.888
3	Incorrect	0.161	0.177	0.081
3	Non-classified	0.053	0.052	0.031
5	Correct	0.775	0.755	0.810
5	Incorrect	0.155	0.149	0.122
5	Non-classified	0.069	0.096	0.068
7	Correct	0.754	0.740	0.792
7	Incorrect	0.137	0.150	0.153
7	Non-classified	0.109	0.110	0.055

assignment performance enabled by the detector array data was quantified using *k*-NN analysis (Table 4.3). Analytes that are members of two analyte classes (fluorobenzene, chlorobenzene, and isopentyl benzoate) were excluded from this analysis. The mean correct class assignment probability for the 72 single-class analytes tested was 0.76-0.82 using Euclidean distances and 0.76-0.81 using Mahalanobis distances for raw data, and 0.79-0.95 using Euclidean distances and FLD-preprocessed data. Classification rates were largely insensitive to whether k= 3, 5, or 7 nearest neighbors was used. The data in Table 4.3 suggested that using 7 neighbors instead of 3 decreased the correct classification rate more than the incorrect classification rate, and significantly increased the number of non-classified exposures. Classifications using 3 neighbors were preferable to those derived from a single neighbor, however, as the increase in the non-classification rate resulted almost exclusively from a decrease in the incorrect classification rate. Therefore, the 3 nearest neighbor algorithm was used in all further analysis.

Tables 4.4 a and b report the performance by analyte class in the form of confusion matrices for the class assignments using Mahalanobis distances derived from raw data and 3 nearest neighbors, where each (X,Y) cell in the table indicates what fraction of exposures belonging to an analyte class X was assigned to the analyte class Y. Perfect classification performance would produce the identity matrix in this representation of the success of class prediction from the detector array response data. Table 4.4a shows the results of all five classes, while 4.4b displays the results having left the halides out of the analysis. The overall the correct classification rate for all 72 single-class analytes using 3 nearest neighbors and Mahalanobis distances was 0.80; excluding halides (retaining all alcohols, aromatics, hydrocarbons, and esters), this rate increased to 0.88. The analytes with the worst classification performances among the alcohols, aromatics, hydrocarbons,

Table 4.4: Confusion matrices developed from *k*-nearest neighbor analysis using three neighbors and Mahalanobis distances.

	Alcohols	Halides	Aromatics	Hydro.	Esters
(a)					
Alcohols	0.9733	0	0	0	0.0267
Halides	0.0077	0.4577	0.1654	0.0192	0.2000
Aromatics	0	0.0885	0.8154	0.0077	0.0346
Hydrocarbons	0.0147	0.0265	0.0324	0.8441	0.0235
Esters	0.0500	0.0714	0.0036	0	0.8464
(b)					
Alcohols	0.9767		0	0	0.0233
Aromatics	0		0.8538	0.0192	0.0808
Hydrocarbons	0.0147		0.0588	0.8382	0.0500
Esters	0.0821		0.0143	0.0036	0.8714

and esters corresponded to isopropyl benzene, cyclopentene, 2,5-dimethyl 2,4-hexadiene, methyl acetate, and trans-2-hexenyl acetate, which resulted in error rates of 1.0, 0.80, 0.80, 1.0, and 0.7, respectively. Removing these five analytes (as well as the halides) from the set resulted in an overall correct classification rate (Mahalanobis distances, 3 neighbors) of 0.95. Thus, by removing 23 of the 75 analytes, the combined rates of non-classification and error was cut by three-fourths. Using FLD-preprocessed data, very few mistakes were made at all.

4.4.2 Determination of Chemical Information in Addition to Class Identity

The structure of the data displayed in Figure 4.2 suggests that classification of additional analyte properties should be possible, because the majority of the non-halide frequently misclassified analytes were polyfunctional, unsaturated, or had low molecular volumes relative to most members of their respective classes. Examples are provided by the responses of isopropyl benzene, cyclopentene (small, cyclic, unsaturated), 2,5-dimethyl 2,4-hexadiene (polyunsaturated), methyl acetate (small), and trans-2-hexenyl acetate (the only ester tested with an unsaturated hydrocarbon chain), as labeled in Figure 4.3. Consequently, further analysis was performed to determine whether such analyte-specific information could be isolated in a systematic fashion from the array vector response data. Figure 4.4 displays a principal components analysis plot of 20 exposures each of the detector array to n-hexane, n-heptane, n-octane, 2,2,4-trimethyl pentane, 2-methyl 2-butene, 2,5-dimethyl 2,4-hexadiene, 1,7-octadiene, and 7-methyl 1,6-octadiene. These analytes contain various degrees of unsaturation in the hydrocarbons yet minimize differences in their molecular weights. The broad, diffuse cloud of array response data in Figure 3 that was produced by the unsaturated hydrocarbons was clearly separable from the tighter cluster of response data that was produced by the saturated hydrocarbons. The only overlap between the two clusters (from the vector angle shown) arose from half of the exposures to 7-methyl 1,6-octadiene.

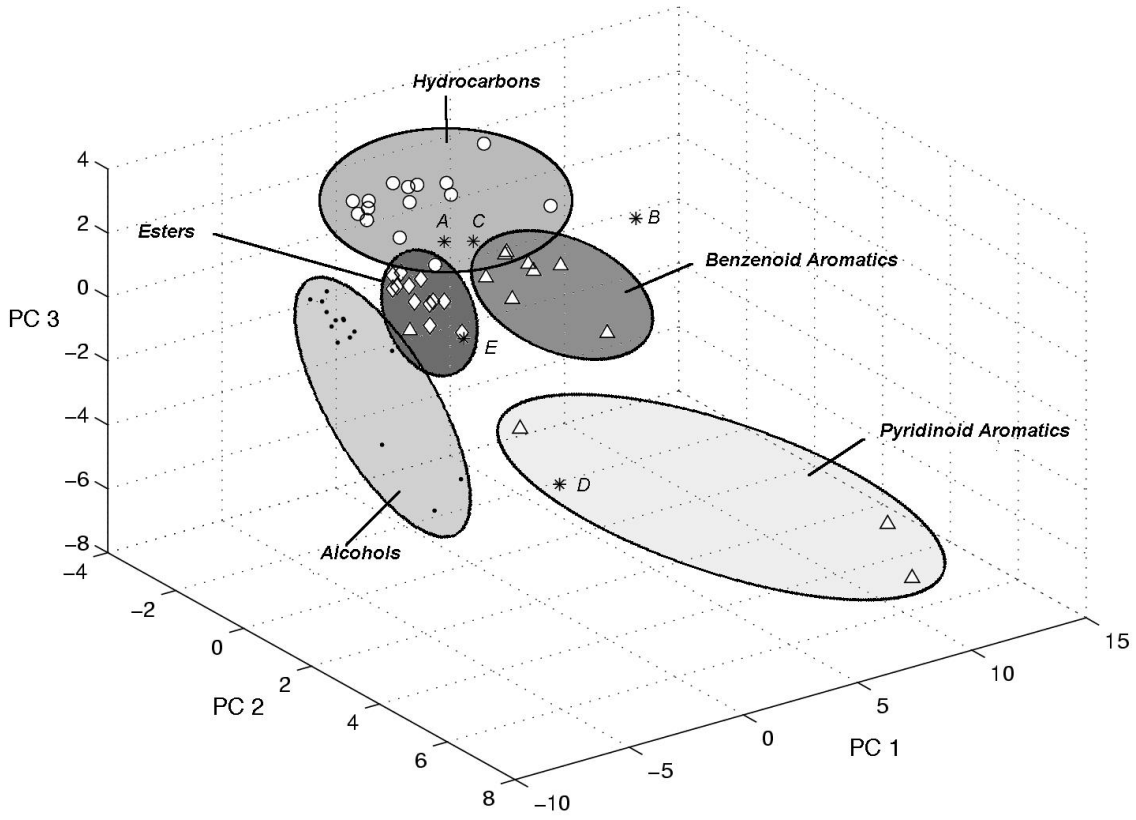


Figure 4.3: Principal components analysis plot of the mean vector termini for each alcohol, aromatic, hydrocarbon, and ester except those that were members of multiple classes (chlorobenzene, fluorobenzene, and isopentyl benzoate). The five analytes that were most frequently misclassified by *k*-nearest neighbor analysis using Mahalanobis distances and three nearest neighbors, isopropyl benzene (A), cyclopentene (B), 2,5-dimethyl 2,4-hexadiene (C), methyl acetate (D), and trans-2-hexenyl acetate (E), are specifically labeled in the plot as (∗).

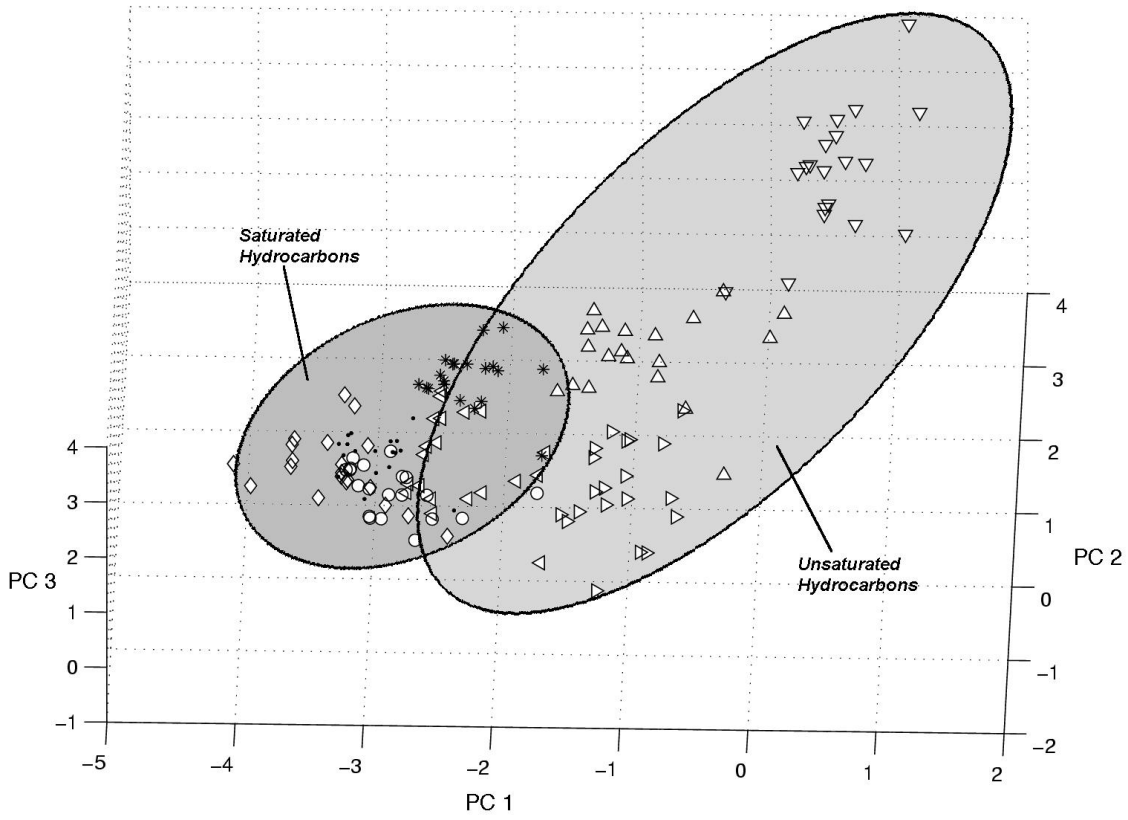


Figure 4.4: Principal components analysis plot of data collected from eight hydrocarbons: n-hexane, n-heptane, n-octane, 2,2,4-trimethyl pentane, 2,5-dimethyl 2,4-hexadiene, 2-methyl 2-butene, 7-methyl 1,6-octadiene, and 1,7-octadiene. All 20 exposures for each analyte are shown. The first four analytes, which are saturated, are represented in the plot by (∗),(●),◇, and ○, respectively. The last four, which are unsaturated, are represented by (△), (▽), (◁), and (▷), respectively. Sufficient separation is achieved between the two clusters to determine saturation with a success rate of over 80%.

A confusion matrix analysis of the *k*-NN determined (Mahalanobis, 3-nearest neighbors) classification performance in this task indicated that the fraction of saturated hydrocarbons correctly identified as such was 0.96; unsaturated hydrocarbons were correctly classified with a probability of success of 0.79. Hydrocarbons which produced responses that were located outside of a single, tight cluster were predicted, with a high degree of success, to be unsaturated. In fact, 10 of the 17 mistakes made in classification of the unsaturated hydrocarbons corresponded to 7-methyl 1,6-octadiene, the only unsaturated hydrocarbon that overlapped with the cluster of saturated hydrocarbons. Figure 4.5 displays a principal components analysis plot of 20 exposures to non-aromatic analytes that contained single or multiple halide groups. The multi-functional halides employed were chloroform, dichloromethane, and 1,1,2-trichloroethane; additionally, all non-aromatic halides (but excluding fluorobenzene and chlorobenzene) from Table 4.1 were included in the analysis. With the exception of four outlier data points that arose from some of the exposures to 1-iodopropane, the analytes that contained single halide functionality were well-separated from analytes that had multiple halide functional groups. This separation persisted despite a wide disparity in molecular weights (78.5-119 g mole^{-1}) and molecular volumes (81.4-126 Å^3) within the group of analytes that contained a single halide functionality. Additionally, although the multi-functional halides had no more than two carbons, a good deal of overlap existed between their molecular weight and molecular volume ranges and those of the monofunctional analytes (96.9-133.5 g mole^{-1} and 58.6-92.6 Å^3, respectively).

Figure 4.6 displays a principal components analysis plot of the clustering that arose from the mean vector response termini from each analyte in the halide and aromatic vapor sets. The clustering is divided into four categories: benzenoid aromatics, pyridinoid aromatics, mono-functional halides, and multi-functional halides. The pyridinoid analytes produced signals that were well-separated from the signals produced by the benzenoid analytes. The clustering of the benzenoid aromatics from the pyridinoids persisted despite the presence of anisole (methoxybenzene), which is similarly polar, so a simple polarity argument is not sufficient to explain the clustering. Furthermore, adding chlorobenzene and fluorobenzene to the set resulted in their responses falling well within the benzenoid aromatic/monofunctional halide cluster. The slight separation between the halide and aromatic clusters (Figure 4.2) was largely produced by the presence of the multi-functional halides and pyridinoid aromatics. The observation that the pyridinoid aromatics were easily separated from the benzenoid is not necessarily surprising, given the increased basicity of the pyridinoid structure, but it is interesting from a classification point of view that the two groups are reasonably well-separated from each other. Furthermore, it is surprising that the degree of overlap between the benzenoids and single-functional halides is so high, even more so than indicated by the PCA and *k*-NN results for aromatics relative to halides.

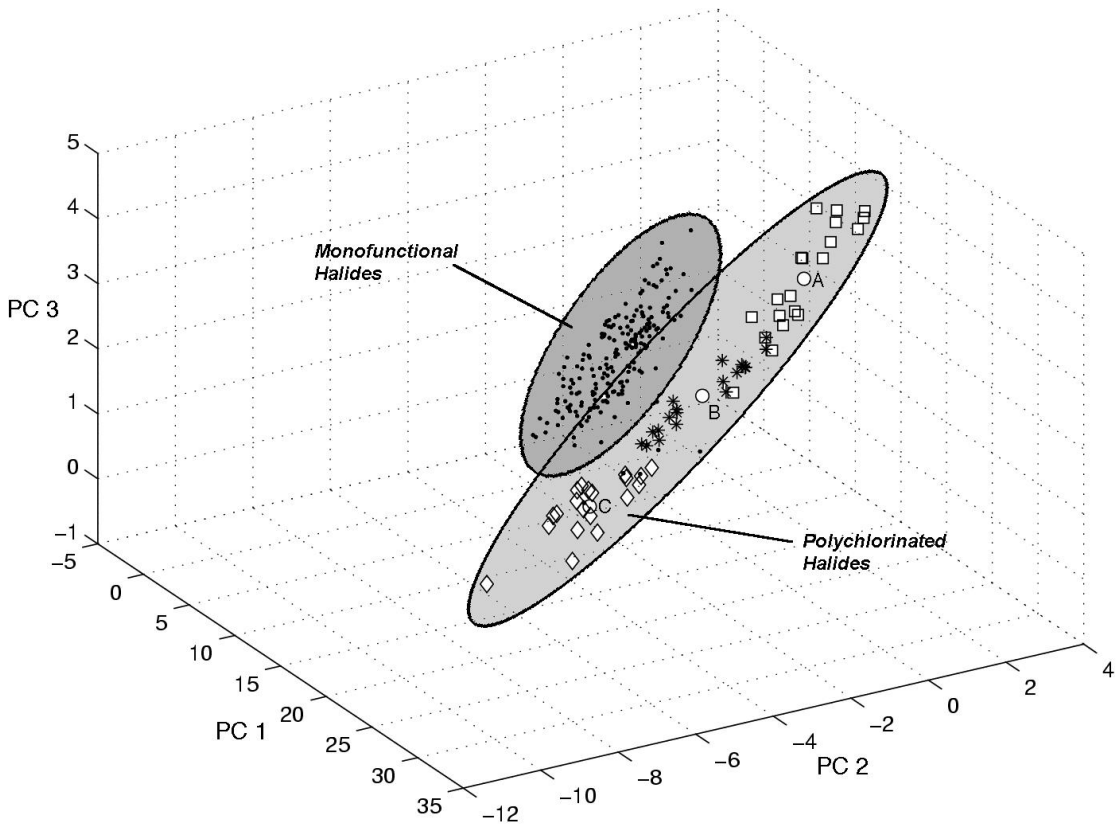


Figure 4.5: Principal components analysis plot of data collected from all of the halides used in this study (20 exposures each). The polychlorinated analytes, including 1,1,2-trichloroethane, dichloromethane, and chloroform are represented in the plot by (□), (∗), and (⋄), respectively. Mean vector response termini for the polychlorinated clusters are denoted by (◯), with A representing 1,1,2-trichloroethane, B representing CH_2Cl_2, and C representing $CHCl_3$. All of the other halides are represented by (•). The polychlorinated analytes are well-separated from the main group of halides.

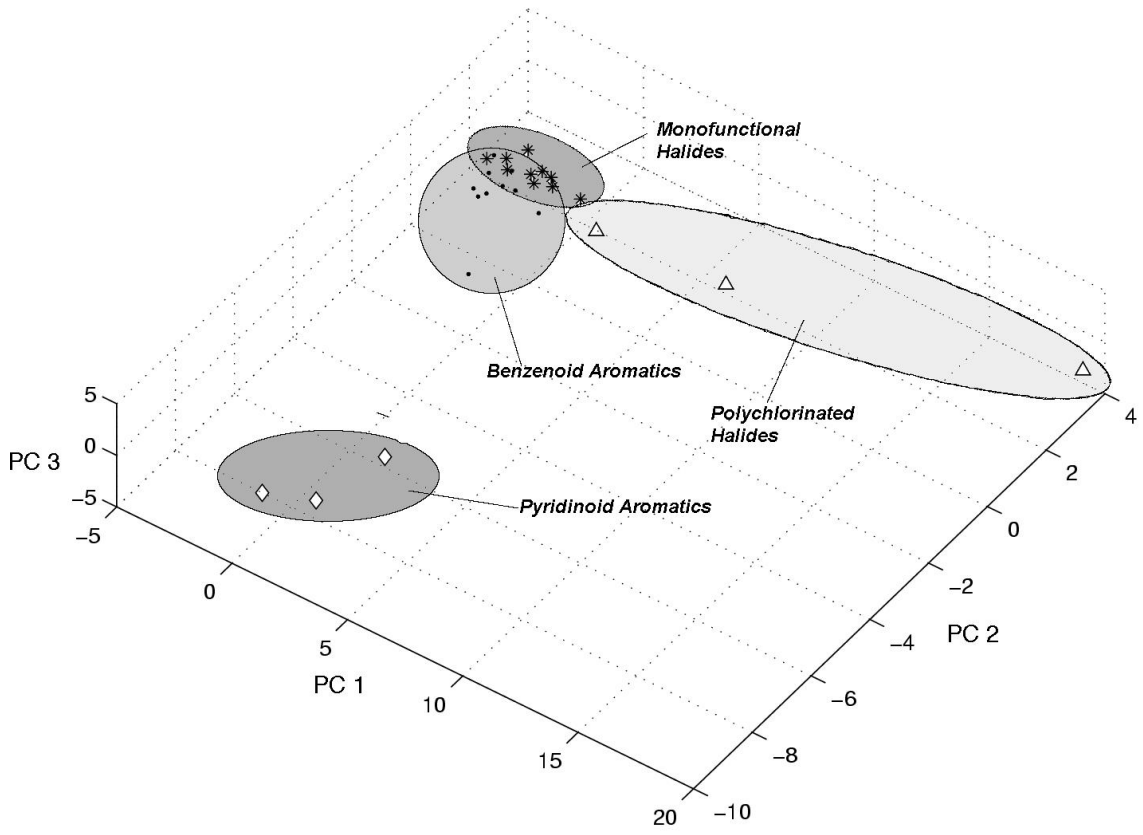


Figure 4.6: Principal components analysis plot of the mean vector response termini for the halides and aromatic analytes used in this study, excluding chlorobenzene and fluorobenzene. Pyridinoid aromatics ($\diamond$) are well-separated from benzenoid aromatics ($\bullet$), as are polyfunctional halides ($\triangle$) from monofunctional halides ($*$). The bulk of the separation between the halides and aromatics in this study arises from the presence of polyfunctional halides and pyridinoid aromaticsseparation between benzenoid aromatics and monofunctional halides is poor.

Table 4.5: Prediction results of six selected analyte properties.

Property Name	# FCs	RMSE/Range	r^2	Slope	Int./Range
K_{ow}	5	0.0783	0.867	1.00	6.0×10^{-4}
Van der Waals Volume	11	0.110	0.664	1.08	-0.0596
Van der Waals Area	10	0.122	0.611	1.07	-0.0543
Dielectric Constant	3	0.0793	0.835	0.978	0.00566
Solubility Parameter	9	0.0737	0.852	1.05	-0.0640
Hansen Dipole	11	0.130	0.595	1.04	-0.104
Hansen Electrostatic	5	0.176	0.535	1.07	-0.0256

4.4.3 Prediction of Physiochemical Values

Although class properties dominated the clustering of the principal component data derived from the analytes investigated, other properties could be predicted from the same response data by weighting each sensor response differently. Rather than determining the weights so that the resulting mappings represented the greatest portion of the variance, weights were determined so that the mappings corresponded with variables of interest.

