Mining the Detector Responses of a Conducting Polymer Composite-Based Electronic Nose

M.C. Burl[1*], *S. Briglin*[3], *B. Doleman*[3], *A. Hopkins*[3],
A. Matzger[3], *D.N. Ortiz*[3], *A. Schaffer*[3], *S. Upchurch*[1],
T. Vaid[2,3†], *N.S. Lewis*[2‡]

Abstract

Arrays of polymer films embedded with conductive or resistive material have attracted significant attention as "electronic noses". Sorption of a vapor into the polymer films causes physical swelling, which leads to a change in the DC electrical resistance of the film. Like the receptors of the mammalian olfactory system, each polymer-based detector responds to more than one analyte and each analyte elicits a response from more than one detector. The DC resistance across the array of detectors is sampled versus time producing a multivariate time-series. Except when rapid response is critical, this multivariate time-series can be converted into a vector representation by focusing on the steady-state behavior and calculating the relative change in resistance in each channel relative to an initial baseline.

In this paper, we provide an overview of our efforts to mine useful information from electronic nose data. Four case studies are presented: rapid detection of DNT (dinitrotoluene) vapors at very low concentrations (< 1 part-per-billion), pairwise discrimination between chemically similar analytes (e.g., H2O and D2O; heptane and hexane), prediction of human percepts of odor quality from electronic nose detector responses, and classification of vapors into the appropriate chemical family (e.g., saying methanol is an alcohol without having previously smelled methanol).

keywords: electronic nose, scientific datasets, case study/application, pattern recognition, classification, statistical methods

---

[*]1. JPL, M/S 126-347, 4800 Oak Grove Drive, Pasadena, CA 91109, `burl@aig.jpl.nasa.gov`
[†]2. WUSTL, Campus Box 1134, St. Louis, MO 63130, `vaid@wuchem.wustl.edu`
[‡]3. Caltech, M/S 127-72, Pasadena, CA 91125, `nslewis@its.caltech.edu`

1   Introduction

Arrays of polymer films embedded with conductive or resistive material have at-tracted significant attention as "electronic noses." There are numerous potential applications for such devices including environmental monitoring, detection of land mines and explosives, customs (detection of drugs or contraband agricultural prod-ucts), quality control, fragrance and wine evaluation, etc. Unlike traditional "lock-and-key" approaches to vapor sensing, in which a detector is very specific to a par-ticular analyte, the polymer-based detectors used here are broadly-tuned so that a given detector responds to many vapors and a single vapor causes a response in many detectors. (Diversity in the detector responses is achieved by including different insulating organic polymers in the array.) Only by analyzing the pat-tern of responses across the array of detectors can specific analytes be identified or discriminated from chemically similar compounds. In other words, data min-ing techniques, including detection, regression, discrimination, and clustering, must be used to transform the raw data into a semantically meaningful representation ("knowledge").

In this paper, we provide an overview of the electronic nose instrument, the type of data it produces, and our efforts to mine useful information from that data. The approach we have taken is quantitative with numerical attributes (sensor response values) and models at the foundation[1] Our focus will be on four case studies that we have undertaken during the past year including: rapid detection of DNT (dinitrotoluene) vapors at very low concentrations (< 1 part-per-billion) [6], pairwise discrimination between chemically similar analytes (e.g., H2O and D2O; heptane and hexane) [7], prediction of human percepts of odor quality from electronic nose detector responses [8], and classification of vapors into the appropriate chemical family (e.g., saying methanol is an alcohol without having previously been exposed to methanol). Interesting directions for further work are also outlined.

2   Experimental Setup and Data Description

Different instantiations of the electronic nose, e.g., with different numbers or types of polymer films, were used in each of the studies conducted. Despite differences in the arrays, the same principles of operation apply in each case. Sorption of a vapor into the polymer films causes physical swelling, which leads to a change in the DC electrical resistance of the film. The DC resistance across each of the films in the array is sampled at approximately uniformly-spaced sample times. The resistance values are digitized with an A-to-D converter and transferred to a central processing unit for analysis.

All analyte exposures are performed using a computer-controlled vapor gener-ation and control system that regulates the identity, concentration, exposure time, and flow rate of the analyte above the detectors [4]. Between exposures, clean air is passed through the system to remove any residue from the previous exposure.