Figure 4.7 shows predicted vs. actual K_{ow} values for the 75 analytes tested. The prediction was accomplished with an r^2 of 0.867, and the regression line exhibited a slope of 1.00, an intercept/range value of 6×10^{-4}, and a root mean squared error/range value (RMSE/range) of 0.07826. These data are summarized in Table 4.5. Figures 4.8-4.13 show similar results for dielectric constant, van der Waals volume and area, Hildebrand solubility parameter, and the Hansen dipole and electrostatic terms, respectively.

Fit effectiveness, as measured by r^2 values, ranges from poor (0.535 for δ_e) to the very good (nearly 0.9 for δ_t and K_{ow}). This analysis suggests that properties more fundamentally related to analyte sorption are more easily predicted by sensor responses.

4.5 Discussion

To our knowledge, little information is avaiable to date in a leave-one-out study protocol regarding whether mapping into functional groups and/or geometric descriptors of a molecule can be robustly performed from the response patterns produced by an array of semi-selective sorption-based vapor detectors. At sufficiently high analyte concentration, the responses of a variety of detectors are sufficiently distinct that unique identification is possible for most analytes. However, because the responses of semi-selective detectors by nature depend on a large number of factors, including molecular volume, branching, dipole, hydrogen bonding, aromaticity as well as many others, the ability to extract any one of these parameters to the exclusion of the others, or at least by limiting them, has not been fully elucidated to date.

Within a single analyte class, isolating certain variables (such as size or saturation) proved

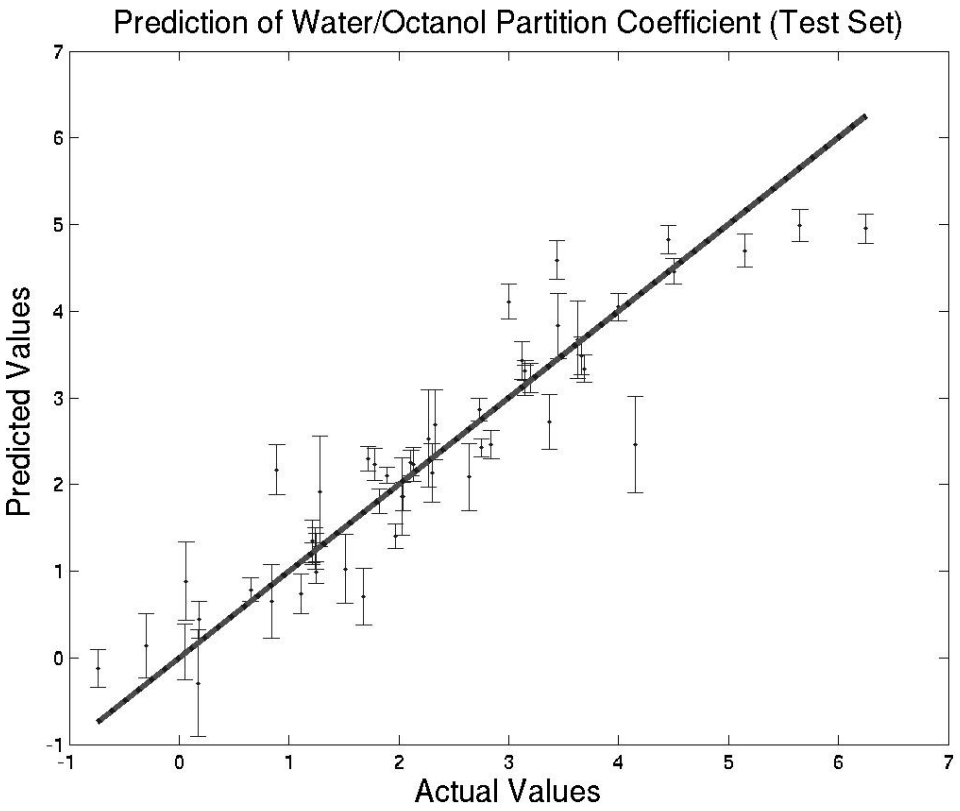


Figure 4.7: Predicted vs. actual K_{ow} values derived from a 5 Fisher component linear model. The solid line represents a best fit for the data; the dashed line is y=x. Error bars are 2σ.

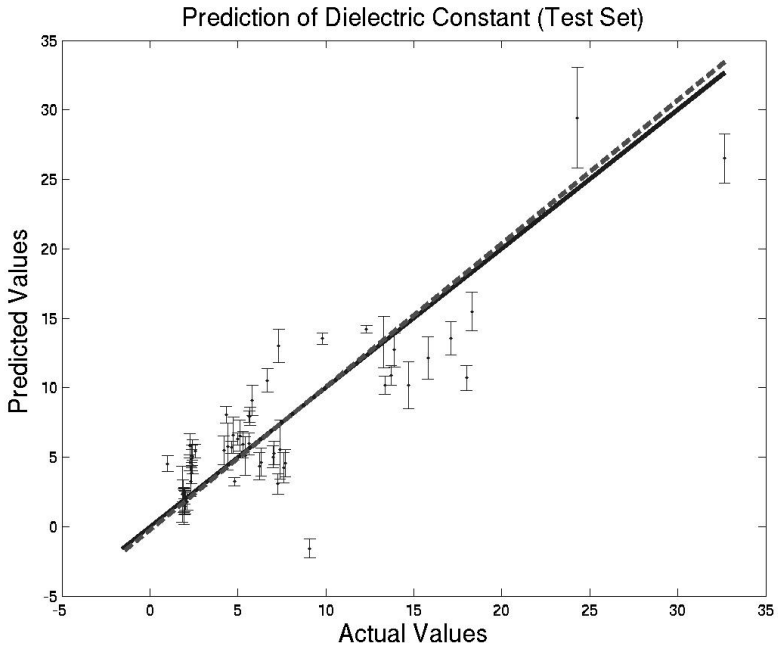


Figure 4.8: Predicted vs. actual ϵ values derived from a 3 Fisher component linear model. The solid line represents a best fit for the data; the dashed line is y=x. Error bars are 2σ.

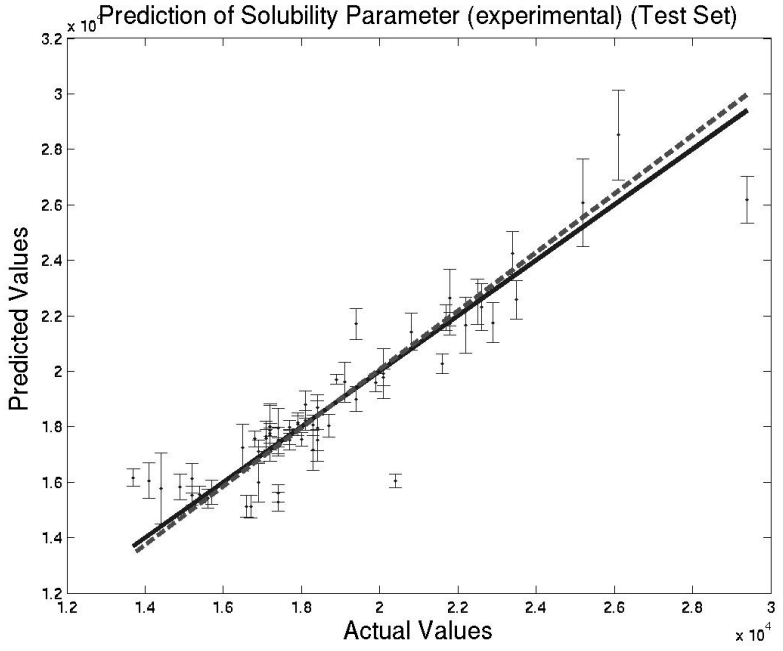


Figure 4.9: Predicted vs. actual Hildebrands solubility parameter values derived from a 9 Fisher component linear model. The solid line represents a best fit for the data; the dashed line is y=x. Error bars are 2σ.

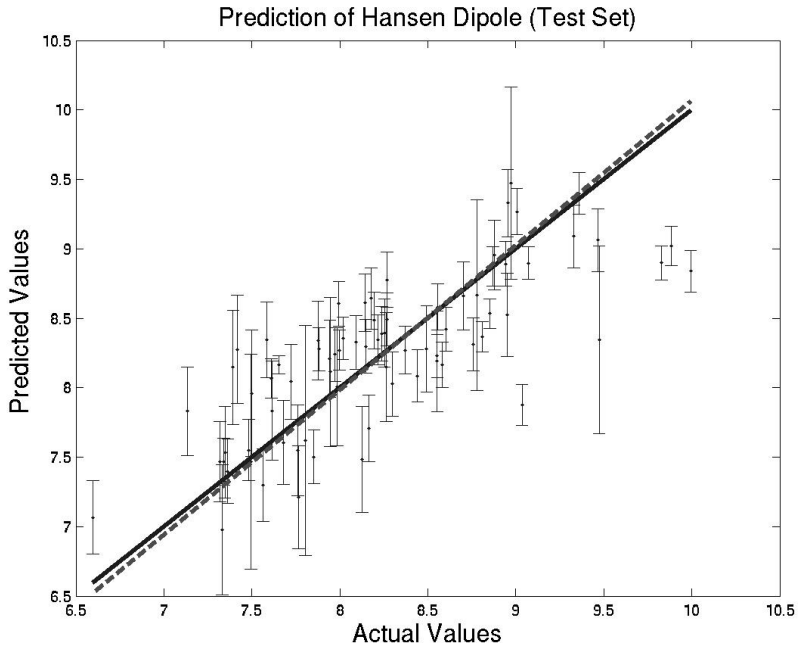


Figure 4.10: Predicted vs. actual Hansen Dipole Parameter values derived from a 11 Fisher component linear model. The solid line represents a best fit for the data; the dashed line is y=x. Error bars are 2σ.

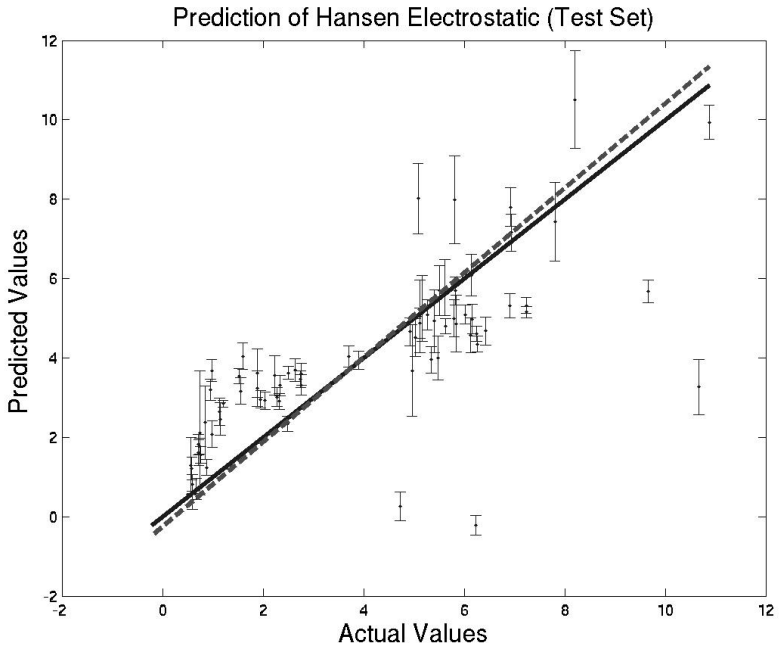


Figure 4.11: Predicted vs. actual Hansen dipole parameter values derived from a 5 Fisher component linear model. The solid line represents a best fit for the data; the dashed line is y=x. Error bars are 2σ.

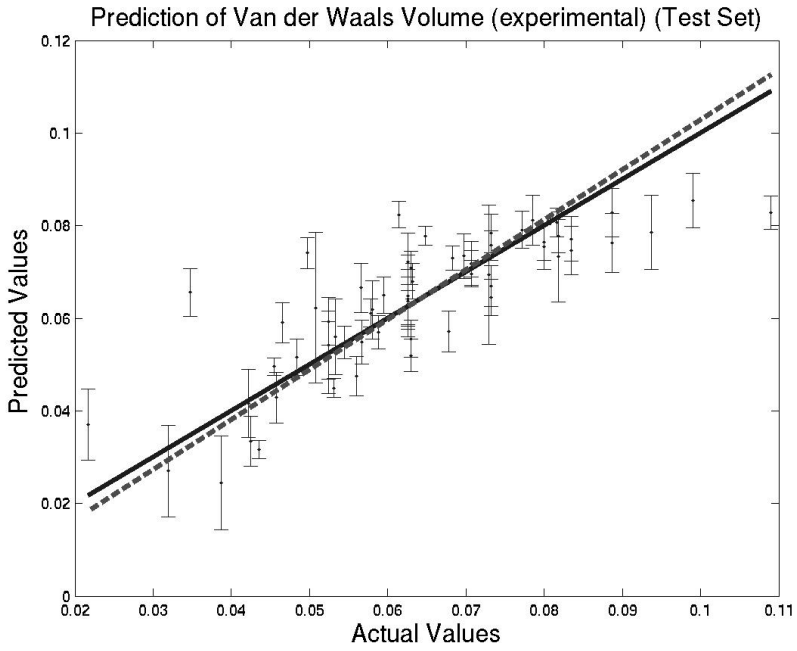


Figure 4.12: Predicted vs. actual van der Waals volume values derived from a 11 Fisher component linear model. The solid line represents a best fit for the data; the dashed line is y=x. Error bars are 2σ.

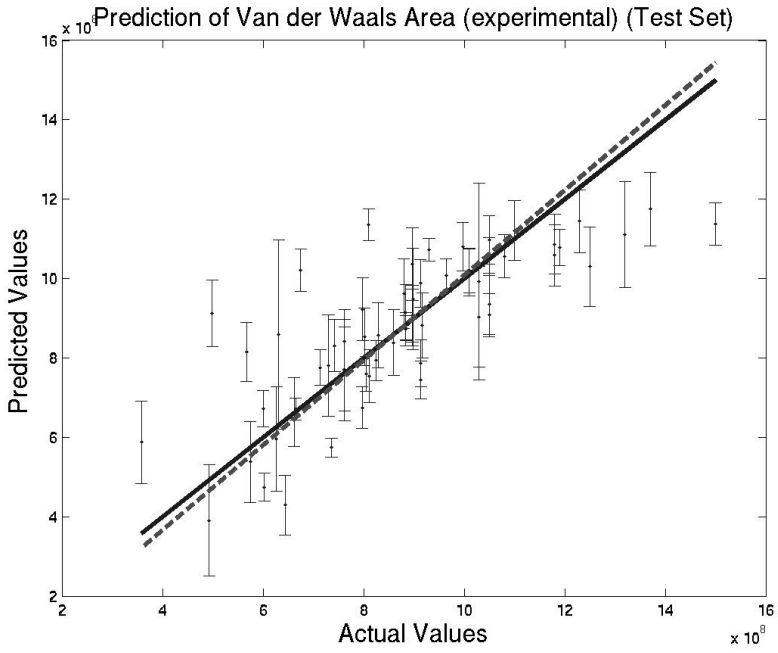


Figure 4.13: Predicted vs. actual van der Waals area values derived from a 10 Fisher component linear model. The solid line represents a best fit for the data; the dashed line is y=x. Error bars are 2σ.

successful because there were, in general, few differences between the analytes in the set except the variable of interest. For instance, within a homologous series of alcohols, only molecular weight and branching of the chain separated the 12 unsaturated members of the set, with three also possessing a unit of unsaturation. Monofunctional halides were successfully separated from bi- or tri-functional halides, benzenoid aromatics were distinct from pyridinoids, and saturated hydrocarbons were not readily confused with unsaturated hydrocarbons. This is largely due to the fact that, of the many factors to which a semi-selective detector is responsive, frequently only one varies significantly within a particular chemical identification task, so the chemical subclasses (such as saturated hydrocarbons) are largely homogeneous relative to the differences between the subclasses.

However, successfully classifying a set of analytes that differ not only in analyte class but also with regard to many other parameters was found to be more difficult. Among the non-halide classes, the bulk of the mistakes in *k*-NN analysis was contained in five analytes, such as methyl acetate or 2,5-dimethyl 2,4-hexadiene, that differed significantly from their base analyte classes. Because classification proved successful for a high proportion of analyte exposures, unique class characteristics such as aromaticity, hydrogen bonding, and others found in this analyte set must largely overwhelm considerations such as saturation, branching, and volume differences, which vary considerably within classes without leading to significant rates of misclassification.

The comparative difficulty in correctly identifying analyte class members that differ from their respective classes is not surprising. However, the result does not imply that recognizing analytes within diverse classes is impossible, as found from the high rate of correct classification among the pyridinoid aromatics, which differ greatly from the benzenoid. Rather, a sufficiently robust analyte basis set must be established to ensure that a sufficiently similar neighbor exists in the analyte database. Expecting correct classification of pyridine as aromatic, for example, would be unreasonable if lutidine or picoline were not present in the set. However, with a well-developed analyte basis set, a great deal of within-class diversity can be tolerated before poor classification performance is obtained.

Though it is beneficial that the detector set employed in this analysis proved more sensitive to solvent class characteristics than physical characteristics, it is not clear that this trend can be generalized to all detector sets. It is certainly conceivable that the overwhelming variable could have been molecular volume. Had this been the case, efficient clustering along class boundaries would likely have not been possible. An array of detectors comprised of a homologous series of polymers that differed only in chain length, such as poly(1-alkenes), might prove more sensitive to molecular volume than to chemical characteristics. Because the detector array in this study was designed to ensure that a variety of polymers with diverse characteristics was represented, it should not be surprising, then, that the detector array used in this study is most sensitive to chemical class.

The polymer/carbon-black composite vapor detectors used in this study have previously been shown to produce a response that is linear with analyte concentration. Hence, the analysis performed herein is essentially independent of the concentration of analyte.[6] Development of a robust model for non-linearly responding detectors is expected to be less straightforward, and likely requires collection of larger datasets along with modeling the detector response as a function of analyte concentration.[29] Such complications are minimized through the use of polymer/carbon-black composite detectors, which facilitates a straightforward development of classification models for various chemical and physical properties of the analytes of interest.

The effectiveness with which each physical property is predicted is related to how strongly it relates to analyte sorption. Hildebrand's solubility parameter is strongly linked to analyte sorption, as is K_{ow}, and both were well predicted from sensor responses. It is interesting to note that neither the Hansen electrostatic or dipole terms are as easily predicted from sensor responses as is solubility parameter; this suggests that the analytes and sensing polymers we used in this study were strongly dependent upon both electrostatic and dipole interactions (as well as hydrogen bonding for the alcohols). Other properties, such as ϵ and van der Waals area or volume, impact analyte/polymer sorption, although not as strongly. As such, these properties are predicted fairly well, though not as well as K_{ow} or solubility parameter, by the manner in which the analytes in question interact with the sensors used in this study.

4.6 Conclusions

In addition to identifying with a high degree of confidence analytes to which it has been previously exposed, an array of semi-selective polymer/carbon-black composite detectors is also capable of qualitatively describing analytes to which it has never been previously exposed. *k*-NN and principal components analysis indicate that, with the exception of the halide analyte class, the detector array used in this work is capable of assigning with a high degree of confidence class descriptors to analytes. In addition to the overall analyte class, a variety of chemical sub-class information can also be extracted from the raw data, including hydrocarbon saturation, mono- vs. poly-functionality among halides, and the nature of aromatic rings (benzenoid vs. pyridinoid). Additionally, in certain cases information regarding the size and dipole moment of molecules can be determined. Ultimately, it is possible that in many cases sufficient knowledge of an unknown analytes characteristics can be established through the extraction of chemical and physical parameters to allow tentative identification of analytes that have not been previously encountered by the sensor array.

Bibliography

[1] Gardner, J.; Pike, A.; Derooij, N.; Koudelkahep, M.; Clerc, P.; Hierlemann, A.; Gopel, W. *Sensors and Actuators B - Chemical* **1995,** *26,* 135–139.

[2] Bartlett, P.; Archer, P.; Ling-Chung, S. *Sensors and Actuators* **1989,** *19,* 125–140.

[3] Shurmer, H.; Gardner, J.; P., C. *Sensors and Actuators B - Chemical* **1990,** *1,* 256–260.

[4] Freund, M.; Lewis, N. *Proceedings of the National Academy of Sciences of the United States of America* **1995,** *92,* 2652–2656.

[5] Lonergan, M.; Severin, E.; Doleman, B.; Beaber, S.; Grubb, R.; Lewis, N. *Chemistry of Materials* **1996,** *8,* 2298–2312.

[6] Severin, E.; Doleman, B.; Lewis, N. *Analytical Chemistry* **2000,** *72,* 658–668.

[7] Hodgins, D. *Sensors and Actuators B - Chemical* **1995,** *27,* 255–258.

[8] Dickinson, T.; White, J.; Kauer, J.; Walt, D. *Nature* **1996,** *382,* 697–700.

[9] White, J.; Kauer, J.; Dickinson, T.; Walt, D. *Analytical Chemistry* **1996,** *68,* 2191–2202.

[10] Dickinson, T.; Michael, K.; Kauer, J.; Walt, D. *Analytical Chemistry* **1999,** *71,* 2192–2198.

[11] Rakow, N.; Suslick, K. *Nature* **2000,** *406,* 710–713.

[12] Ronot, C.; Archenault, M.; Gagnaire, H.; Goure, J.; Jaffrezicrenault, N.; Pichery, T. *Sensors and Actuators B - Chemical* **1993,** *11,* 375–381.

[13] Ballantine, D.; Rose, S.; Grate, J.; Wohltjen, H. *Analytical Chemistry* **1986,** *58,* 3058–3066.

[14] Rose-Pehrsson, S.; Grate, J.; Ballantine, D.; Jurs, P. *Analytical Chemistry* **1988,** *60,* 2801–2811.

[15] Patrash, S.; Zellers, E. *Analytical Chemistry* **1993,** *65,* 2055–2066.

[16] Lang, H.; Baller, M.; Berger, R.; Gerber, C.; Gimzewski, J.; Battiston, F.; Fornaro, P.; Ramseyer, J.; Meyer, E.; Guntherodt, H. *Analytica Chimica Acta* **1999,** *393,* 59–65.

[17] Duda, R.; Hart, P. *Pattern Classification and Scene Analysis;* John Wiley and Sons: New York, 1984.

[18] Jurs, P.; Bakken, G.; McClelland, H. *Chemical Reviews* **2000,** *100,* 2649–2678.

[19] Park, J.; Groves, W.; Zellers, E. *Analytical Chemistry* **1999,** *71,* 3877–3886.

[20] Patel, S.; Jenkins, M.; Hughes, R.; Yelton, W.; Ricco, A. *Analytical Chemistry* **2000,** *72,* 1532–1542.

[21] Doleman, B.; Lonergan, M.; Severin, E. *Analytical Chemistry* **1998,** *70,* 4177-4190.

[22] Burl, M.; Sisk, B.; Vaid, T.; Lewis, N. *Sensors and Actuators B - Chemical* **2002,** *87,* 130–149.

[23] Matzger, A.; Lawrence, C.; Grubbs, R.; Lewis, N. *Journal of Combinatorial Chemistry* **2000,** *2,* 301–304.

[24] Vaid, T.; Burl, M.; Lewis, N. *Analytical Chemistry* **2001,** *73,* 321–331.

[25] Vaid, T.; Lewis, N. *Bioorganic and Medicinal Chemistry* **2000,** *8,* 795–805.

[26] Hildebrand, J. *The Solubility of Nonelectrolytes;* Dover Publications: New York, 1964.

[27] Hansen, C. *Hansen Solubility Parameters: A User's Handbook;* CRC Publications: Boca Raton, Florida, 1999.

[28] *DIPPR Chemical Database;* Brigham Young University DIPPR Thermophysical Properties Laboratory: `http://dippr.byu.edu/public/chemsearch.asp`, 2004.

[29] Osbourn, G.; Bartholomew, J.; Ricco, A.; Frye, G. *Accounts of Chemical Research* **1998,** *31,* 297–305.

Chapter 5

Comparison of Fisher's Linear Discriminant and Multi-Layer Perceptron Networks for Classification of Analytes Exposed to a Chemical Vapor Detector Array

5.1 Abstract

Two different classification methods, Fisher's linear discriminant (FLD) and a multilayer perceptron neural network (MLP), were directly compared with respect to their abilities to differentiate response patterns arising from arrays of chemical vapor detectors. The algorithms were compared in five different types of tasks that had been selected because they produced classification problems of varying character and difficulty. In one task, an array of 20 compositionally distinct polymer/carbon-black composite vapor detectors was exposed to P/P°=0.0075 1-propanol and P/P°=0.0083 2-propanol, where P and P° are the vapor pressure and standard vapor pressure, respectively, of a given analyte. The second task consisted of classification of a mixture of P/P°=0.011 1-propanol and P/P°=0.0090 2-propanol vs a mixture of P/P°=0.00901-propanol and P/P°=0.011 2-propanol. A third task consisted of multiple concentrations of three hydrocarbons, and a fourth task involved clustering two hydrocarbons in the presence of a variable background composition. An additional dataset was generated by exposing an array of five thin-film metal-oxide sensors to the headspace of seven different coffee blends. In each case, the MLP and FLD techniques were compared using the 5-sensor subset of the 20 available sensors that proved optimal for that dataset. The FLD and MLP algorithms yielded comparable performance on straightforward classification tasks, whereas the MLP technique yielded better performance on tasks that involved non-linear classification boundaries. In addition, for the four datasets produced by the polymer/carbon-black

composite detector array, the performance of each possible 5-sensor subset was evaluated using both signal processing approaches. The performance of the best 5-sensor subset selected with MLP was found to be slightly better than the performance of the FLD-selected subsets, and the performance of the median 5-sensor subset using MLP was nearer to that of the optimal subset than the median sensor array selected by FLD. In one case, the optimal test set performance distribution was found to be significantly better with MLP than with FLD: MLP had a clear advantage (86% vs. 57% correct classification rate) when applied to the "coffees" dataset, and this trend is likely applicable to other multi-cluster classification tasks that consisted of non-Gaussian shaped data in lower-dimensional spaces.

5.2 Introduction

Arrays of broadly responsive chemical sensors have attracted significant attention due to their inherent portability, ability to accommodate a diversity of detector materials, and wide-ranging possible application areas. Of course, the utility of such array-based methods is intimately linked to the efficacy of the signal processing algorithm that is used for the identification, classification, and quantification of analytes. In this work, we have compared the performance of two different methods of feature selection, Fisher's linear discriminant (FLD) and a multilayer perceptron neural network (MLP), on a variety of vapor detection tasks.

A great deal of prior work has been devoted to evaluating many different methods for analysis of data[1–4] produced by arrays of broadly responsive vapor detectors based on a variety of different architectures.[5–9] Vaid *et al.* used polymer/carbon-black composite detectors to compare a variety of lower-order methods, including FLD and quadratic discriminant analysis, as well as k-nearest-neighbor analysis, and found that FLD performed best with the data included in that study.[10] In that study, the sensor responses to a given analyte were largely Gaussian in character, with the variability dominated by nonsystematic fluctuations in the vapor generation system. Longer-term signal classification performance has also been addressed for such systems, and such studies have revealed that only small, nonsystematic, changes in response occurred over time periods as long as two months.[11] When the responses of different analytes are sufficiently distinct, and when the clusters of response data are easily represented by a parametric form (i.e., responses that are linear with concentration and Gaussian for each sensor), the choice of algorithm used is hardly relevant, because even a simple Euclidean distance metric should be sufficient for perfect classification. In fact, for binary analyte separations of Gaussian data, a FLD model can be proven to coincide with that of the optimal Bayes classification model.[12] This is typically the case for the arrays of polymer/carbon-black composites vapor detectors, which have been previously shown to elicit highly linear responses with respect to analyte concentration.[13] For example, even for tasks such

as separation of n-hexane from n-heptane (each at P/P=0.01) FLD provided correct classification >95% of the time over extended periods of use. Thus, more challenging analyte classification tasks are required to determine the relative quality of different data analysis algorithms on such vapor detector arrays.

For such challenging tasks, as well as for sensors that inherently do not produce a Gaussian distribution of response and/or do not produce signals that are linear with increasing analyte concentration, it is of interest to compare the performance of more flexible, non-linear algorithms, such as artificial neural networks (ANNs), to that of linear algorithms such as FLD. In this work, two sets of detector arrays, one based on polymer/carbon-black composites, and one based on metal oxide sensors, have been used in a variety of different tasks. For the polymer/carbon-black composite sensors, standard chemical solvents were used as test analytes. As expected, these analytes produced highly linear and Gaussian responses on the chemiresistor detector array. However, the classification tasks were very difficult and involved very chemically similar analytes at low concentrations, very similar mixtures of analytes, and sets of similar analytes in the presence of various chemical interferents. On the array of metal oxide vapor detectors, a series of coffees was studied, and these complex analyte mixtures on non-linear sensors produced nonlinear and non-Gaussian responses that also presented difficult classification issues for linear discriminant algorithms.

The effect of using sub-optimal feature (sensor) selection has also been evaluated with respect to the performance of linear vs. highly nonlinear algorithms. The simulation of sub-optimal feature selection is important because any sensor array of any size is a subset of a potentially larger sensor array, and is almost certainly not the most capable sensor set that could in principle be constructed for any specific analyte recognition task. Thus, evaluating the performance of different algorithms on nonoptimal sensor arrays is essential to ensuring the viability of these algorithms under real-world conditions. Additionally, smaller sensor subsets (as small as four or five sensors of some larger array) are frequently used to reduce power consumption, increase sampling frequency, minimize device size, or reduce computational overhead.[14,15] As such, it is useful to compare FLD and MLP with respect to reducing any performance penalties that may arise when fewer sensors are used in data collection and/or analysis.

Table 5.1: Detectors used in this study.

1	poly(caprolactone)
2	poly (ethylene-co-vinyl acetate)
3	poly(ethylene oxide)
4	poly(ethylene glycol)
5	poly(styrene-co-butadiene)
6	poly(methyl octadecylsiloxane)
7	poly(vinyl stearate)
8	ethyl cellulose
9	poly(styrene-co-maleic anhydride)
10	poly(methyl vinyl ether-co-maleic anhydride)
11	poly(4-vinyl phenol)
12	poly(vinyl acetate)
13	cellulose acetate
14	polycarbonate
15	polystyrene
16	polysulfone
17	poly(methyl methacrylate)
18	poly(vinyl butyral)
19	hydroxypropyl cellulose
20	poly(styrene-isoprene)

5.3 Experimental

5.3.1 Sensor arrays

5.3.1.1 Polymer/Carbon-black Composite Detectors

Twenty compositionally distinct polymer/carbon-black composite chemically sensitive resistors (Table 5.1) provided one test array. The substrate for each sensor consisted of a glass slide, onto which two Au leads had been evaporated. Detector films were cast from mixtures of 40% polymer, 40% di(ethylene glycol) dibenzoate, and 20% by weight of carbon-black (Black Pearls 2000, Cabot Inc), as described previously.[11,16] Sensors were fabricated by using an airbrush to spray coat the polymer/carbon-black mixture onto the substrate. The array of sensors was housed in a stainless steel assembly that was connected by teflon tubing to a computer-controlled, calibrated vapor generation and delivery system.[13] When exposed to an analyte vapor, each sensor exhibited an increase in resistance that was transduced as a single descriptor, $\Delta R_{eq}/R_b$, where ΔR_{eq} is the equilibrium differential resistance of the sensor after exposure to vapor and R_b is the resistance of the sensor when exposed to clean laboratory air.

5.3.1.2 Metal Oxide Detectors

The Pico-2 EN, developed at the Sensor Lab of the University of Brescia, made use of five thin film semiconducting gas sensors. The sensors were either classical catalyzed SnO_2-based sensors

or Ti-Fe sensors. All of these sensors were formed by sputtering using the Rheotaxial Growth and Thermal Oxidation (RGTO) technique.[17] The surface of the film after the thermal oxidation step of the RGTO technique consisted of porous, nano-sized agglomerates that are well-suited for gas absorption. A thin layer of noble metal was deposited as a catalyst on the three SnO_2 sensors. To improve sensitivity and selectivity, one sensor was coated with Au, one with Pd and one with Pt. Prior to data analysis, responses from these sensors were transduced to $\Delta R_{max}/R_b$, where ΔR_{max} is the maximum resistance change upon exposure to vapor and R_b is the baseline resistance under laboratory air.

5.3.2 Datasets

5.3.2.1 Polymer/Carbon-black Composite Detectors

Four datasets (Table 5.2) were collected on polymer/carbon-black composite detector arrays. Dataset 1 ("low concentration"), consisted of 200 exposures in random order to either 1-propanol at P/P° = 0.0075 or 2-propanol at P/P° = 0.0083. Slightly different activities of the two analytes were used because these values produced approximately equal amplitude responses on the polymer/carbon-black composite sensor array. Dataset 2 ("close mixtures") consisted of 200 exposures to a gas mixture of 1-propanol at P/P° = 0.0090 and 2-propanol at P/P° = 0.011 and of 200 exposures to a second gas mixture that consisted of 1 propanol at P/P° = 0.011 and 2-propanol at P/P°= 0.0090. Dataset 3 ("multiple concentrations") consisted of exposures to either n-hexane, ethyl acetate, n-heptane, or n-octane, each at P/P° = 0.010, 0.020, 0.040, and 0.070. Each analyte and concentration combination was exposed 100 times to the detector array, with exposures occurring in random order during the data collection run. Dataset 4 ("interferents") consisted of 200 exposures to each of n-heptane at P/P° = 0.022 or n-octane at P/P° = 0.022, with all exposures performed in the presence of a series of interferents that were introduced into the background carrier gas. The interferents were: 1) nothing; 2) ethanol at P/P° = 0.0060; 3) ethanol at P/P° = 0.010; 4) ethanol at P/P° = 0.016; 5) tetrahydrofuran (THF) at P/P° = 0.0060; 6) THF at P/P° = 0.010; 7) THF at P/P° = 0.016; or 8) a mixture of THF and ethanol at a combined partial pressure of P/P° = 0.016. For interferent 8, the analyte bubblers that contained THF and ethanol were driven by a single mass flow controller, so the exact fraction of the two interferents is unknown. Each of the 16 analyte/interferent combinations was exposed in random order 25 times to the detector array.

5.3.2.2 Metal Oxide Detectors

For this classification task, seven distinct blends of Italian coffees were used as analytes (Table 5.2). To generate analyte vapors, gas was sampled via an autosampler (static headspace extraction) that contained 2g of ground coffee. 36 measurements each were collected for seven coffee blends.[18]

Table 5.2: Test-set classification rate for the classification tasks studied.

Dataset #, Name	Classification Problem	MLP Performance	FLD Performance
1,"Low concentration"	Binary, no normalization	94%	94%
1,"Low concentration"	Binary, normalization	68%	65%
2,"Low concentration"	Binary, no normalization	84%	81%
2,"Low concentration"	Binary, normalization	80%	79%
3,"Low concentration"	Binary, n-hexane vs. n-heptane	100%	97%
4,"Low concentration"	Binary	97%	95%
4,"Low concentration"	Binary, all sensors, n-heptane vs. n-octane	99%	98%
4,"Low concentration"	4 classes, determination of ethanol interferent concentration	100%	80%
4,"Low concentration"	4 classes, all sensors, determination of ethanol interferent concentration	100%	83%
5,"Low concentration"	7 classes	86%	57%
5,"Low concentration"	Binary, coffee 1 vs. coffee 5	100%	89%

All classification performances derived from the best 5-sensor subset unless otherwise noted.

5.3.3 Algorithms and Feature Selection

5.3.3.1 Linear Methods

Data were first visualized using Principal Components Analysis (PCA).[12] FLD was then used to determine the optimal linear separation between any set of analyte data clusters. In all cases, a train/test scheme was employed, in which the optimal FLD transformation was determined using the training data, and this transformation was then applied to the test data.

5.3.3.2 Multi-layer Perceptron Networks

The raw data were preprocessed using principal components analysis (PCA) yielding projected data that were then used as inputs to the MLP system. Since only five sensors were employed, all five PCA projections were used without further extraction. As a result, the role of the PCA transformation for this data set was to de-correlate the inputs, as the dimensionality of the data was unchanged. The train-test division was the same as that used for the Fisher discriminant. To prevent overfitting,[19] both early stopping (ES)[20] and weight decay regularization.[21] were initially employed. The two methods gave similar results, so the faster ES method was used. The use of a complexity control method dispensed with the need to select the optimal number of hidden units for the MLP network. The only requirement for this method was to use a sufficiently large network to avoid underfitting.[22] Trials were performed with 5, 7, and 9 hidden units, and 7 units were found to give sufficient flexibility and minimal computational time. The computational time remained an important issue, as performing MLP analysis for each 5 sensor subset of 15 total

sensors required 12 hr, whereas FLD required only 20 min to perform the same task. The MLP network was trained using the backpropagation method, with the error function minimized using the Levenberg-Marquardt algorithm.[19,23] Ten network initializations were usually performed for scanning the error space, and the net that produced the best result on the test set was used.

5.4 Results

5.4.1 Classification of Low Concentration Exposures of Chemically Similar Analytes

Dataset 1 was analyzed for both normalized and raw response data. The normalization process was performed using eq 5.1:

$$S'_{ij} = \frac{S_{ij}}{\sum_{i=1}^{n} S_{ij}} \tag{5.1}$$

where S_{ij} refers to the $\Delta R_{eq}/R_b$ response of the jth detector (out of n total detectors) to the ith analyte exposure, and S'_{ij} represents the sum-normalized analog of S_{ij}. If the sensor response is linearly related to the concentration of analyte, as is the case for the polymer/carbon-black composite sensors, this normalization procedure scales the sensor responses to be independent of analyte concentration. Discrimination between different analytes can then be based only on the direction of the feature vector and not on its amplitude.

The PCA plot for Dataset 1 (Figure 5.1) yielded two fairly Gaussian-shaped, partially superimposed data clusters. Classification results on these data were similar for FLD and MLP, with each algorithm yielding 94% correct classification for unnormalized data (Table 5.2). Such behavior is expected because for a pair of perfectly Gaussian distributed classes the optimal Bayes classifier is obtained with FLD.[12] Both algorithms yielded much lower classification performance on normalized data (Table 5.2), indicating that slight differences in the concentrations delivered for two analytes played a role in the enhanced classification performance of the unnormalized data set.

5.4.2 Classification of Highly Similar Binary Analyte Mixtures

The PCA plot for Dataset 2 ("close mixtures") (Figure 5.2) clearly showed that these data were not Gaussian-distributed. In spite of this, for unnormalized data, both FLD and MLP yielded relatively high classification rates using the best 5-sensor subset (Table 5.2). The FLD and MLP methods yielded nearly identical performance for this task as well. The plane generated by the first two principal components (PC 1 and PC 2), which is derived from some contribution of all sensors used, evidently did not contain a significant contribution from the dimension on which the FLD

Figure 5.1: Principal components analysis plot of Dataset 1 ("low concentrations"), comprised of (◯) $P/P°$=0.0075 1-propanol and (×) $P/P°$ =0.0083 2-propanol.

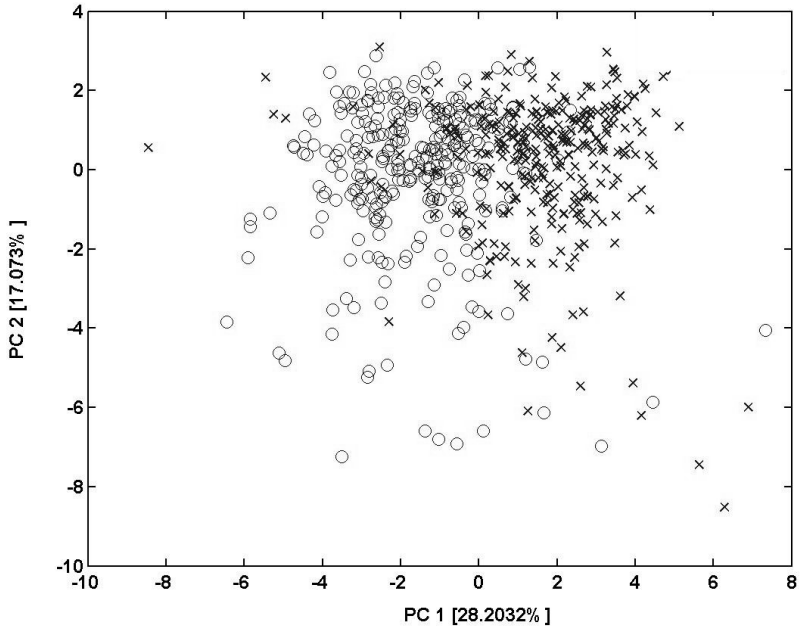


Figure 5.2: Principal components analysis plot of Dataset 2 ("close mixtures"), comprised of a mixture of $P/P°$=0.011 1-propanol and $P/P°$=0.0090 2-propanol (◯) and a mixture of $P/P°$=0.0090 1-propanol and $P/P°$=0.011 2-propanol (×). The first two principal components shown contained 66% of the total sample variance.

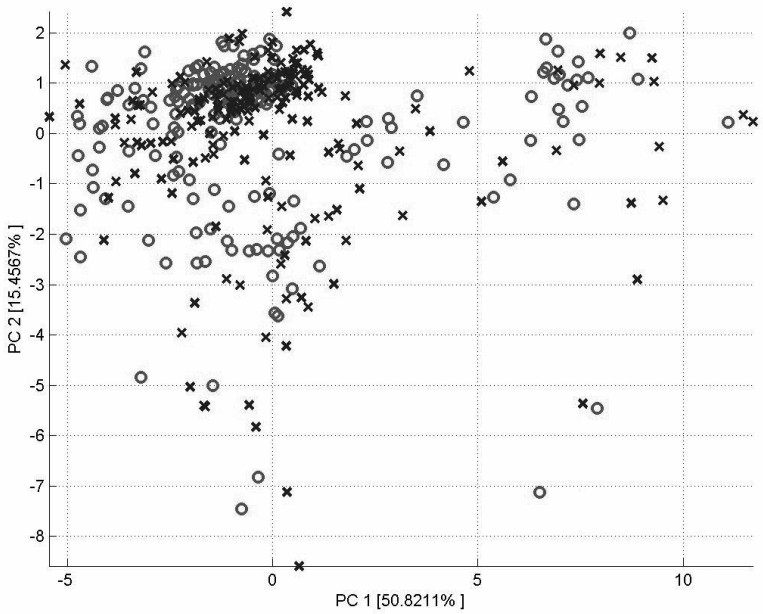
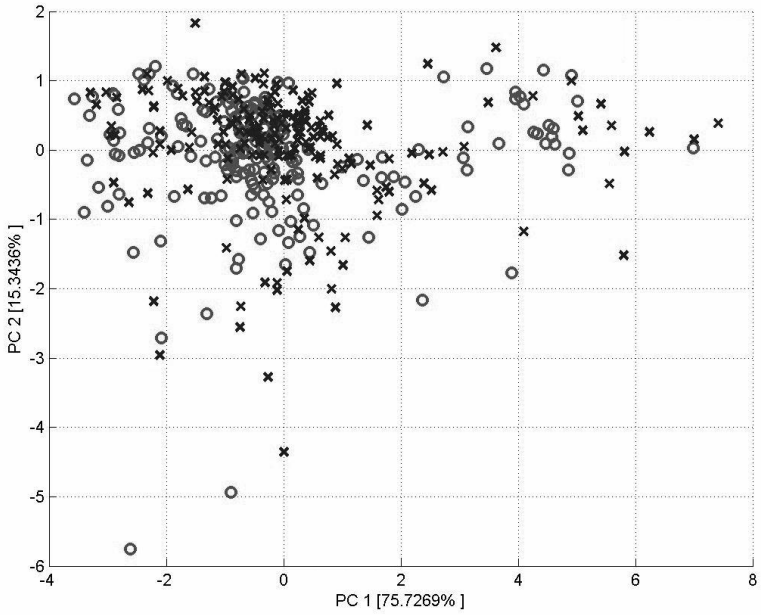


Figure 5.3: Principal components analysis plot of the two most significant principal components (PCs 1 and 2) of Dataset 2("close mixtures"), comprised of a mixture of $P/P°$=0.011 1-propanol and $P/P°$=0.0090 2-propanol (◯) and a mixture of $P/P°$=0.0090 1-propanol and $P/P°$=0.011 2-propanol (×). Only data from the best 5 sensors (6, 8, 12, 18, and 19 from Table 5.1) were used to generate this plot.



algorithm projected the data in order to produce its 81% classification rate. Had it done so, the data of Figure 5.2 would have shown significant clustering for each analyte. In an attempt to find a PC plane that did contain such a dimension, the search for the discrimination boundary was restricted to the feature space spanned by the top-scoring subset of 5 detectors, as determined by FLD (sensors 6, 8, 12, 18, and 19 from Table 5.1). Figure 5.3 shows the PCA plane generated by the first two principal components derived from the array of 5 optimal detectors. This plot exhibited no better clustering than that in Figure 5.2. A different distribution, and a much better separation, was only found when the fifth principal component was included in the PCA plot (Figure 5.4). Note that principal component 5 (out of 5) contains a very small fraction of the total variance (0.14%) in the dataset, and therefore the sensors must be very stable if this separation is to be preserved.