---

[1]As pointed out by one of the reviewers, there is a different approach involving qualitative reasoning that has been applied with some success to various chemical, physiological, and phar-maceutical problems [1, 2, 3].
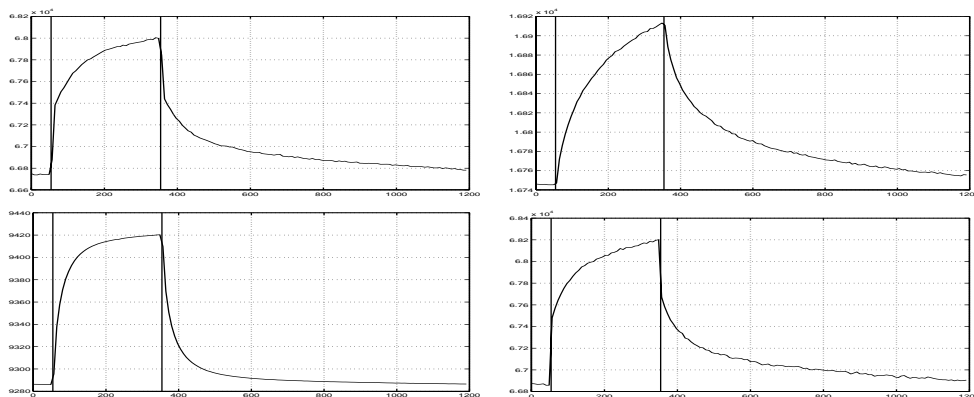
Figure 1. *The raw data produced by the electronic nose consists of a multivariate time-series. Each figure shows the resistance versus time profile in a different polymer channel. During the initial 60s, clean air is flowing across the sensors. At 60s, the analyte is introduced. At 360s, the flow of analyte is stopped and the flow of clean air is resumed.*

Analytes are typically presented to the system in a randomized order to prevent biases in the results.

The raw data produced by the electronic nose consists of sampled resistance versus time for each polymer channel, i.e., a multivariate time-series. Figure 1 shows raw data produced by a few sensors of one array. A typical array may include up to 40 sensors. In the example shown, the sensors were exposed to $60s$ of clean air flow $[0, 60]$, followed by $300s$ of exposure to *methanol* $[60, 360]$, followed by $840s$ of exposure to clean air $[360, 1200]$.

Except when rapid decisions are required, we have typically discarded the temporal information in the multivariate time-series and focused on the equilibrium response (steady state) data. Specifically, the data are reduced to produce a $\frac{\Delta R}{R_b}$ value for each channel, where R_b is the drift-corrected baseline response of the detector during the pre-exposure period immediately before the analyte is introduced into the system and ΔR is the difference between the steady-state response of the detector measured at the end of the analyte exposure period and the drift-corrected baseline. Thus, the raw time-series response of the electronic nose to a given analyte can be converted to a d-dimensional vector where d is the number of channels (polymer films). Note that for rapid detection of a vapor, the steady-state characterization is not useful and the actual time-series must be analyzed. (See, for example, the problem considered in the next section.)

For a broad range of concentrations and analytes, the electronic nose arrays behave like a linear system. Increasing or decreasing the concentration of an analyte produces a proportional increase or decrease in the signature, and the response to mixtures of analytes is approximately the weighted average of the response to the individual analytes [4].

3 Rapid Detection of Low Concentrations of DNT

As part of a DARPA-sponsored program to identify and detect the chemical signature of land mines [], we performed a laboratory experiment to study rapid detection (1-2s) of very low concentrations of dinitrotoluene (DNT) using a conducting polymer composite-based electronic nose. DNT is an impurity that results during the manufacture of military grade TNT. DNT has been shown to accumulate in the soil above buried land mines [5]; hence, detection of DNT is of keen interest as a possible tool for demining [6, 9]. The laboratory apparatus was set up to periodically deliver approximately 5s pulses of DNT to the detectors at a maximum concentration less than 7 parts-per-billion(ppb). The maximum concentration value was based in part on theoretical calculations; actual measurements conducted independently by Sandia National Laboratory showed that the actual delivered concentration was less than 1 ppb.