5.4.3 Classification of Unnormalized Analyte Exposures Independent of Concentration

A PCA plot of Dataset 3 ("multiple concentrations") indicated that ethyl acetate produced very distinct responses from those produced by the three hydrocarbons n-hexane, n-heptane, and n-octane. Figure 5.5 shows a PCA plot of Dataset 3, in which the ethyl acetate data have been omitted for clar-

Figure 5.4: Principal components analysis plot of the two least significant principal components (PCs 4 and 5) of Dataset 2 ("close mixtures"), comprised of a mixture of 0.011 $P/P°$ 1-propanol and $P/P°$=0.0090 2-propanol (◯) and a mixture of $P/P°$=0.0090 1-propanol and $P/P°$=0.011 2-propanol (×). Only data from the best 5 sensors (6, 8, 12, 18, and 19 from Table 5.1) were used to generate this plot.

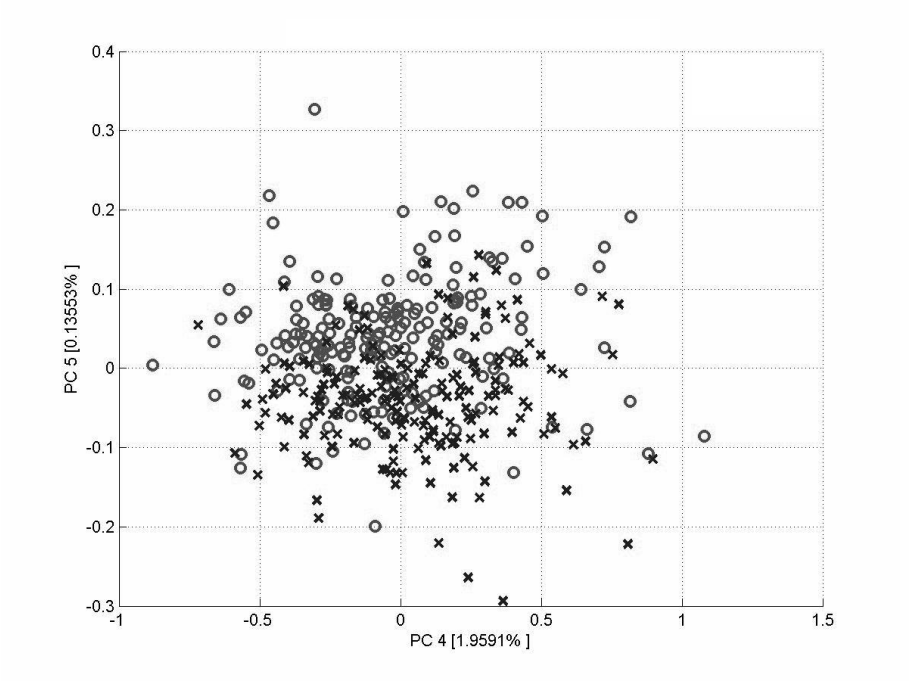
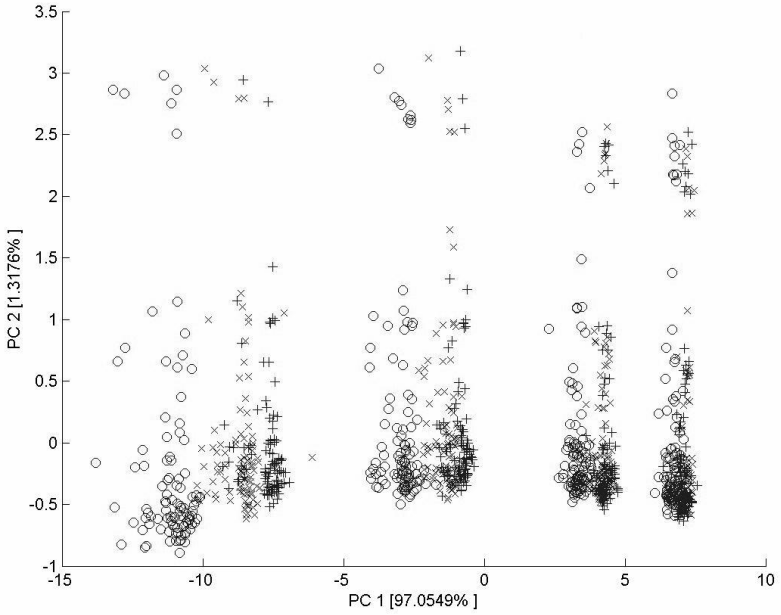


Figure 5.5: Principal components analysis plot for Dataset 3 ("multiple concentrations"), comprised of 400 exposures to n-hexane (◯), n-heptane (×), and n-octane (+). Each analyte was presented to the detector array at four different concentrations. Concentrations in the plot increased from right to left, with a high degree of clustering within the individual analytes. Data were not normalized prior to PCA analysis.

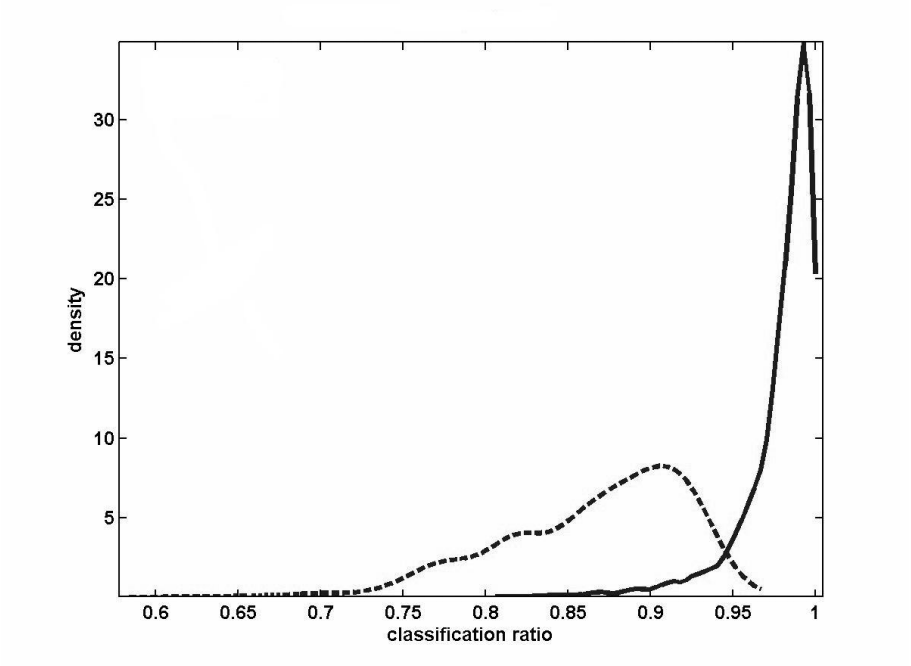


ity. The clusters were separated primarily based on the differences in concentration of the different analytes, and not on differences between the identities of the analytes at that test concentration. At very high concentrations, the various hydrocarbon analytes were somewhat distinguishable from each other in the space defined by the first two principal components. Despite the obvious similarities between the response patterns for n-hexane and n-heptane at each concentration, MLP showed 100% correct classification of analyte for normalized data collected over a variety of analyte concentrations, and the linear FLD classifier yielded a 97% correct classification rate. This strongly implies that whatever concentration differences existed were extremely linear, and therefore could be robustly removed by normalization. These results also suggest that the MLP method was able to effectively define its own non-parametric normalization scheme, and was as effective toward that end as FLD with standard normalization.

5.4.4 Classification

The MLP and FLD methods exhibited similar classification performance when optimal detector array subsets were used for the data input. However, for suboptimal sensor subsets, MLP yielded a high rate (> 90%) for most of the 5-sensor subsets, whereas the classification performance using FLD

Figure 5.6: Distribution of the Fisher Linear Discriminant (dashed line) and Multi-Layer Perceptron (solid line) classification rates over all 5-sensors subsets for Dataset 3 ("multiple concentrations").



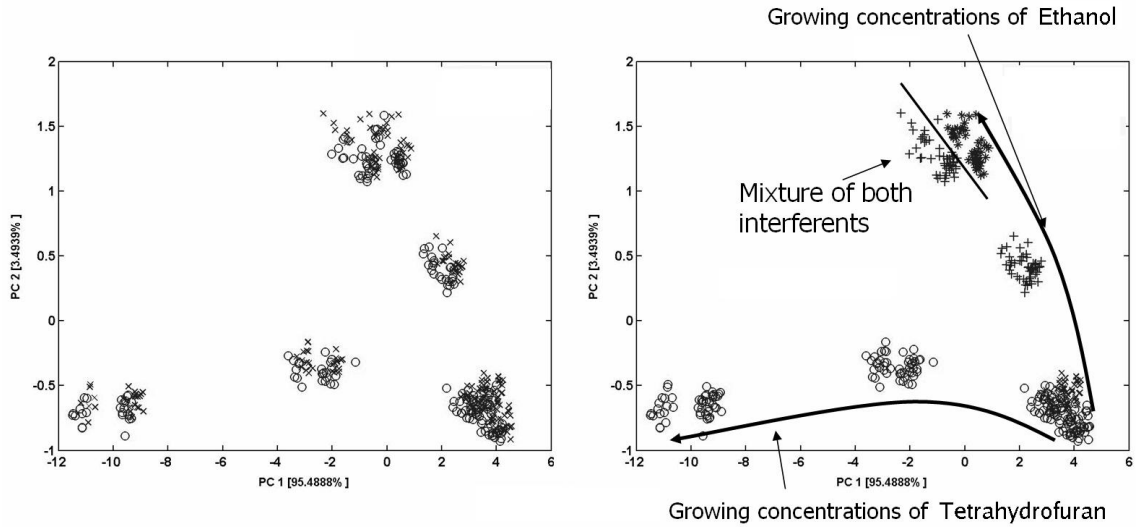
degraded much more quickly when non-optimal sensor subsets were used (Figure 5.6). Apparently, only some subsets are sufficiently insensitive to the concentration differences between analytes that they span a linearly separable space, whereas the bulk of the sensor sets possess some significant sensitivity to these nonlinearities.

5.4.5 Classification of Similar Analytes in the Presence of Variable Analyte Backgrounds

Dataset 4 ("interferents") was collected with the intention of making the discrimination task very difficult and additionally to attempt to simulate some of the conditions that might exist in poorly controlled field use of such sensor arrays. In this data run, two base analytes (n-heptane and n-octane) were used in fixed concentrations and various concentrations of two background interferents (ethanol and tetrahydrofuran) were added to the base analyte stream. The resulting sensor array data were analyzed in the context of two separate tasks: the first task was to identify the base analyte (n-heptane or n-octane) without regard to the concentration of the interferent; the second task was to identify the concentration level of the ethanol "interferent," without regard to the identity of the base analyte or to the concentration of the tetrahydrofuran interferent.

Figure 5.7a presents a PCA plot for the first task, discrimination between n-heptane and n-octane in the presence of various interferents, while Figure 5.2b presents a PCA plot for the second task,

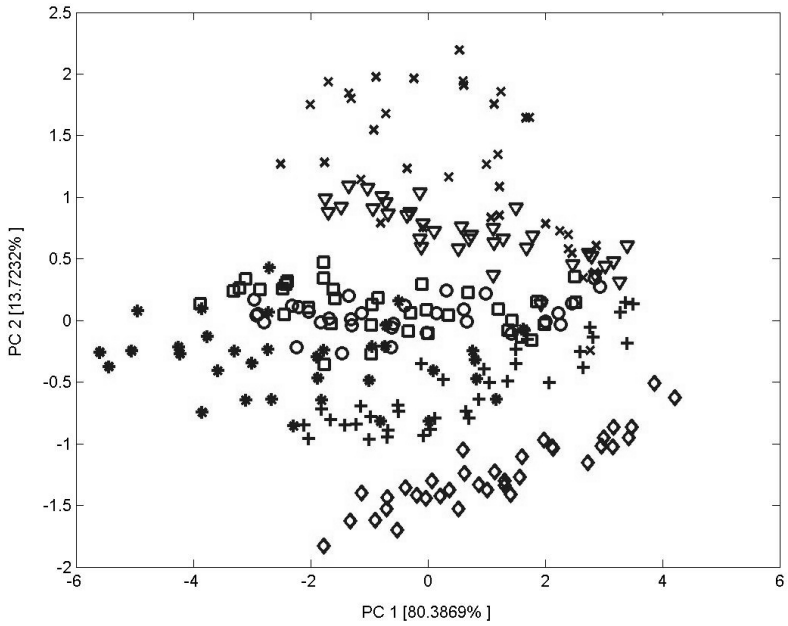
Figure 5.7: Principal components analysis of Dataset 4 ("interferents"), which consisted of n-heptane or n-octane in the presence of variable levels of ethanol and/or tetrahydrofuran. Plots a) and b) display the same PCA projection with two different labelings of the data, corresponding to the two classification tasks that were performed with the data. Plot a) shows discrimination between two analytes, n-heptane (◯) and n-octane (×) irrespective of interferents. Plot b) shows the concentration level of the ethanol interferent (◯,×, +, and ∗, indicating lowest to highest ethanol concentration, respectively), without regard to the base analyte identity or presence of the tetrahydrofuran interferent.



distinguishing among the four different ethanol interferent concentrations without regard to the THF interferent concentration or the identity of the base analyte. The PCA plots (PCs 1 and 2) indicate that the variance in the data was produced almost entirely by the different concentration levels of the interferents. These figures indicate that the class distribution of each analyte was multimodal, where each mode corresponded to a different concentration level of an interferent. For the first "interferents" task, excellent classification performance was observed for both MLP and FLD classifiers (97% vs. 95% classification rate, respectively). The excellent performance of the FLD in this task clearly is related to the linearity of response of the detectors as a function of analyte concentration, as observed for the tasks of Datasets 1 and 3. When the classification performance of a collection of nonoptimal detector subsets was considered, the distributions had a similar shape but the MLP distribution was shifted towards somewhat higher performance values, with the means over all subsets being 82% and 75% for the MLP and FLD classifiers, respectively.

The second "interferents" task was effectively a multi-class problem. For the best 5-sensor subsets, MLP determined the ethanol interferent concentration with perfect accuracy (473 5-sensor subsets of the sensors in Table 5.1 yielded perfect classification), but the FLD algorithm produced no higher than 80% classification when using a 5-sensor subset. The FLD algorithm produced an 83% classification rate when data from all of the sensors were used. The Fisher projection is

Figure 5.8: Principal components analysis of Dataset 5 ("coffees"), which was derived from analysis of the headspace of 7 different coffee blends. Each of the blends was nominally collected under the same conditions.



known to be suboptimal for more than two classes because it tries to preserve distances of classes that are already well-separated, thereby often resulting in a large overlap of, or even occlusion between, neighboring classes.[24] Hence, for such multi-class problems, non-linear discriminants have a performance advantage over the FLD approach.

5.4.6 Classification of Coffees

Figure 5.8 displays a PCA plot of Dataset 5 ("coffees"), revealing effectively no linear or Gaussian character in any analyte cluster or PC dimension. Consequently, a significant disparity was observed between MLP and FLD, with the algorithms producing 86% and 57% correct classification performance, respectively. FLD is most efficient for binary separation problems; however, even when only two of the coffees in Dataset 5 were selected, MLP significantly outperformed FLD (100% vs. 89% correct classification) due to the form of the data in this Dataset.

5.5 Discussion

Overall, the Fisher linear discriminant method yielded good to excellent classification performance even in these extremely challenging discrimination tasks. The FLD method is relatively simple

to implement computationally and also imposes a relatively low burden on a training set while allowing for facile updating of the database in response to the addition of response patterns for new analytes. Hence, the FLD is an algorithm of choice for sensors and tasks that are well matched to the use of linear discriminant methods for classification. Clearly, the success of the FLD approach in the present examples is related to the linear response vs. analyte concentration behavior of the polymer/carbon-black composite vapor sensors.

In contrast, the MLP method (and by extension, other ANN methods) has the flexibility to establish highly nonlinear decision boundaries, as evidenced by its success with the interferents, multiple concentrations, and coffees datasets. The drawback of using such an approach, however, involves the number of fittable parameters associated with such models. For a multi-class implementation of FLD, a dataset consisting of D features and C classes will require $D \times (C-1)$ fittable parameters. Assuming no preprocessing, the simplest possible ANN model, utilizing no hidden layer, would require $D \times C$ parameters. In practice, such methods rarely prove useful, and hidden layers are nearly always necessary. For typical fully connected, feed-forward ANN methods (such as the MLP investigated here), the number of fittable parameters can be found from eq 5.2, in which n represents the number of layers (including inputs and outputs) and $size_l$ represents the number of fittable parameters in layer l.

$$parameters = \sum_{l=1}^{n} size_l \times size_{l+1} \tag{5.2}$$

Here, the size of the first layer (inputs) is equal to D, and the size of the last layer (outputs) is equal to C. Even assuming a small number of inputs and classes, of 5 and 3 respectively, FLD will require 10 fittable parameters, whereas an ANN with no biases or hidden layers will require 15 parameters. However, adding even a single modestly sized 4 units hidden layer increases the number of fittable parameters to 32.

As a result, ANN systems must typically reduce the number of inputs for all but the smallest feature sets. In the case of MLP technique employed here, preprocessing using PCA reduces the number of features from 40 to 5. Reducing the number of fittable parameters is critical to reducing the training burden, as more complicated models require more training data to establish a necessary goodness of fit. Alternatively, for a training set of a given size, a more complicated model increases the risk of overtraining, in which the model is attuned to non-systematic information in the training set and thus develops a fit to information that will likely not be present in any test set, degrading the ultimate predictive classification performance of the model.

One could reasonably thus assume that the simplest model that performs a given task appropriately is the model of choice. Because FLD typically involves far fewer features than do ANN methods, it would likely be more appropriate for better-controlled systems in which the data clus-

ters have at least some hyperplane that provides a linearly separable decision boundary for the task of interest. Additionally, for systems in which relatively small datasets are available, FLD would be preferred because of the reduced risk of overtraining. However, for systems in which the variance is expected to be highly nonlinear and non-Gaussian, and for which relatively large data libraries are available, more flexible models such as ANNs will likely be required to obtain good classification performance.

This process is not without cost, as shown by Dataset 2 in this study, in which the fifth (and last) principal component contained the bulk of useful information for separating the two analytes. Had the very last PC not been recognized to include such useful information, it might have been rejected in the MLP method (as is often the case), and the MLP performance on this task may have suffered. In this way, ANN methods often reduce the number of fittable parameters required by trading some information loss for greater flexibility in the resulting model. This type of trade is typically beneficial for nonlinear, non-Gaussian data clusters.

The possibility that sub-optimal sensor arrays will be chosen for real-world use must also be considered. For the datasets considered in this study, FLD showed similar classification performance for optimally performing 5-sensor subsets on a given task to the performance of the MLP in conjunction with its optimally performing 5-sensor subset (with the exception of the extremely nonlinear Datasets 4 and 5). However, even for tasks in which FLD performed similarly to MLP, fewer sub-arrays showed near-optimal classification performance with FLD than with MLP. This is likely because, while some of the sensors showed at least somewhat Gaussian responses to a given analyte set, and some sensors were selective with regard to those analytes, few sensor sub-arrays were able to combine Gaussianity with selectivity to yield high performance. As such, not only did FLD perform very poorly in highly nonlinear classification tasks, it required optimal sensor arrays (or large sensor arrays) to perform well with datasets that were somewhat nonlinear. Thus, more plastic models such as ANN provide much improved classification performance for data having significantly less than ideally distributed variance.

5.6 Conclusions

The Fisher linear discriminant method performed as well as a non-linear multi-layer perceptron approach in most of the tasks evaluated in this work. The non-linear MLP method yielded superior performance for multi-class problems and in some other instances. The degree to which the MLP method outperformed the FLD approach varied by the nature of the task. Binary classification tasks that contained data clusters that were largely Gaussian in all or most dimensions were almost optimally classified by FLD, as proscribed by theory. However, tasks that were multi-class, or which contained a number of highly non-Gaussian dimensions could only be optimally solved by

a nonlinear method, for example the MLP. Although a nonlinear method is generally preferable from a performance perspective, such methods are generally more computationally intensive and require a greater ratio of samples/features to prevent overtraining. Therefore, for systems in which many features are required, relatively few samples are available, or little computational power is available, a linear method such as FLD is preferable. Clearly, the success of the FLD approach in the systems evaluated in this work is related to the linear response vs. analyte concentration behavior of the polymer/carbon-black composite vapor sensors used to generate the array response patterns for the various tasks of interest.

Bibliography

[1] Nakamoto, T.; Fukunishi, K.; Moriizumi, T. *Sensors and Actuators* **1990,** *1,* 473–476.

[2] Grate, J.; Patrash, S. *Analytical Chemistry* **1995,** *67,* 2162–2169.