An electronic nose consisting of eight nominally identical methyloctadecylsiloxane (MODS)-carbon black composite detectors was used to try to detect the DNT pulses. The experimental protocol consisted of one hour of exposure to air only, followed by ten control exposures to 5s DNT pulses spaced every 605s, followed by a randomized sequence of twenty exposures/non-exposures spaced every 605s. The data set was then independently analyzed without knowledge of the ground truth for the randomized sequence of exposures/non-exposures. A control run was also performed to rule out the possibility that responses were obtained due to small changes in the flow rate of gas to the detectors in the exposure/non-exposure cases.

Figure 2 shows the resistance versus time profile computed by averaging over the bank of eight nominally identical MODS detectors. The dark vertical lines show the ground truth of when the DNT puffs were applied. The first ten lines represent the control set. Note that the time axis spans over 6 hours (22e03 s). The series of "bumps" that are visible on this long time scale plot are not related to the DNT pulses and instead represent environmentally-induced oscillations in the baseline resistance of the detector. The DNT-induced behavior occurs on a 5 second time scale that is not discernible on this plot.

The resistance versus time profile from Figure 2 was processed using what is essentially a ramp-sensitive matched filtering algorithm with adaptive background estimation and subtraction. Figure 3 shows the detector output versus time. As in Figure 2, the dark vertical lines show the ground truth of when the DNT puffs were applied. The circles show local maxima of the detector output that exceed a given threshold. All DNT exposures and non-exposures were correctly identified within both the control and randomized sequence. In fact, a much stronger result was obtained: the separability at the detector output was sufficient for all DNT exposures (in both the control set and the randomized set) to be correctly identified with *zero* false alarms over the entire 6 hour duration of the experiment.

The highest detector output value (15.3) occurred at the 6th control sample. The resistance versus time profile for this sample is shown in detail in Figure 4a. The detector output versus time is shown in Figure 4b. The lowest detector output value (7.35) occurred at the 7th control sample. The resistance versus time profile for this sample is shown in detail in Figure 4c. The detector output versus time is
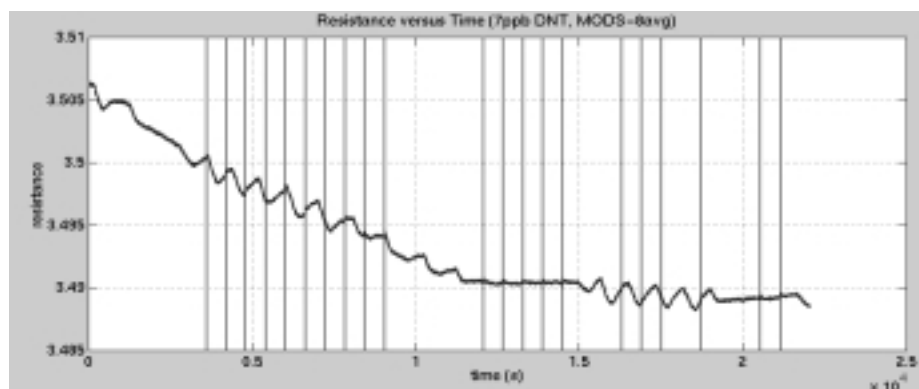
Figure 2. *Resistance versus time profile computed by averaging over the bank of eight nominally identical methyloctadecylsiloxane sensors. Dark vertical lines indicate exposures to 5s DNT puffs.*
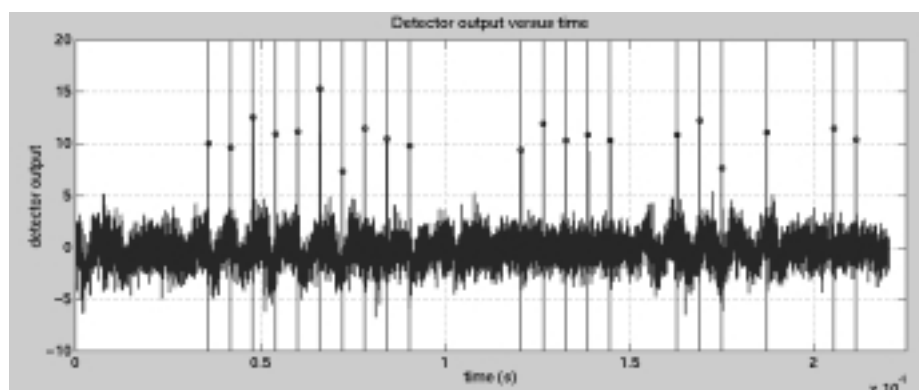

Figure 3. *Detector output versus time over the 6+ hours of the experiment. The dark vertical lines show the ground truth of when DNT puffs were applied.*

shown in Figure 4d. In part, the detector output for this exposure is lower because the decreasing resistance trend prior to the exposure results in a biased estimate of the pre-exposure resistance level.