[3] Albert, K.; Lewis, N.; Schauer, C.; Sotzing, G.; Stitzel, S.; Vaid, T.; Walt, D. *Chemical Reviews* **2000,** *100,* 2595–2626.

[4] Jurs, P.; Bakken, G.; McClelland, H. *Chemical Reviews* **2000,** *100,* 2649–2678.

[5] Ballantine, D.; Rose, S.; Grate, J.; Wohltjen, H. *Analytical Chemistry* **1986,** *58,* 3058–3066.

[6] Shurmer, H.; Gardner, J.; P., C. *Sensors and Actuators B - Chemical* **1990,** *1,* 256–260.

[7] Patrash, S.; Zellers, E. *Analytical Chemistry* **1993,** *65,* 2055–2066.

[8] Lonergan, M.; Severin, E.; Doleman, B.; Beaber, S.; Grubb, R.; Lewis, N. *Chemistry of Materials* **1996,** *8,* 2298–2312.

[9] Dickinson, T.; White, J.; Kauer, J.; Walt, D. *Nature* **1996,** *382,* 697–700.

[10] Vaid, T.; Burl, M.; Lewis, N. *Analytical Chemistry* **2001,** *73,* 321–331.

[11] Sisk, B.; Lewis, N. *Sensors and Actuators B - Chemical* **2004,** *104,* 249–268.

[12] Duda, R.; Hart, P. *Pattern Classification and Scene Analysis;* John Wiley and Sons: New York, 1984.

[13] Severin, E.; Lewis, N. *Analytical Chemistry* **2000,** *72,* 2008–2015.

[14] Park, J.; Groves, W.; Zellers, E. *Analytical Chemistry* **1999,** *71,* 3877–3886.

[15] Burl, M.; Sisk, B.; Vaid, T.; Lewis, N. *Sensors and Actuators B - Chemical* **2002,** *87,* 130–149.

[16] Koscho, M.; Grubbs, R.; N.S, L. *Analytical Chemistry* **2002,** *74,* 1307–1315.

[17] Sberveglieri, G. *Sensors and Actuators B - Chemical* **1995,** *23,* 103–109.

[18] Pardo, M.; Sberveglieri, G. *IEEE Transactions on Instrument and Measurement* **1995,** *51,* 1334–1339.

[19] Bishop, C. *Neural Networks for Pattern Recognition;* Oxford University Press: Oxford, UK, 1995.

[20] Sarle, W. In *Symposium of the Interface of Computer Science and Statistics*; 1995.

[21] Forsee, F.; Hagan, M. In *Joint International Conference for Neural Networks*; 1997.

[22] Pardo, M.; Sberveglieri, G. *IEEE Sensor Journal in press,*.

[23] Demuth, H.; Beale, M. *Manual of the Neural Network Toolbox;* The Mathworks, Inc.: 1998.

[24] Loog, M.; Duin, R.; Haeb-Umbach, R. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2001,** *23,* 762–766.

Chapter 6

Comparison of Analytical Methods and Calibration Methods for Correction of Detector Response Drift in Arrays of Polymer/Carbon-black Composite Vapor Detectors

6.1 Abstract

The responses of 15 polymer/carbon-black composite chemiresistors have been analyzed during exposure to 8 different analytes (n-hexane, tetrahydrofuran, ethanol, ethyl acetate, cyclohexane, n-heptane, n-octane, and isooctane) in random order at low concentration (0.5% of the vapor pressure of analyte at room temperature) over 4 months (8000 total analyte exposures) of data collection. Data were collected for periods during which the array was continuously exposed periodically to analytes and after long periods during which no analyte exposures had been performed. All but the most difficult separation tasks (for example, discrimination between low concentrations of straight-chain hydrocarbons) could be performed robustly over the entire 4 month time period based only on the use of a decision boundary formulated from an initial training set of 200 exposures, indicating the sensor drift had minimal effect on system performance in such classification tasks. For the remaining classification tasks, modeling the dynamics of sensor drift either through a linear regression or Fourier transform decomposition of the individual relative differential resistance responses vs time of each sensor yielded little improvement in classification performance, indicating that external events were largely responsible for changes in sensor response vs time. Six analytes that were not treated as unknowns for a binary separation task were individually treated as calibrants whose response was intermittantly used to renormalize the response of the sensor array. A simple linear sensor-by-sensor calibration scheme proved effective at restoring the classification performance of

difficult binary separation tasks to the performance that was observed in the initial training set period. Calibrants that were mutually similar to the analytes being differentiated tended to be more effective than calibrants that were very chemically different from the analytes of interest. Evaluation of various calibration protocols indicated that an optimal tradeoff existed between the number of calibration exposures and the frequency of calibration periods. Condition-based calibration, in which calibration was only performed when the classification model exhibited a decline in classification performance below a predetermined threshold value, was observed to be superior to a time-based calibration approach or to interval-based, cyclic calibration protocols for this set of analytes exposed under the chosen analysis conditions.

6.2 Introduction

Robust vapor sensing systems all have some method of algorithmically correcting or physically minimizing sensor drift. For example, flame ionization detectors are calibrated with known standard gas mixtures, and near-infrared detectors are typically calibrated with calibrant sets[1] or even with a single analyte.[2] Recently, arrays of broadly cross-reactive sensors have received significant attention for their possible use in detection and classification of analyte vapors. These systems can be based on many signal transduction modalities, including polymer-coated quartz-crystal microbalances (QCM) or surface-acoustic wave (SAW) devices,[3–5] glass beads or optical fibers coated with dye impregnated polymers,[6–10] conducting polymer[11–13] or polymer composite[14–16] chemically sensitive resistors, polymer-coated micromachined cantilevers,[17] polymer-based capacitors and FETs,[18,19] and metal oxide chemiresistors.[20–23]

To maximize sensor diversity, the sensors in such systems are typically as uncorrelated as possible. However, sensor drift can ultimately invalidate some of the classification models, requiring correction or retraining. The effects of sensor drift can be significant, and complex to treat analytically, because of the large number of ways in which the array response can change with time. Without correction for sensor drift, analyte recognition libraries may be limited in size because of the significant burden that frequent retraining of large libraries would impose on the user. The best approach would be to minimize the drift in the sensors themselves, only applying corrections or calibrations to situations in which it was demanded for the successful application of the technology to a specific problem of interest.

Because drift is such a serious problem in many types of sensors, often preventing reliable analyte identification over long timescales, a great deal of work has been directed towards the development of drift correction methods. Effectively, the solutions fall into three categories: a) attempts to attune classifiers to signal while ignoring drift; b) the use of adaptive models that update the classifier based on assigned identities; and c) the use of a calibrant to return the classifier to its original

state. Attempts to attune the classifier to specific signals of interest have been used in conjunction with either independent components analysis (ICA)[24] or principal components analysis (PCA)[25] to determine which dimensions of the analyte space correlate most highly with the differences between the analytes in the set. These presumably represent the dimensions that are least noisy and/or are least affected by drift, and are the only dimensions retained in constructing a classification model. The attuning methods can provide significant improvements in classification over a fixed time period. However, it is not clear what would be done upon addition of new analytes to the recognition library, as the previously rejected dimensions might be necessary to robustly identify new analytes. Additionally, these methods contain no provisions for updating the model, and thus may ultimately be invalidated by sensor drift, requiring some sort of retraining or calibration.

Adaptive models, frequently neural networks such as self-organizing maps (SOM), have also been useful, because newly recognized data that match the stored analyte signatures can be continuously used to retrain the classifier.[26–29] This technique has the advantage of simplicity because no actual recalibration is performed by the user. However, most of the analytes encountered must be pure, to ensure that the sensors, rather than the analytes, drift in time. Also, the drift must be gradual, as a discontinuity in response between consecutive exposures (regardless of the time interval between the exposures) would immediately invalidate the classification model and would prevent adaptation.

Use of a calibrant or set of calibrants to retrain a classifier is a more time-intensive method of drift correction, but may well be the only robust method for determining precise information regarding the degradation of the classification model regardless of the degree of sensor drift. Ultimately, calibration is also therefore the only method that is capable of sustaining a high degree classification performance in the face of inconsistent sensor drift over extremely long time periods and with potentially intermittent use. Calibration methods have proven successful for transferring classifier models between sensor arrays, and presumably this approach could also be used to correct for drift in a given sensor array, because after a sensor array drifts sufficiently, it effectively becomes a new, slightly different array.[30] Neural methods, which are computationally intensive, might be less suited to drift correction of arrays having large numbers of sensors than they are to a one-time array transfer. Frank et al. developed a method of using an array of calibrants to periodically update a classifier by relating the linear change of the each single sensor to that of the average sensor.[31] This method has the advantage of allowing for retraining whenever necessary until the sensors cease providing meaningful output signals.

The goal of this work was not to develop an optimal classifier or to develop the best method for transferring/updating that classifier naturally, different classification and transfer schemes are more useful in different situations. Rather, the goal of this work was to explore analytically the effects of drift on quantification and classification of analytes in an array of chemically sensitive

vapor detectors formed from composites of carbon-black and insulating organic polymers. Such sensor arrays have been shown to yield excellent classification performance between organic vapors and additionally have been shown to provide a relatively linear response as a function of analyte concentration. Ultimately, we seek to determine an acceptable method for updating a classifier at the lowest level possible a single sensor at a time. Such a method would be a useful transfer method in itself, or could be used as a pre-processing stage before application of most of the techniques mentioned above. In this work, we have analyzed the response of a 15-detector array over a 120-day time period to 1000 exposures each of 8 different analytes (8000 total analyte exposures).

A specific question of interest is how important is drift in affecting the quantification and classification of analytes based on the response signals of polymer/carbon-black composite vapor detector arrays. Fishers Linear Discriminant[32] was selected as the classifier of choice, because it has been previously shown to be effective for the types of sensors and classification tasks considered in this work.[33] In principle, many other non-reductive schemes (i.e., not PCA or ICA) could have been used, so the approach described herein should be applicable to other classifiers as well. We have determined how much of the sensor drift is linear vs. time for each sensor as well as other functional forms that can be corrected analytically at the sensor level. Additionally, we have evaluated how often one needs to perform calibration and for what period in order to develop a successful sensor-level-based drift-correction model. For the component of sensor array drift that is not analytically correctable at the sensor level, we have evaluated whether the array response can be calibrated with a known standard analyte, and if so, how similar the calibrant should be to the test analytes in order to be useful for updating the classifier model. Finally, we have analyzed whether calibration should be performed at regular intervals or whether analytical methods can be used to indicate when calibration is necessary.

6.3 Experimental

6.3.1 Detectors

The detector array used in this study consisted of 15 compositionally distinct polymer/carbon-black composite chemically sensitive resistors (Table 6.1). Detector films were cast from mixtures of 40% polymer, 40% di(ethylene glycol) dibenzoate (a plasticizer), and 20% by weight of carbon-black (Black Pearls 2000, Cabot Inc), as described previously.[15,16,34] The detector films were deposited between two Au leads that had been evaporated onto a glass slide, and the array was housed in a stainless steel assembly that was connected by Teflon tubing to a computer-controlled, calibrated vapor generation and delivery system.[15]

Table 6.1: Detectors used in this study.

1	poly(caprolactone)
2	poly (ethylene-co-vinyl acetate)
3	poly(ethylene oxide)
4	poly(ethylene glycol)
5	poly(styrene-co-butadiene)
6	poly(methyl octadecylsiloxane)
7	poly(vinyl stearate)
8	ethyl cellulose
9	poly(methyl vinyl ether-co-maleic anhydride)
10	poly(4-vinyl phenol)
11	polycarbonate
12	polystyrene
13	poly(methyl methacrylate)
14	poly(vinyl butyral)
15	poly(styrene-isoprene)

6.3.2 Analytes and Data Collection

Five hydrocarbons (cyclohexane, n-hexane, n-heptane, n-octane, and isooctane) and three other analytes (tetrahydrofuran, ethanol, and ethyl acetate) were presented to the detectors. Within a single run, these eight analytes were presented 200 times each, in random order, to the detector array. Each analyte presentation consisted of 70 s of clean laboratory air, followed by 80 s of analyte vapor, followed by another 60 s of clean air to purge the system. The laboratory air contained 1200 ppm of water vapor. All analytes were presented to the detector array at concentrations of approximately P/P°=0.0050, where P is the partial pressure and P° is the vapor pressure of the analyte at room temperature (21±1°C). Five total runs of 200 exposures per analyte were performed, with each run requiring approximately 4 consecutive days of data collection. Breaks of less than 1 week occurred after runs 1 and 2, while longer breaks of 70 days and 20 days followed runs 3 and 4, respectively.

6.3.3 Data Pre-processing

The response of a vapor detector to a particular analyte was expressed as $\Delta R_{eq}/R_b$, where R_b is the baseline resistance of the detector in the absence of analyte, and ΔR_{eq} is the baseline-corrected steady-state resistance change upon exposure of the detector to analyte.[35] Baseline correction was performed by fitting a spline to the data obtained during the pre-exposure period, and subtracting the spline over the entire exposure. In some cases, the responses from each of the datasets were subsequently sum-normalized to remove any linear concentration dependence on the resulting

array signals. This process was performed using eq. 6.1

$$S'_{ij} = \frac{S_{ij}}{\sum_{j=1}^{n} S_{ij}} \tag{6.1}$$

where S_{ij} refers to the R/R_b response signal of the j^{th} detector (out of n total detectors) to the i_{th} analyte exposure, and S'_{ij} represents the sum-normalized analog of S_{ij}.

6.3.4 Quantification of Classification Performance

The Fisher Linear Discriminant (FLD) algorithm was used to analyze the data for classification performance.[32] In the FLD approach, analyte exposures in a training set are used to select a hyperplanar decision boundary that maximizes the separation between the two data clusters of interest. For normalized data (eq. 6.1) produced by the responses of an n-detector array, this hyperplane (or classification model) has the form:

$$D_i = \sum_{j=1}^{n-1} c_j S'_{ij} \tag{6.2}$$

where c_j represents one of the $n-1$ weighting factors from the hyperplane determined by the FLD algorithm. The value of D_i (hereafter referred to as the "D-value") is a single, scalar metric that characterizes the position, along a vector normal to the hyperplane discrimination boundary, of the detector array data produced by an individual analyte exposure. The function of the FLD algorithm is to maximize the separation, or clustering, of the two distinct populations of D-values that arise from a single binary separation task. This clustering is measured by the resolution factor (rf) characteristic of a separation task, as given in eq. 6.3:[32]

$$rf = \frac{\delta}{\sigma_1^2 + \sigma_2^2} \tag{6.3}$$

Here, δ is the difference in the population means of D-values, and σ_1 and σ_2 are the standard deviations of the two populations of D-values that correspond to the two analytes of the separation task. The FLD algorithm was used to evaluate the separation of two analytes at a time for each possible combination of two analytes in the data set. Because a supervised algorithm inherently introduces some bias into the analysis, a train/test scheme was employed. For each pair of two analytes that comprised a single separation task, the first 200 exposures to each analyte were used to generate a training set, and a set of coefficients (comprising a classification model) as described in eq. 6.2 was established. Population means and standard deviations based on the characteristic Fisher scalar values were also calculated. A decision boundary was then developed by defining the hyperplane at which an unknown analyte exposure would have an equal probability (according to

eq. 6.3) of belonging to either analyte population of the given binary separation task. All subsequent data were treated as test data, in that the Fisher algorithm was not performed after the training phase, and analyte identities were classified according to their positions relative to the fixed FLD decision boundary.

6.3.5 Calibration

For any given experiment, two of the eight analytes were treated as unknowns, with their identities masked to the classification algorithm. A third analyte was treated as a calibrant, with its identity known to the algorithm. To insure that all linear variance, consisting of (for example) incidental concentration variations, was already removed from the dataset, only normalized data were used for calibration. Calibration was performed by determining the mean and normalized response of a calibrant to a specific detector during training, and then determining these quantities during some later time interval (or, even, to a single later calibrant exposure). The response of each detector was adjusted by a multiplicative calibration factor to produce the same amplitude on the calibration data as would have been observed during the initial exposures to the calibrant (eq. 6.4):

$$S_{a,t} = S_{c,t} \times \frac{S_{a,0}}{S_{c,0}} \tag{6.4}$$

$S_{a,t}$ and $S_{c,t}$ indicate the $\Delta R_{eq}/R_b$ response signals for an analyte a or calibrant c, respectively, at an arbitrary time t after training. The original FLD coefficients established during training ($t = 0$) can be used indefinitely if the theoretical response of a detector at $t = 0$ is correctly predicted from its actual response at time t using the relationship in eq. 6.4. Better correlation between Sa and Sc over substantive time intervals will allow for a better fit over time between the calibrated data and the original FLD model. Calibrant data were not collected strictly contemporaneously with exposures to unknown analytes, so a block of data was used for calibration and the updated Fisher model was then used to classify the analytes whose identities were masked during development of the model. This scheme was used to simulate likely use in the field, in which calibration after each unknown exposure is not practical.

6.3.6 Method of Evaluation

Typically, for a binary separation, the only information of interest is the identity of the unknown analyte, with the knowledge that the unknown belongs to one of the two populations. Once the analyte identity has been established by determining its location relative to the FLD decision boundary, the analyte concentration can be determined using a linear regression because the response of polymer/carbon-black composite detectors is linear with analyte concentration over a significant range of conditions. In this approach, a significant amount of detector drift could in principle occur

without affecting the analyte classification performance, if the drift occurs in a direction completely parallel to the decision boundary. To account for components of drift perpendicular to the boundary, any deviation of an analyte away from the original population means determined during training was penalized, regardless of whether the deviation pushed the population toward or away from the decision boundary. To accomplish this, the root-mean-square (RMS) values of the deviations between the actual and predicted mean-population D-values were calculated both before and after calibration. These values were treated as indicative of the effectiveness of the calibration, with the goal being a reduction of the RMS deviations. To prevent outliers from unduly affecting the results, these results were recalculated after removing from each analysis the 2% of data points having the highest and lowest deviations. These metrics will be referred to as $RMSD$ and $RMSD_{rej}$, which refer to the RMS deviations before and after outlier rejection, respectively.

6.4 Results

6.4.1 Analysis of Drift

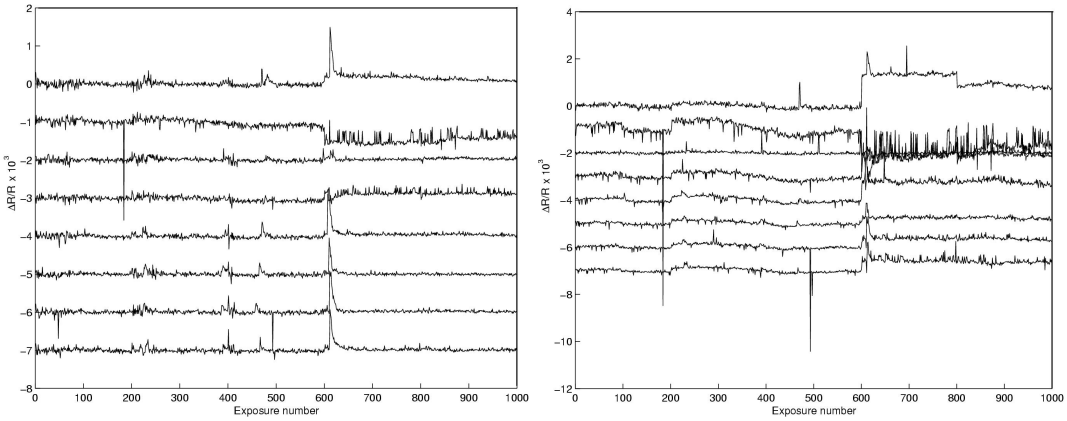
6.4.1.1 Analysis of Raw Data

At moderate analyte concentrations, previous studies have indicated that polymer/carbon-black composite vapor detector responses in conjunction with the FLD algorithm can robustly separate even extremely similar analyte pairs over extended periods of time [33, 35, 36]. To obtain a data set in which the inherent drift in detector response drift did affect classification, and therefore provide a testbed for analyzing the efficacy of various drift-correction methods, all of the analytes investigated in this study were presented at low concentration, with $P/P^\circ = 0.0050$, to the detector array. These relatively low concentration values produced less than 100% classification performance for at least some of the binary separation tasks.

Figures 6.1 a-b plot the unnormalized $\Delta R_{eq}/R_b$ responses of two sensors (detectors 2 and 3 from Table 6.1, respectively) for all 8 analytes as a function of exposure number. The drift of the sensors was clearly not continuous or monotonic, and instead the bulk of the variance clearly resulted from specific events. For example, sensor responses were clearly affected by the two-month break after exposure 600; additionally, sensor 9 (not shown) also showed significant spikes after the other, shorter breaks that followed exposures 200, 400, and 800. All of the sensors additionally exhibited less immediately explainable spikes and response changes at other times during data collection.

The average raw $\Delta R_{eq}/R_b$ response of a detector to a specific analyte changed by 71.5% of its mean value over the entire period of data collection, with the greatest change being 381% and the lowest being 14.3%. These values were based on an 11-point moving average after removal of 4 outlier data points out of the 8000 total data points. For normalized data, the average $\Delta R_{eq}/R_b$ response of a

Figure 6.1: Unnormalized (raw) $\Delta R_{eq}/R_b$ values, for all eight analytes, as a function of exposure number for a) Sensor 2 and b) Sensor 3. For clarity, each analyte data series is offset by a constant. Each plot contains 8 data series, corresponding to sensor responses to n-hexane (-1.2, -0.27), tetrahydrofuran (-5.1, -2.4), ethanol (-2.6, -2.4), ethyl acetate (-5.5, -2.9), cyclohexane (-6.1, -4.5), n-heptane (-6.6, -5.3), n-octane (-7.4, -6.3), and iso-octane (-8.6, -6.3) listed top to bottom, with constant offsets [$\times 10^3$] for sensor 2 and sensor 3, respectively, in parentheses



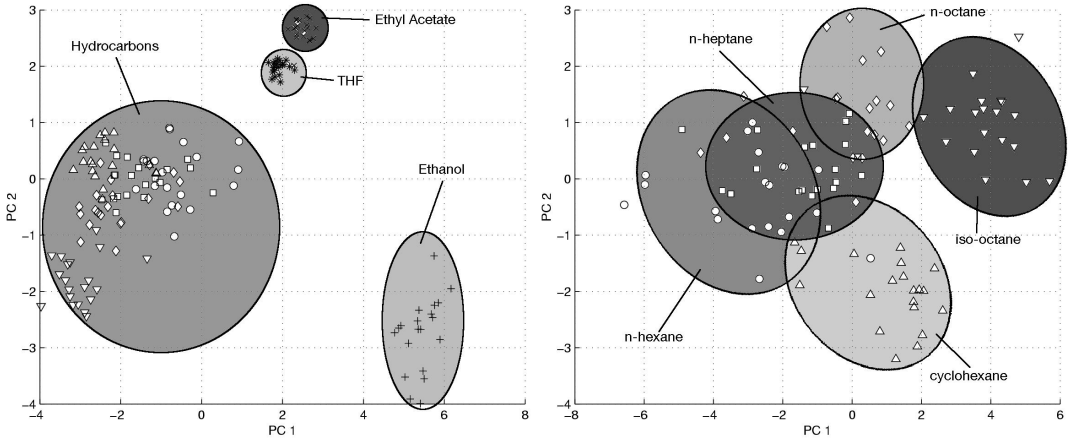
detector to a specific analyte changed by 21.3%, with the greatest change being 115% and the lowest change being 5.05%. In contrast, for the first 200 exposures to each analyte, which were collected without any discontinuity in time between subsequent exposures, the highest, lowest, and mean detector response changes for normalized data were 63.6%, 1.68%, and 7.85%, respectively. This result suggests that discontinuities were a major source of drift.

The normalized responses of n-hexane and ethanol on individual sensors differed by 55.6%, averaging over all sensors and considering the full range of data. This value is significantly higher than the average change in normalized sensor responses over time of 21.3%, so the performance of a classification model trained to separate n-hexane and ethanol should be reasonably unaffected by such time-dependent detector response changes. However, for a pair of analytes whose response differences are less than the threshold of average sensor change, such as n-hexane and n-heptane (which differ by only 6.50%, on average), classification performance is expected to be strongly time-dependent, and will require some correction for detector drift even within a run having no temporal break between analyte exposures.

6.4.1.2 Separability of Analytes During Training Phase

Figure 6.2 a shows data in principal components space collected for all 8 analytes during the initial training set run. Figure 6.2 b shows a similar plot only for the 5 hydrocarbon analytes. The data presented in Figure 6.2 2 are based on normalized detector responses, so changes in amplitude that may affect the ability to quantitate analyte concentration are absent from such a representation. Clearly, even at these relatively low analyte concentrations, each of the analyte clusters was robustly

Figure 6.2: Principal components analysis (PCA) plot of a) all 8 analytes considered in this analysis and b) of only the hydrocarbons. Data for this plot were taken from 20 random exposures of the first 200 acquired from each analyte, which were obtained from the first unbroken run of data collection. The hydrocarbons considered in this analysis were n-hexane (∘), n-heptane (□), n-octane (⋄),iso-octane (▽), and cyclohexane (△). Other analytes considered were tetrahydrofuran (*), ethanol (+), and ethyl acetate (×). For clarity, ellipses containing at least 80% of the data for each analyte are shown.



separated from all of the other clusters, although the hydrocarbons are mutually more similar to each other than to any of the other three analytes. This behavior is in accord with previous studies of the resolution between analytes at low concentrations on arrays of polymer/carbon-black composite vapor detectors.

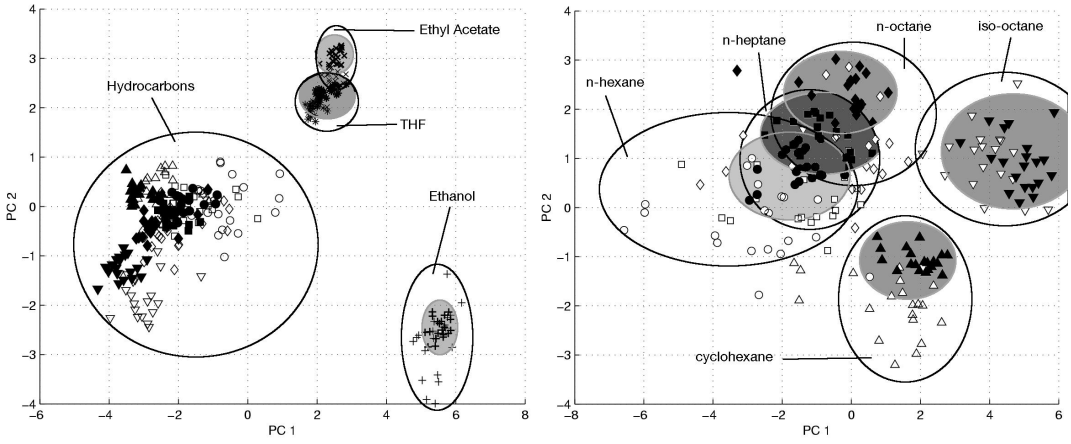
To determine the effectiveness of the Fisher linear discriminant classifier without considering any temporal differences in the data, for each pairwise analyte separation task the training set was subdivided into two nominally equivalent subsets of analyte exposures. Exposures were randomly assigned to these two subsets, ensuring only that each contained the same number of exposures to each analyte. The FLD analysis was performed on one of the subsets, and the resulting model then applied to the second subset, with the fraction of correctly predicted analyte exposures then being measured from the second subset. The subsets were then switched and the entire process repeated. The classification statistics from the two subsets were then averaged to yield an expected classification performance for the training set. This quantity can then be compared to statistics from subsequent test sets while minimizing issues related to training-set overfitting. This process was repeated for each binary analyte separation task in the data set.

Table 6.2 shows the training-set separability (as described above) of the hydrocarbon analyte pairs. All other binary analyte separation tasks produced no errors in the training period. Appreciable error rates were only present for binary separations that involved exceptionally similar analytes, such as the straight-chain hydrocarbons differing in length by only a single carbon (e.g., n-hexane and n-heptane). Other hydrocarbon separations, such as cyclohexane vs. n-octane, produced nearly perfect classification performance even at these low analyte concentration values.

Table 6.2: Train/test performance of hydrocarbon binary separations from training set (first 200 exposures).

	n-hex	c-hex	n-hept	n-oct	i-oct
n-hex	N/A	0.9950	0.8325	0.9500	1.000
c-hex	-	N/A	1.000	1.000	1.000
n-hept	-	-	N/A	0.8275	1.000
n-oct	-	-	-	N/A	0.9950
i-oct	-	-	-	-	N/A

Figure 6.3: Principal components analysis (PCA) plot of a) all 8 analytes considered in this analysis and b) of the hydrocarbons only. Data for this plot were taken from 20 random exposures to both the first and last 200 exposures acquired from each analyte. These exposure series derive from the first and fifth (last) unbroken runs of data collection, respectively. Hydrocarbons considered in this analysis were n-hexane (○), n-heptane (□), n-octane (◇),iso-octane (▽), and cyclohexane (△). Other analytes considered were tetrahydrofuran (*), ethanol (+), and ethyl acetate (×). Data points collected during the final run are filled/bolded. For clarity, an unfilled ellipse is drawn for each analyte containing at least 80% the data corresponding to both runs. Within each unfilled ellipse, a shaded ellipse is drawn containing the data from the final run.

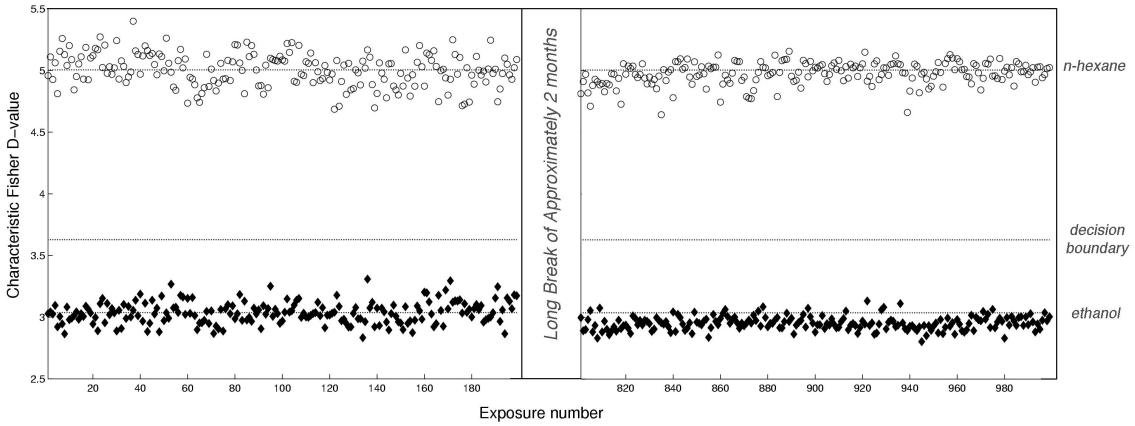


Surprisingly, n-octane vs. i-octane also produced near-perfect classification performance, suggesting that the detector array is highly tuned to analyte structure as well as analyte functionality and molecular weight. The FLD analysis therefore quantifies and largely corroborates the PCA data for the training set exposures.

6.4.1.3 Separability of Analytes After the Training Phase Using Uncorrected Data

Figure 6.3 a shows PCA data for the 5 hydrocarbons collected during two different runs, the first and the last, between which over 3 months elapsed. The PCA rotational matrix that was used to map the original, autoscaled data onto the PC axes was determined from the original (first run) data, and the 3-D viewing angle was chosen to maximize the separation of the 5 analytes in the training set. Then, the rotational matrix and viewing angle were applied without modification to the subsequent (last run) data set. Figure 6.3 b shows similar results for all 8 analytes.

Figure 6.4: Waterfall plot of the FLD "D-values" as a function of exposure number for the n-hexane/ethanol binary separation task. The first 200 exposures were used to train the model, yielding the population means for the n-hexane (○) and ethanol (◆) clusters as well as the decision boundaries between the two clusters. D-values for the final 200 exposures (801-1000), the last collected as a part of this study, were determined by applying the FLD model determined during training to the normalized $\Delta R_{eq}/R_b$ data for each exposure. Dotted lines are shown representing the mean D-value of each analyte and the decision boundary between them as determined during training.

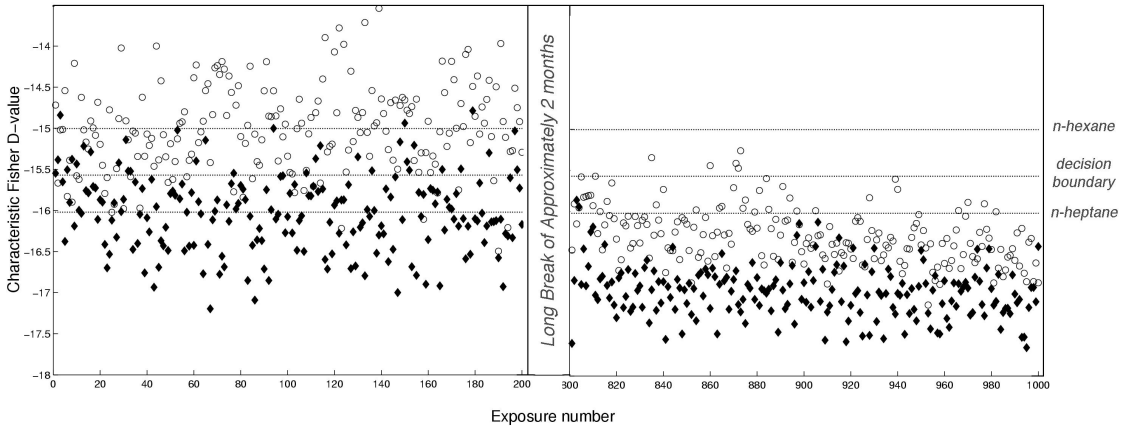


The array responses to polar analytes clearly were sufficiently different from each other and from the straight-chain hydrocarbons that, even at relatively low analyte concentration, sensor drift did not affect classification performance over a nearly 4-month period. However, under such conditions the hydrocarbons exhibited better clustering by run than by analyte. Hence for hydrocarbon separations some type of drift correction approach is needed to retain the initial classification performance throughout the entire period of data collection.

A quantitative representation of the change in separation distance between analyte clusters is presented in Figures 6.4 and 6.5 for two limiting cases of performance. Figure 6.4 presents a waterfall plot in which the array response data for a separation of n-hexane and ethanol are projected onto the Fisher discriminant vector that was obtained by use of a training set of 200 exposures/analyte for the task of interest. The analytes are very well separated, and essentially perfect classification can be achieved using the indicated decision boundary arising from the FLD. Data projected onto the FLD vector were chosen from the first 200 responses (on which the FLD model was based) and last 200 responses to each analyte, with a gap of approximately 100 days and 4800 analyte exposures between the two groups. Even after this period of months and over 8000 analyte exposures, the data did not drift sufficiently to invalidate the initially developed classifier.

In contrast, Figure 6.5 shows a corresponding plot for a separation of n-hexane and n-heptane. Clearly, this separation task produces more classification errors, particularly for the last 200 exposures. Classification performance using a fixed decision boundary is additionally more sensitive to

Figure 6.5: "Waterfall" plot of the FLD "D-values" as a function of exposure number for the n-hexane/n-heptane binary separation task. The first 200 exposures were used to train the model, yielding the population means for the n-hexane (○) and n-heptane (◆) clusters as well as the decision boundaries between the two clusters. D-values for the final 200 exposures (801-1000), the last collected as a part of this study, were determined by applying the FLD model determined during training to the normalized $\Delta R_{eq}/R_b$ data for each exposure. Dotted lines are shown representing the mean D-value of each analyte and the decision boundary between them as determined during training.



drift in the detector response, as expected from the PC plot of Figure 6.3 a.

Table 6.3 shows that, even after sensor drift over an extended time period, very few errors are made when comparing analytes that do not share a functional group. In many cases, few errors are made even for the straight-chain hydrocarbon separations. The n-hexane/THF separation was a statistical anomaly that arose because the THF population cluster was much tighter than that of n-hexane; consequently, the decision boundary produced by FLD analysis on the training set was located so much closer to the THF population that a slight detector response drift shifted the population over the decision boundary. Another decision boundary could clearly have been drawn (with less statistical validity on the initial training set data) that would have yielded essentially perfect classification performance for this separation over the entire data set. Other than this anomaly, however, only binary separations between very chemically similar analytes presented at very low concentrations to the detector array generated classification errors even after significant elapsed time and use of the detectors.

6.4.2 Statistical Analysis of Correlations Between Detectors

To correct for the signal drift in the array of polymer/carbon-black composite sensors and thereby improve the classification performance under the most demanding analyte separation conditions, the nature of the drift must be understood statistically, if not phenomenologically. It is clearly of interest to ascertain whether the drift in response of the entire set of detectors and/or analyte

Table 6.3: Performance of all separation tasks from final 200 exposures.

	n-hex	THF	ethanol	EtOAc	c-hex	n-hept	n-oct	i-oct
n-hex	N/A	0.5875	1.0	1.0	1.0	0.5125	0.5725	0.8900
THF	-	N/A	1.0	0.9875	1.0	1.0	0.9875	0.9650
ethanol	-	-	N/A	1.0	1.0	1.0	1.0	1.0
EtOAc	-	-	-	N/A	1.0	1.0	1.0	1.0
c-hex	-	-	-	-	N/A	1.0	1.0	0.5000
n-hept	-	-	-	-	-	N/A	0.5050	0.6725
n-oct	-	-	-	-	-	-	N/A	0.5900
i-oct	-	-	-	-	-	-	-	N/A

exposures can be estimated from the behavior of a subset of sensors and/or analytes. Additionally, the time scale over which such predictable drift manifests itself is of interest.

6.4.2.1 Correlation Between Sensors Exposed to Each Individual Analyte

To investigate whether any correlation existed between the drift of the various individual sensors, a 15×15 correlation matrix was developed from a base 8000×15 matrix of unnormalized sensor responses vs. time. Because the global data matrix contained the responses of the sensors to eight different analytes, the global mean response to each analyte/sensor combination was subtracted from each individual sensor response. In this way, only the changes in response vs. time were analyzed, independent of the mean response of any sensor to a particular analyte. The mean correlation (averaged over all possible sensor combinations) of unnormalized data was found to be 0.9993 over the full 8000 analyte exposures. In contrast, the mean correlation between sensors after normalization was 0.1459. The large decrease in correlation after normalization implies that most of the drift was linearly correlated between sensors. This strongly suggests that the correlated variation was a result of systematic external perturbations such as slight changes in analyte concentration at nominally the same settings of the mass flow controllers, temperature fluctuations, or other external variables that affected the response of essentially all of the sensors in the same fashion at the same time.

Next, the data were split by analyte, to form a series of 1000 exposures to each analyte. The analysis described above was then independently repeated for each of the 8 analytes tested, yielding 8 15×15 correlation matrices. On average, the correlation before normalization for any pair of sensors exposed to a single analyte was 0.7826, and was 0.2896 after normalization. The unnormalized correlations are lower than those for the unsplit, 8000-exposure data series presumably because concentration differences between analytes are rendered moot upon sorting by analyte. The significantly lower normalized variance for the unsplit series indicates that a significant proportion of the variance arising from analyte diversity is nonlinear in nature, and hence not correlatable using the FLD classification model. These results indicate that the nonlinear portion of the sensor drift trends

vary significantly by analyte, and as such the trends must be evaluated either analyte-by-analyte or at most within groups of similar analytes. Ultimately, however, the greatest portion of drift is linear in nature and is easily removed by sum-normalization.

6.4.2.2 Correlation Between Analytes to Each Individual Sensor

Having determined the correlation relationships between sensors, a similar investigation was performed for the different analytes in the test set. For this portion of the study, the data were first split by sensors into 15 separate 1000×8 matrices that represented the sensor responses as a function of time for all 8 analytes. Correlation matrices can not however be directly established from these data, because the 8 analyte data series in a given matrix were not temporally synchronized in the sense that the nth exposure to analyte 1 may have occurred right before, right after, or 20 minutes apart from the nth exposure to analyte 2. A 5-point moving average was therefore constructed from the data series, assuming that, over 5 exposures, the individual analyte series should have similar temporal characteristics. This process yielded 15 new 200×8 matrices.

Averaging the correlation matrices over all analyte pairs as well as all 15 sensors yielded global values of 0.1294 and 0.2535 before and after normalization, respectively. Because normalization occurs over sensors and not analytes, there is no requirement that the normalized correlations be lower than the unnormalized values. The correlations between the normalized data tended to be higher than those between unnormalized data, presumably because any concentration-influenced variance would be completely independent of analyte identity and even independent of exposure number to a given analyte. Therefore, normalization should effectively remove a significant amount of almost completely non-correlatable variance, resulting in better correlation coefficients between the normalized analyte responses compared to the unnormalized responses.

Table 6.4 displays the 8×8 correlation matrix derived from the normalized, moving-averaged 200x8 data matrix. The relationships between individual analytes (or similar groups of analytes) tended to be stronger than between sensors. Considering normalized data, the average correlations were 0.3444 between straight-chain hydrocarbons, 0.3293 between all hydrocarbons, 0.2484 between the polar non-hydrocarbons, and 0.1971 considering all possible comparisons between a hydrocarbon and non-hydrocarbon. This clearly indicates that drift is most similar among analytes that are themselves similar. In addition to evaluating the correlations over the entire data set, correlations and variances during unbroken runs were additionally compared to those for exposure blocks that spanned multiple runs; i.e., those blocks that contained at least one major discontinuity. To this end, covariance and correlation matrices were generated for each of the 5 separate collection runs and, within each run, for each sensor. For this analysis, the first 50 exposures of each run were rejected to ensure that the sensors sufficiently recovered from the most recent discontinuity. For comparison, similar statistics were collected for blocks of 200 exposures that spanned 2 separate runs (for exam-

Table 6.4: Correlation matrix of analyte responses vs. time averaged over all sensors.

	n-hex	THF	ethanol	EtOAc	c-hex	n-hept	n-oct	i-oct
n-hex	1.0	0.32	0.061	0.33	0.40	0.35	0.35	0.26
THF	0.32	1.0	0.11	0.53	0.35	0.27	0.20	0.22
ethanol	0.06	0.11	1.0	0.11	0.044	0.040	0.040	0.46
EtOAc	0.33	0.53	0.11	1.0	0.37	0.29	0.22	0.25
c-hex	0.40	0.35	0.044	0.38	1.0	0.43	0.31	0.30
n-hept	0.35	0.27	0.040	0.29	0.43	1.0	0.33	0.31
n-oct	0.35	0.20	0.040	0.22	0.31	0.33	1.0	0.25
i-oct	0.26	0.22	0.046	0.25	0.30	0.31	0.25	1.0

ple, the block consisting of exposures 101-300, spanning the discontinuity after exposure 201). For each block, the covariances and correlations were averaged over sensors and analyte pairs, with the autocorrelations and autovariances removed from analysis.

The results of this analysis are displayed in Figure 6.6. The exposure blocks within single runs showed low average variances and correlations as compared to the values observed for the blocks that spanned runs. Three of the 4 spanning blocks showed covariances and correlations that were higher than any of the 5 non-spanning blocks, with the last spanning block being the lone exception. This indicates that specific events, here discontinuities in sampling, produced the greatest portion of the variance. However, as shown in Figure 6.6, a proportional amount of this additional variance is also reflected in the correlation statistics. Thus, information regarding the changes in one (or a few) analytes following a discontinuity could be used to predict changes in the other, unknown analytes. In this way, the problem of predicting and/or calibrating drift can be further reduced.

6.4.2.3 Correlation Trends As a Function of the Range of Exposure Data and Number of Exposures Used

Naturally, the time spent calibrating the system should be as short and infrequent as possible. Therefore, it would be preferable to have high short-term correlation between analytes. To this end, the correlation analysis described above was repeated for data subsets of all possible sizes as low as 3 data points (corresponding to 30 exposures, because this analysis, unlike the previous one, consisted of 10-point moving average data due to computer memory constraints). Additionally, to minimize bias, 20 subset samples were chosen for each subset size, spreading each of the samples out as much as possible through the data. The goal was to investigate how relations between analytes might be dependent on the range of data collected, independent (as much as possible) of the presence of sampling discontinuities.

Figure 6.7 (solid line) shows a plot of the mean correlation, averaged over all sensors and over the straight-chain hydrocarbon analyte pairs, as a function of the range of the data (number of exposures). The results indicate that the correlations show a peak at short time intervals and then

Figure 6.6: Average covariance and correlation values between all analyte pairs for specific exposure blocks of 150 or 200 exposures. 150-exposure blocks that were collected without any sampling discontinuity (■, solid line) generally showed less average covariance between sensor pairs than did 200-exposure blocks collected symmetrically around a discontinuity (■, dashed line). Additionally, most of the additional variance was correlatable, as shown from the plots of mean correlation (● dashed line reflecting sampling discontinuities, solid line none).

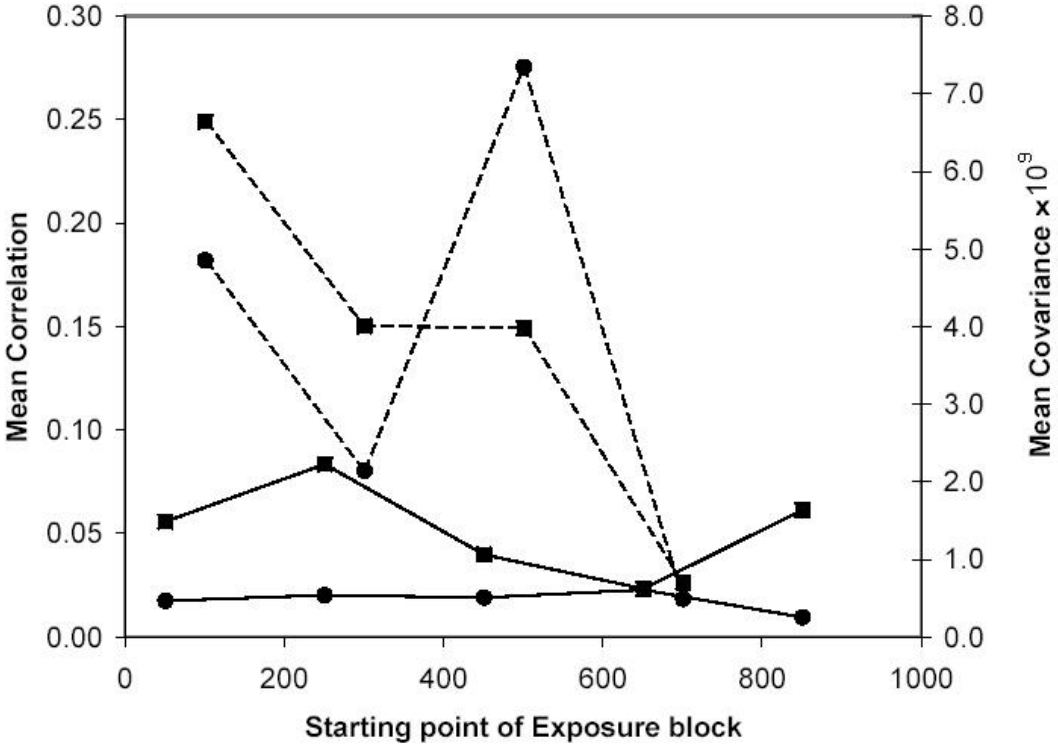
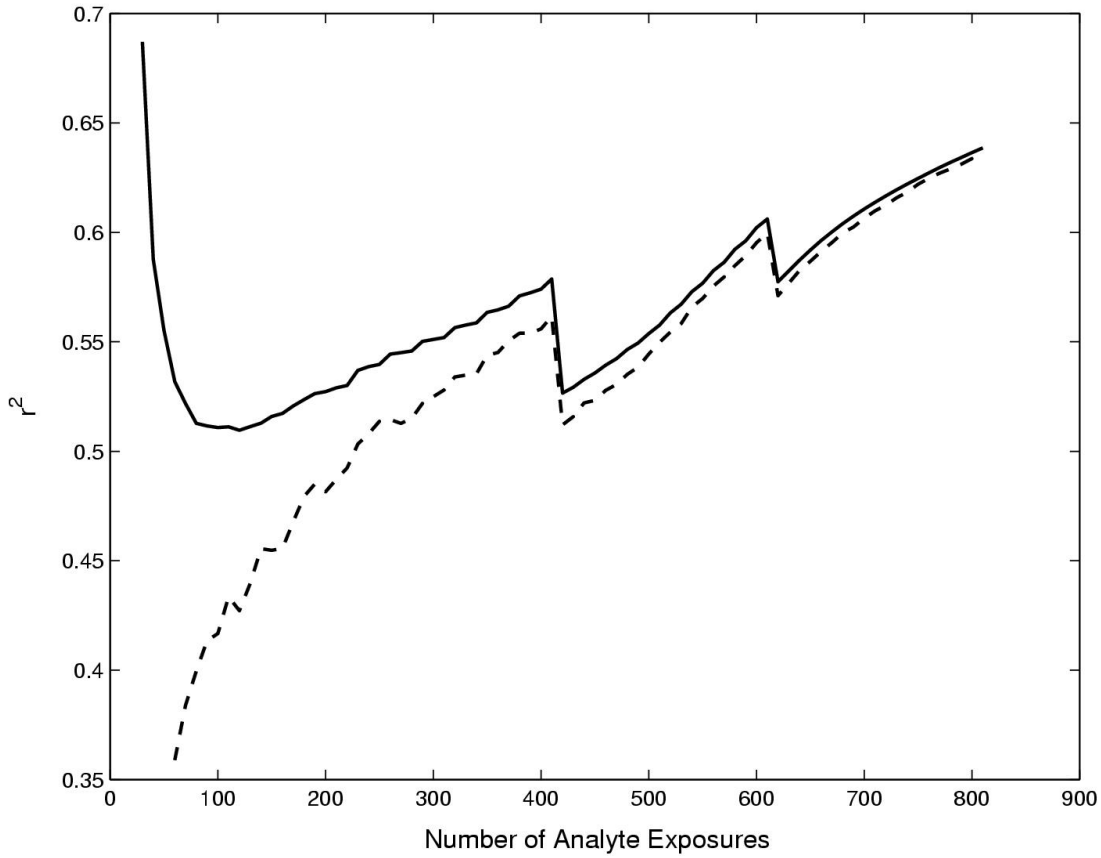


Figure 6.7: Average correlations between straight-chain hydrocarbon analyte pairs as a function of the length of the exposure window over which they were correlated. Correlation values (r^2) determined from normalized $\Delta R_{eq}/R_b$ data are shown as calculated (solid line), and as extrapolated to remove data artifacts (dotted line).



drop quickly, followed by a gradual increase with yet a larger number of exposures. However, this analysis was a bit deceptive in that fewer data points tended to show higher correlations due to overfitting. Thus the dependence of correlation on the range of the data became convoluted with the number of data points used. In other words, the desired quantity is $r^2(E)$, in which r^2 is dependent solely on the exposure interval, E; however, the r^2 yielded by the analysis is $r^2(E,p)$, where p is the number of data points per exposure interval. For the data used in this study, the most appropriate functional form between these two quantities (based on best fit) is shown in eq 6.5:

$$r^2(E,p) = \frac{a}{p} + r^2(E) \tag{6.5}$$

where both a and $r^2(E)$ are fittable parameters, and $r^2(E,p)$ and p are known. Because this analysis is extremely computationally intensive, it was only performed on the three straight-chain hydrocar-

bons considered (n-hexane, n-heptane, and n-octane), which due to their mutual chemical similarity are of greater interest than other (more well-separated) separation tasks. Figure 6.7 (dashed line) shows a plot of $r^2(E)$ as a function of the data range (in number of exposures), with $r^2(E)$ increasing effectively monotonically with the size of the exposure range. This indicates that analytes will correlate better over longer times, at which point more predictable variance is manifested. On the other hand, at shorter intervals, any variance can effectively be considered white noise, or nearly so. As such, one might expect to be able to effectively calibrate drift over longer intervals, less so over shorter intervals.

In addition to the general upward trend of mean correlations vs. exposure length, a few discontinuities are clearly present. Though the subset windows are spaced evenly, certain lengths will unavoidably begin to pick up sampling discontinuities, which have a dramatic effect on the result. This result supports the conclusion that specific events, not time or consistent use, are responsible for most of the noise in the full data set.

6.4.3 Correction of Sensor Data Without Calibration

Analytical methods of drift correction would optimally only require determining the response of a single sensor to a single known analyte and using that value to adjust the remaining sensor responses in the array. Accordingly, three methods were explored to evaluate the extent to which the drift could be corrected through a measurement-derived approach: a) the drift of individual sensors were fit to a regression analysis; b) a Fourier transform analysis was used to determine any periodicity in the drift of individual sensors, and c) the short-term analyte history of the sensor responses was analyzed to investigate whether the drift was dependent on the recent history of exposure to various analytes.

6.4.3.1 Non-periodic Functional Detrending

If the response signal of a sensor to a given analyte at time t, $S(t)$, is a function of time, and if changes in S are due to continuous processes, then it might be possible to model $S(t)$ as $S(t) = S_{t=0} + mt$, where m is equal to $\Delta S/\Delta t$. Alternatively, if $S(t)$ does not arise from continuous processes but varies due to a number of processes that are fairly consistent over relatively short time intervals, then the same model can still be used but with m and $S_{t=0}$ being defined piecewise rather than globally for each sensor/analyte combination.

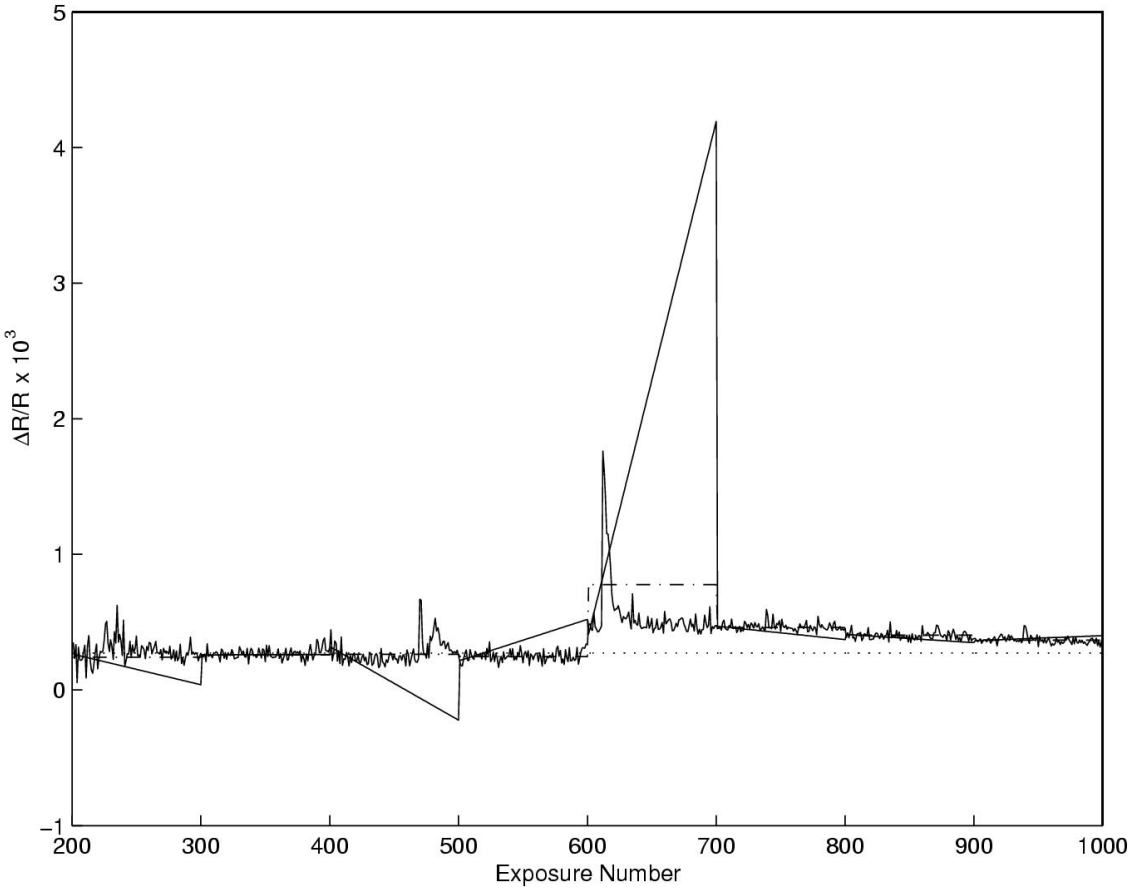
To determine whether this description of sensor response data was applicable to the data collected in this study, each of the normalized, 1000-exposure sets that corresponded to a single sensor/analyte response combination was fitted by a series of linear regressions. For each of the sets, the first 200 exposures were used to determine the mean response for that sensor/analyte

combination. Subsequently, the remaining 800 exposures were analyzed by alternately building a regression model on 20 exposures, testing that model for the next 80 exposures, and repeating the process until the end of the data set was reached.

A regression-derived model for sensor drift would only be of value if it were more predictive than simply using the mean response of a sensor over the same range of data, and should certainly be more predictive than the mean response value during the original, 200-exposure training phase. Following each regression period, the data were extrapolated using the regression model, and this extrapolation was compared to the response means measured during both the previous regression period and during the training period. After the 800 exposures had been exhausted, data from all of the extrapolation periods were collected and compared to the actual response data that was collected during the intervals of interest. Three sets of deviations (or residuals) were thus collected one for the model based on linear regression, and two based on the respective mean-response-based models. Each set of deviation values was then divided by the appropriate mean sensor/analyte response as determined during the training phase. Finally, RMS values of the deviations for each model and analyte/sensor combination were determined. These RMS values are indicative of the degree of predictive power of each model, with a high RMS deviation implying a poor model.

For the regression-based models applied to unnormalized data, the mean RMS deviation (over all analyte/sensor combinations) was 0.849. This indicates that, on average, the scatter in the regression-predicted responses was roughly 84.9% of the actual response. The corresponding values were 20.2% and 24.8% for the mean-based models based on the continually updated and non-updated mean response values, respectively. A similar analysis applied to normalized data yielded values of 34.7%, 10.1%, and 12.1%, respectively. This suggests that a frequently updated mean value is often a very capable predictor, and was better than simply using the original training-phase mean response. Additionally, the analysis indicates that a regression model introduced a significant amount of error in prediction ability. Figure 6.8 shows data that were derived from the responses of sensor 3 to hexane, overlaid with the regression model as well as both of the mean-value based models. Clearly, the occasional benefit from using the trending information did not outweigh the errors that arose in regions in which a temporary trend resulted in a model that was completely non-predictive. Many of the discrepancies in regression prediction of course originated from the breaks in data collection. However, even when the first 50 data points of each run were rejected and regression periods were chosen not to overlap breaks in data collection, the regression approach did not perform as well as simply using the most recently determined mean values of the sensor response. For normalized data, this revised method yielded fractional RMS deviations of 19.0% for regression, with lower values of 8.71% and 12.1% for the predictions based on the continuously updated and non-updated mean values, respectively. Thus, even under optimal circumstances, regression was not suitable for prediction of analyte responses, as it was worse than simply doing

Figure 6.8: $\Delta R_{eq}/R_b$ responses of sensor 3 to n-hexane at $P/P°$=0.005. Curves shown are the actual data (rough line), response predictions predicted by regression as updated using the first 20 exposures out of every 100 (solid line), by using the mean response as updated using the first 20 exposures out of every 100 (dash-dot line) and by using the mean response as determined from the first 200 exposures (dotted line).



nothing.

Both the regression and updated-mean schemes are of course impractical as methods of correcting a classifier because application of the model requires prior knowledge of the behavior of an unknown analyte. However, both schemes are indicative of the best possible performance that could be attained by similarly implemented calibration schemes because an ideal calibrant would correlate with an unknown as highly as possible, allowing their histories to be somewhat interchangeable. Hence, these results may be taken as best case scenarios for similarly implemented calibration schemes. The recent history of a known calibrant may thus be useful information if the calibrant correlates suitably with the unknown. However, the data also suggest that the responses R(t) in this study were not sufficiently reliant on continuous processes, but rather arose primarily from discontinuous and unpredictable events, as shown in Figure 1. Thus, it is not surprising that attempting regression with this dataset is actually deleterious to any predictive ability, and that

using a continuously updated mean value is preferable.

6.4.3.2 Spectral Character of Signal Variability in Individual Analyte Signal Series

Since regression analysis of the drift over short timescales lacked any significant predictive power, the variability in the $\Delta R_{eq}/R_b$ vs. time signals is expected, over the short term, to lack any significant, linearly predictable behavior. However, these sensor response series might still show periodic variability over similar timescales. To probe this aspect of the sensor response dynamics, the full, 1000-exposure signal stream from each analyte/sensor pair was deconvoluted using a fast Fourier-transform (FFT) technique.

On average, the signal spectral power vs. exposure period trends determined from FFT were highest for higher exposure periods such as 500 and 1000 exposures, suggesting that little short-term predictive ability could be expected. The spectral power was consistently low below 300 exposures for all sensor/analyte combinations, with differences between sensors in the spectral character of the signal variability arising only over longer time intervals.

In general, the spectral power increased with length of period (in exposures). The low-frequency spectral power of the signal for sensor 3 was not significantly higher than high-frequency power the average signal spectral power for sensor 3 between periods of 2 and 5 exposures was 0.0235, whereas the power at a period of 1000 exposures was 0.436 an increase of very slight magnitude and a factor of less than 20. Sensor 2, on the other hand, showed a high-frequency value of 0.0471 (similar to that of sensor 3), but a low-frequency power of 17.5 an increase factor of 372. The differences in the character of the spectral power between these two sensors can be attributed to the behavior of the two sensors after the two-month discontinuity in data collection. Sensor 3 showed a temporary discontinuity at this point, and Sensor 2 showed a seemingly permanent discontinuity. As expected from above results, the discontinuities are the greatest source of variability in the system. It should be therefore expected that inconsistent use of a detector array would yield a significant amount of low-frequency variability, varying by degree with sensor identity. Considering that the periods that showed any appreciable signal spectral power were equal to at least half the range of the data, ascribing this behavior to any actual periodicity would likely be inaccurate.

6.4.3.3 Effect of Sensor History Upon Non-correlatable Short-term Trends

The order of analyte exposure in this study was effectively random, with the pattern repeating every 1600 exposures. Any effect arising from analyte history would therefore be difficult to deconvolute without prior knowledge of the specifics of the analyte exposure protocol. To investigate these effects, the first 1600 consecutive exposures (200 per analyte) were first de-meaned by analyte and then independently sorted by the identities of the parent and grandparent analytes.

On average, exposures following high $\Delta R_{eq}/R_b$ analyte responses were themselves higher than expected given the mean $\Delta R_{eq}/R_b$ value for that analyte/sensor combination. This effect varied in that some sensors responded more rapidly or slowly than others to a given analyte, causing greater lag/overlap with the next analyte exposure. Ultimately, the grandparent effect was nearly zero; hence the parent effect was the only one of significance.

For some sensors, the parent effect was sufficient to overwhelm the differences in response between very similar analytes. For example, considering only sensor 13, the average difference in analyte exposures following a THF exposure vs. an iso-octane exposure was approximately five times greater than the actual response differences between any of the straight-chain hydrocarbon exposures. Hence, a great algorithmic reliance upon this sensor would be very detrimental to a separation task that involved similar hydrocarbons. Fortunately, because the periodicity of the $\Delta R_{eq}/R_b$ vs. time signal should be rather short, a reasonably long calibration cycle should remove such artifacts from the analysis, and a similarly reasonable training set should prevent reliance on such sensors at all. However, the short effect of history also implies that it cannot be the origin of the long-term drift. Removing this effect would be at best useful for making all separations slightly better.

6.4.4 Correction of Sensor Data with Calibration

Effectively, the data obtained in this study cannot be reliably corrected without some type of prior knowledge. Even then, recent knowledge of analyte history is insufficient to correct for drift, and functional modeling of responses is no more successful. As such, removal of drift will require fairly detailed information about the analyte/sensor combination to be corrected. Fortunately, calibration over longer time frames (over which drift manifests) is fairly correlated between similar sensors, as shown above. Therefore, we have treated certain analytes as known calibrants to correct the signal response vectors for other analytes whose identities have been masked.

6.4.4.1 Effectiveness of Calibration As a Function of Calibrant Identity

Optimal calibrants were sought for each of the following binary separation tasks: hexane vs. heptane, heptane vs. octane, and THF vs. ethanol. For each separation task, each of the 6 analytes not treated as an unknown analyte was used, individually, as potential calibrant. A cycle of 150 use exposures and 50 calibration exposures was used to test each calibrant. Table 6.5 displays the $RMSD_{rej}$ values before and after calibration for each calibrant and each task.

As expected, calibrants that are more similar to the two "unknown" analytes in question are more effective. For both the hexane/heptane and heptane/octane separations, all of the hydrocarbon calibrants yielded lower (i.e., beneficial) $RMSD_{rej}$ values than did any of the non-hydrocarbons.

Table 6.5: RMSDrej values of selected binary separations using all possible calibrants.

Separation Task	Calibrant	Fraction Correct	RMSD	$RMSD_{rej}$
n-hexane/	THF	0.6525	1.3009	1.2412
n-heptane	Ethanol	0.6512	1.3056	1.2541
	Ethyl Acetate	0.6687	1.0274	0.9358
	Cyclohexane	0.6475	1.0314	0.9236
	n-octane	0.7612	0.9004	0.7931
	i-octane	0.6588	1.0296	0.9134
	NONE	0.6687	1.2348	1.1250
n-hexane/	n-hexane	0.6900	0.8613	0.8005
n-octane	THF	0.6013	2.0537	2.0050
	Ethanol	0.6025	1.4045	1.3793
	Ethyl Acetate	0.8025	0.7196	0.6275
	cyclohexane	0.7300	0.8046	0.7931
	i-octane	0.6588	1.0296	0.9134
	NONE	0.6687	1.2348	1.1250
THF/	n-hexane	1.0	0.1266	0.1232
Ethanol	Ethyl Acetate	1.0	0.1031	0.0967
	Cyclohexane	1.0	0.2078	0.2019
	n-heptane	1.0	0.1419	0.1378
	n-octane	1.0	0.1497	0.1457
	i-octane	1.0	0.2500	0.2435
	NONE	1.0	0.6916	0.9313

Also, for a separation of THF and ethanol, ethyl acetate yielded a lower $RMSD_{rej}$ value than any of the hydrocarbons. These results indicate that using a calibrant similar to the analytes in question provides the most reliable way of correcting for drift, assuming a single calibrant is to be used. Fortunately, it is frequently possible to derive sufficient information about an unknown analyte to effectively determine its functional identity (i.e., alcohol, hydrocarbon, etc). Therefore, choosing appropriate calibrants should frequently be possible even with no prior knowledge of analyte identity.[36]

6.4.4.2 Effectiveness of Calibration As a Function of Calibration Frequency and Number of Calibration Exposures

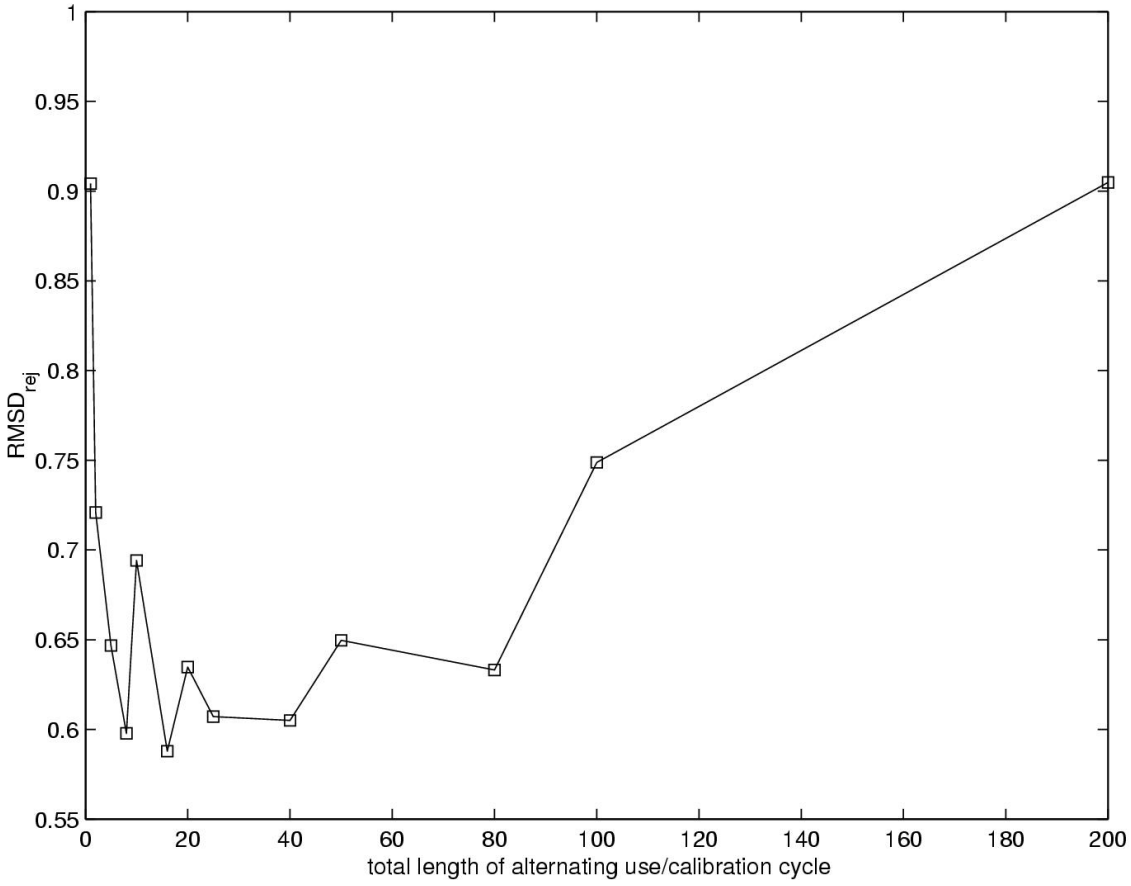
The previous FFT and correlation vs. exposure range analysis indicated that calibration should be ineffective on an exposure-to-exposure basis because the variability in response for exposures separated by so little time tended to be mostly highly temporally uncorrelated white noise. The same analysis indicated that the total variance of a series of n consecutive response signals to a single detector analyte combination increased with n, as did the correlation between two such series. This result implies that lengthy calibration would be preferable to shorter calibration. Naturally, frequent calibration should also be preferable, as the state of the system during use must be similar to the

state during calibration. Unfortunately, calibration cannot be both frequent and lengthy without unacceptably increasing the total burden of calibration upon the user. From a use standpoint, calibration should be short and a small fraction of total use, while the opposite is desired from a statistical standpoint. Assuming a predetermined fixed fraction of calibration vs. use exposures, it is not clear from the analyses described above whether short and frequent calibration is better or worse than longer but rarer calibration runs, if a compromise between length and frequency of calibration must be made.

These questions were addressed for the three most challenging separation tasks evaluated in this work, hexane vs. heptane, heptane vs. octane, and hexane vs. octane. For all three tasks, the straight-chain hydrocarbon not treated as an unknown was used to calibrate the array. A calibration/use cycle was simulated in which the first 200 exposures were used to train the model, the next x exposures of a pair of unknowns were classified by the model, and finally the next y exposures of a known calibrant were used to re-train the model. The final two steps were then repeated until all of the data had been exhausted. This method simulates a likely mode of use of a detection device in the field. The variables x and y were then varied, with RMS deviation values with and without outlier rejection (*RMSD* and $RMSD_{rej}$, respectively) determined for the entire "use" populations as a function of x and y. The variable y was adjusted to be between a single exposure and x, at which point a user would be spending as much time calibrating the device as using it.

Generally, the $RMSD_{rej}$ value was minimized by maximizing y, as might be expected the longer the calibration stage, the more robust it should be. Additionally, for a constant calibration period, a shorter use period (x) resulted in better performance. This behavior ostensibly arose because each data point is, on average, less removed from a calibration, limiting the degree of drift that could manifest itself. Therefore, a duty cycle with a higher fraction of calibration exposures would be preferred from a performance standpoint. Naturally, however, a calibration-heavy duty cycle is not preferable from a usability or economy standpoint. Therefore, it is of interest to determine the optimal cycle length ($x + y$) for a given calibration fraction. Figure 6.9 displays the $RMSD_{rej}$ vs. cycle length for a duty cycle of 50% calibration/50% use. For this analysis, the data were collected and averaged from each of the three possible tasks in which two straight-chain hydrocarbons were classified using a third such straight-chain hydrocarbon. The resulting average $RMSD_{rej}$ dropped somewhat with increasing length of the duty cycle as more consecutive calibration exposures are used. However, this trend ultimately reversed as the use exposures became too far removed from the most recent calibration, increasing $RMSD_{rej}$. The optimum calibration/use length was approximately 25 exposures, although this result is likely peculiar to the detector set and experimental conditions employed in this work.

Figure 6.9: Root mean square deviation ($RMSD_{rej}$) between calibrated D-values and those determined during training, as a function of the length of the alternating use/calibration cycle used in the system. For this plot, only binary separations of straight-chain hydrocarbons also calibrated with such were considered. Calibration was performed on a cyclic basis, varying the length of the cycle while maintaining the ratio of use/calibration exposures.



6.4.4.3 Event-driven Calibration

As illustrated in Figure 6.1, any persistent trending in the system studied was driven more by specific events than any continual process. As such, periodic calibration may prove less appropriate than calibrating after an event that renders the FLD classification model less able to correctly classify analyte members of a particular separation task. If one could determine when calibration needs to be performed, fewer exposures would be sorted with an inferior classifier model, and fewer (possibly needless) calibration cycles would be performed.

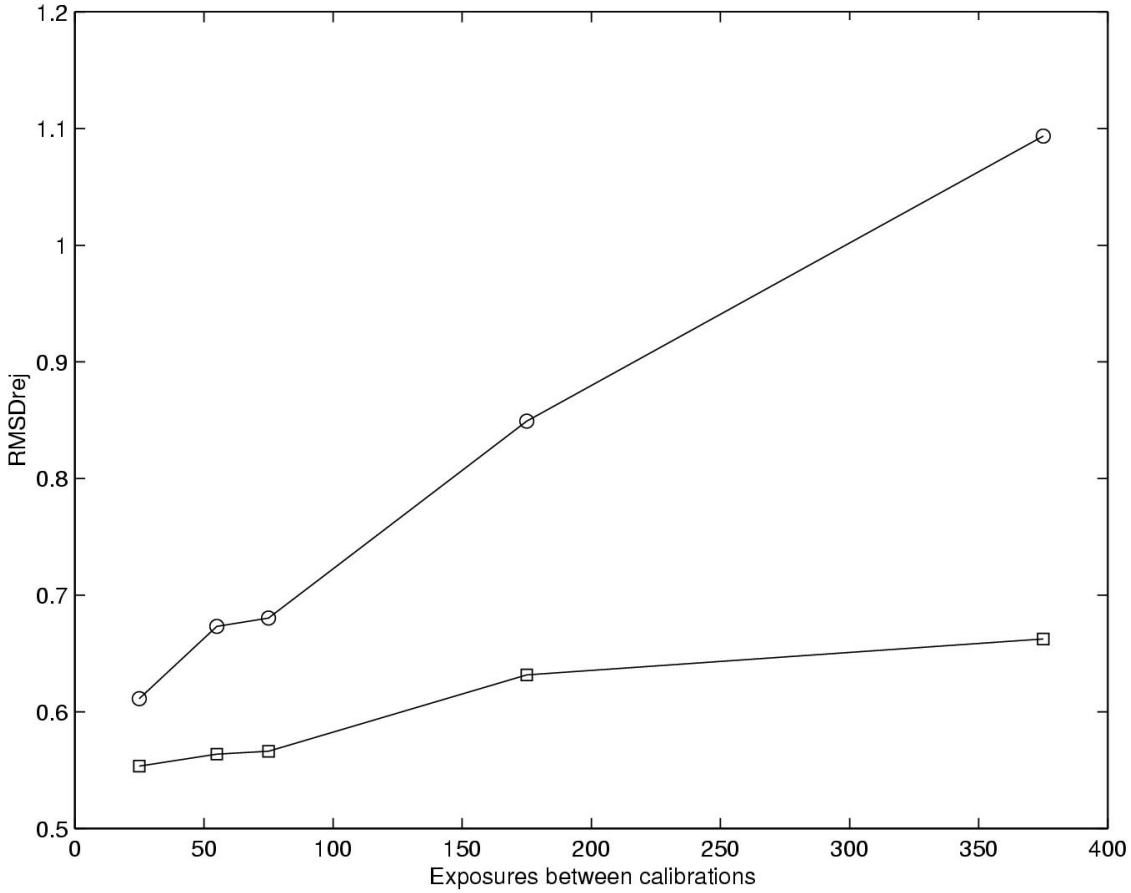
For the binary classification tasks used in this study, re-calibration was deemed necessary when X recent, unclassified test exposures out of the most recent Y tested fell a distance $> Z\sigma$ from both of the training populations. Thus, when the model consistently failed to reliably classify new analyte exposures as either possible analyte, re-calibration must be performed. The parameters X, Y, and Z are variable, with higher values of Y and lower values of X and Z increasing the frequency of re-calibration.

For all further analysis, values of 5 and 10 were used for X and Y, respectively, and 25 exposures were used to recalibrate the FLD model whenever necessary. The analytes n-hexane and n-heptane were treated as unknowns, and n-octane was used as a calibrant. As previously, 200 exposures were used to train the FLD models. The non-cyclic model was allowed to update itself as often as was warranted, provided that the model allowed Y exposures between calibrations to provide for the appropriate statistics to be collected. For comparison, cyclic calibration schemes were evaluated using 25, 55, 75, 175, or 375 exposures between calibrations, with the number of exposures between calibrations constant for a particular experiment. Additionally, to ensure that the cyclic vs. event-driven calibration approaches were directly compared, the values of Z used were adjusted so they produced, on average, the same average number of exposures between calibrations as listed above.

In all cases, the event-driven calibration showed a higher performance and lower $RMSD_{rej}$ values, indicating overall better calibration models. Figure 6.10 depicts a plot of $RMSD_{rej}$ vs. the average number of exposures between calibrations for both the cyclic and event-driven calibrations. When given the choice, the event-driven calibration algorithm chose to calibrate after breaks in data collection and after outliers of the data set.

Naturally, the curves of Figure 6.10 cannot continue to diverge with increasing numbers of exposures between calibrations, as eventually no calibration occurs and poor performance results. Similarly, with very frequent calibration, the two curves will be effectively the same, performing calibration after a very few exposures. Between these two extremes, when an algorithm can hasten or delay calibration significantly, event-driven calibration shows much better performance than cyclic calibration. This is particularly advantageous for data of the sort evaluated in this study, in which the degradation of the FLD model is not easily modeled or predictable (outside of the

Figure 6.10: Root mean square deviation ($RMSD_{rej}$) between calibrated D-values and those determined during training, as a function of the number of use exposures between calibration periods. Cyclic calibration using a cyclic calibration scheme (○) is compared to that using a scheme in which calibration is performed after a certain degree of deterioration of the classification model (□).



obvious breaks between data collection periods), but rather is seemingly random (assuming no knowledge of external events). From the data presented in this experiment, it would be appropriate to assume that calibration would be necessary after a long break between data collection; however, the algorithm was quite capable of determining other situations in which calibration is necessary as well.

The data collected in this study was obtained under relatively controlled laboratory conditions. When used in the field, one might assume that other, less predictable events would tend to dominate, and that breaks in data collection would possibly be more frequent than calibration reasonably could be. Under such circumstances, the recalibration method would be useful because predictions of drift cannot be based heavily on sampling dynamics, but must instead be determined from the behavior of the sensors themselves.

6.5 Conclusions

The detectors used in this study seemed rather immune to drift in all but the most difficult binary classification problems. Even for such separations, drift was typically only problematic after weeks of use. For such challenging separations, a simple, linear method has been developed for correcting sensor drift over time, without significant regard as to the nature or dynamics of the drift. The method is most effective at dealing with large degradations in binary separations that were previously at least somewhat robust.

Because of the lack of a simple, predictive model to estimate future sensor drift with the sensors used in this study, a calibration system was necessary to prevent frequent and complete retraining of the Fisher classifier models. Such a calibration scheme can be effective because a given detector will exhibit similar drift trends to similar analytes. Therefore, if the response of a given detector is known at any given time to a certain analyte, it is possible to estimate how any similar analyte would have responded at any time by simply sampling it once. This phenomenon allows a static model to be used for a dynamic system while we have no fundamental understanding of the dynamics that affect the analyte/sensor response, we can at least measure the dynamics on a known, similar analyte.

Because a suitable calibrant needs to be similar to the analyte it is used to calibrate, this scheme might at first seem daunting, as each binary separation task would appear to require two calibrants, each of which need be similar to one of the two analytes of the task. Without prior knowledge, this requirement might seem more difficult than simply determining the identity of the unknown analytes themselves. However, any separation task that is difficult enough to generate errors will necessarily involve two very similar analytes, as shown above. This means that two such similar analytes can readily share the same calibrant. Additionally, suitable calibrants can be easily

chosen based upon similarities between the unknown analyte signatures and other, stored analyte signatures. This selection of similar analytes is possible despite any sensor drift that might occur, as our sensors do not drift sufficiently to cause similar analytes to appear dissimilar over reasonably long time periods.

Of course, efficient calibration requires that a relatively large database of chemical signatures be collected. Naturally, a classifier model trained to distinguish n-hexane from n-heptane will not be effectively calibrated if hydrocarbon data has never been previously collected. However, such a library database need only be collected once for a given sensor set to be useful over the operation periods of four months evaluated in this study.

Finally, there is no specific determination of how much or how often calibration needs to be performed. Naturally, longer and more frequent is always better. In a tradeoff between frequency and duration, the optimal mixture will likely vary on the sensor set used as well as the ratio of calibration to use exposures employed. It is reasonable to assume, however, that extremely short cycles will not be effective because of the lack of predictive ability of the response variability at short timescales. Similarly, extremely long cycles are not optimal, as a great deal of drift might occur in the meantime. It is best to evaluate the model constantly, and calibrate when it has clearly deteriorated, as the drift/noise in this system is based on specific events more than an elapsed time or number of exposures.

Ultimately, this method provides a simple, linear, computationally manageable method of correcting a previously determined classification model. Additionally, by intelligently determining when calibrations are required, the least possible time can be spent calibrating, allowing for more time spent using the device.

Bibliography

[1] Wang, Y.; Veltkamp, D.; Kowalski, B. *Analytical Chemistry* **1991,** *63,* 2750–2756.

[2] Blank, T.; Sum, S.; Brown, S.; Monfre, S. *Analytical Chemistry* **1996,** *68,* 2987–2995.

[3] Ballantine, D.; Rose, S.; Grate, J.; Wohltjen, H. *Analytical Chemistry* **1986,** *58,* 3058–3066.

[4] Rose-Pehrsson, S.; Grate, J.; Ballantine, D.; Jurs, P. *Analytical Chemistry* **1988,** *60,* 2801–2811.

[5] Patrash, S.; Zelleres, E. *Analytical Chemistry* **1993,** *65,* 2055–2066.

[6] Ronot, C.; Archenault, M.; Gagnaire, H.; Goure, J.; Jaffrezicrenault, N.; Pichery, T. *Sensors and Actuators B - Chemical* **1993,** *11,* 375–381.

[7] Rakow, N.; Suslick, K. *Nature* **2000,** *406,* 710–713.

[8] White, J.; Kauer, J.; Dickinson, T.; Walt, D. *Analytical Chemistry* **1996,** *68,* 2191–2202.

[9] Dickinson, T.; Michael, K.; Kauer, J.; Walt, D. *Analytical Chemistry* **1999,** *71,* 2192–2198.

[10] Dickinson, T.; White, J.; Kauer, J.; Walt, D. *Nature* **1996,** *382,* 697–700.

[11] Bartlett, P.; Archer, P.; Lingchung, S. *Sensors and Actuators* **1989,** *19,* 125–140.

[12] Gardner, J.; Pike, A.; Derooij, N.; Koudelkahep, M.; Clerc, P.; Hierlemann, A.; Gopel, W. *Sensors and Actuators B - Chemical* **1995,** *26,* 135–139.

[13] Shurmer, H.; Gardner, J.; Corcoran, P. *Sensors and Actuators B - Chemical* **1990,** *1,* 256–260.

[14] Freund, M.; Lewis, N. *Proceedings of the National Academy of Sciences of the United States of America* **1995,** *92,* 2652–2656.

[15] Lonergan, M.; Severin, E.; Doleman, B.; Beaber, S.; Grubb, R.; Lewis, N. *Chemistry of Materials* **1996,** *8,* 2298–2312.

[16] Severin, E.; Doleman, B.; Lewis, N. *Analytical Chemistry* **2000,** *72,* 658–668.

[17] Lang, H.; Baller, M.; Berger, R.; Gerber, C.; Gimzewski, J.; Battiston, F.; Fornaro, P.; Ramseyer, J.; Meyer, E.; Guntherodt, H. *Analytica Chimica Acta* **1999,** *393,* 59–65.

[18] Torsi, L.; Dodabalapur, L.; Sabbatini, L.; Zambonin, G. *Sensors and Actuators B - Chemical* **2000,** *67,* 312-316.

[19] Cornila, C.; Hierlemann, A.; Lenggenhager, R.; Malcovati, P.; Baltes, H.; Noetzel, G.; Weimar, U.; Gopel, W. *Sensors and Actuators B - Chemical* **1995,** *25,* 357–361.

[20] Gardner, J.; Shurmer, H.; Corcoran, P. *Sensors and Actuators B - Chemical* **1991,** *4,* 117–121.

[21] Corcoran, P.; Shurmer, H.; Gardner, J. *Sensors and Actuators B - Chemical* **1993,** *15,* 32–37.

[22] Watson, J. *Sensors and Actuators* **1984,** *5,* 29–42.

[23] Yamazoe, N. *Sensors and Actuators B - Chemical* **1991,** *5,* 7–19.

[24] Di Natale, C.; Martinelli, E.; D'Amico, A. *Sensors and Actuators B - Chemical* **2002,** *82,* 158–165.

[25] Artursson, T.; Eklov, T.; Lundstrom, I.; Martensson, P.; Sjostrom, M.; Holmberg, M. *Journal of Chemometrics* **2000,** *14,* 711–723.

[26] Holmberg, M.; Davide, F.; DiNatale, C.; DAmico, A.; Winquist, F.; Lundstrom, I. *Sensors and Actuators B - Chemical* **1997,** *42,* 185–194.

[27] Holmberg, M.; Winquist, F.; Lundstrom, I.; Davide, F.; DiNatale, C.; DAmico, A. *Sensors and Actuators B - Chemical* **1996,** *36,* 528–535.

[28] Marco, S.; Ortega, A.; Pardo, A.; Samitier, J. *IEEE Transactions on Instrumentation and Measurement* **1998,** *47,* 316–321.

[29] DiNatale, C.; Davide, F.; D'Amico, A.; Gopel, W.; Weimar, U. *Sensors and Actuators B - Chemical* **1994,** *19,* 654–657.

[30] Balaban, M.; Korel, F.; Odabasi, A.; Folkes, G. *Sensors and Actuators B - Chemical* **2000,** *71,* 203–211.

[31] Frank, M.; Ulmer, H.; Ruiz, J.; Visani, P.; Weimar, U. *Analytica Chimica Acta* **2001,** *431,* 11–29.

[32] Duda, R.; Hart, P. *Pattern Classification and Scene Analysis;* John Wiley and Sons: New York, 1984.

[33] Vaid, T.; Burl, M.; Lewis, N. *Analytical Chemistry* **2001,** *73,* 321–331.

[34] Koscho, M.; Grubbs, R.; Lewis, N. *Analytical Chemistry* **2002,** *74,* 1307–1315.

[35] Burl, M.; Sisk, B.; Vaid, T.; Lewis, N. *Sensors and Actuators B - Chemical* **2002,** *87,* 130–149.

[36] Sisk, B.; Lewis, N. *Sensors and Actuators B - Chemical* **2003,** *96,* 268–282.