Although all DNT exposures were perfectly detected with zero false alarms, it is interesting to look at the "close calls" or "near false alarms". In Figure 2 there are four places where the detector output on the background exceeds a threshold of 5 (but still well below the minimum target value of 7.35). The detailed resistance versus time profiles for these four "near false alarms" are shown in Figure 5a-d.
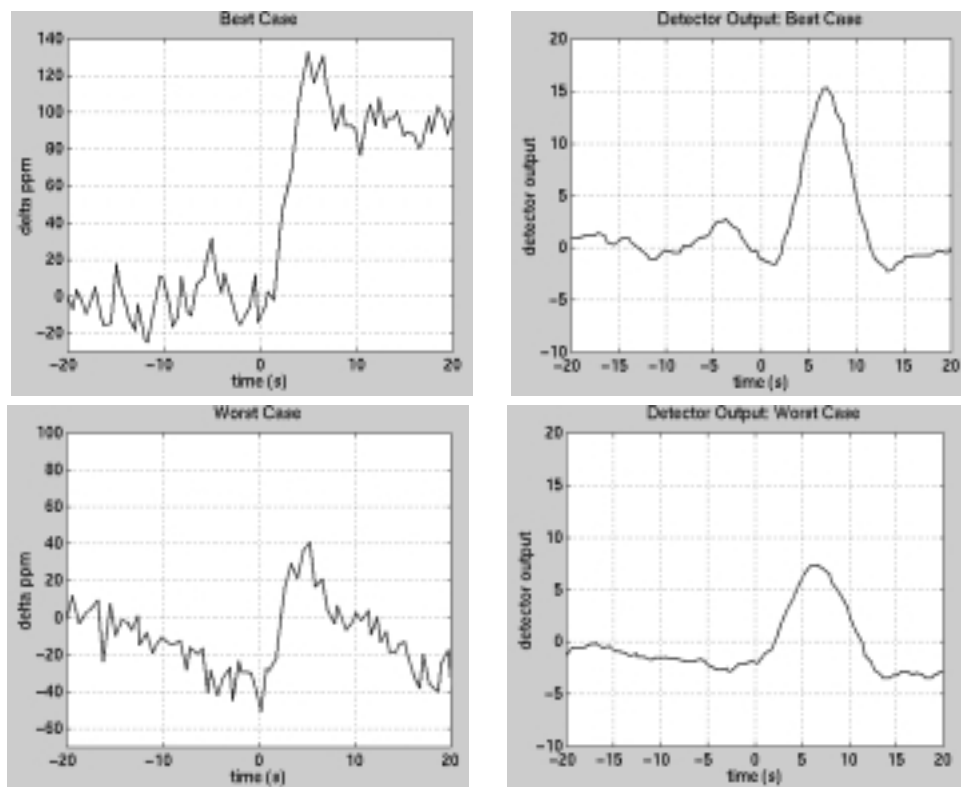
Figure 4. *(a) Resistance (delta ppm) behavior of the DNT exposure that produced the largest detector output value (control sample 6). (b) Detector output value (unitless) versus time. (c) Resistance (delta ppm) behavior of the DNT exposure that produced the smallest detector output value (control sample 7). (d) Detector output value (unitless) versus time.*

4   Discrimination of Chemically Similar Analytes

A second study was conducted to evaluate the performance of various "classical" pattern recognition algorithms on electronic nose data. In previous work, we demonstrated that excellent pairwise discrimination performance could be achieved, even between closely related sets of analytes, using Fisher's Linear Discriminant [10]. In particular, pairs of simple organic vapors were essentially perfectly separated from each other, including structural isomers such as o- and m-xylene [11]. To better understand the relative effectiveness of different discrimination algorithms, it was necessary to formulate several very difficult test cases in which at least some analyte samples were misclassified by at least some of the algorithms. We established three such test cases: (1) discrimination of a pair of compounds that are chemically very similar, H2O and D2O, (2) discrimination of similar analytes (1-propanol vs 2-propanol and n-hexane vs n-heptane) at low concentrations (% of vapor pres-
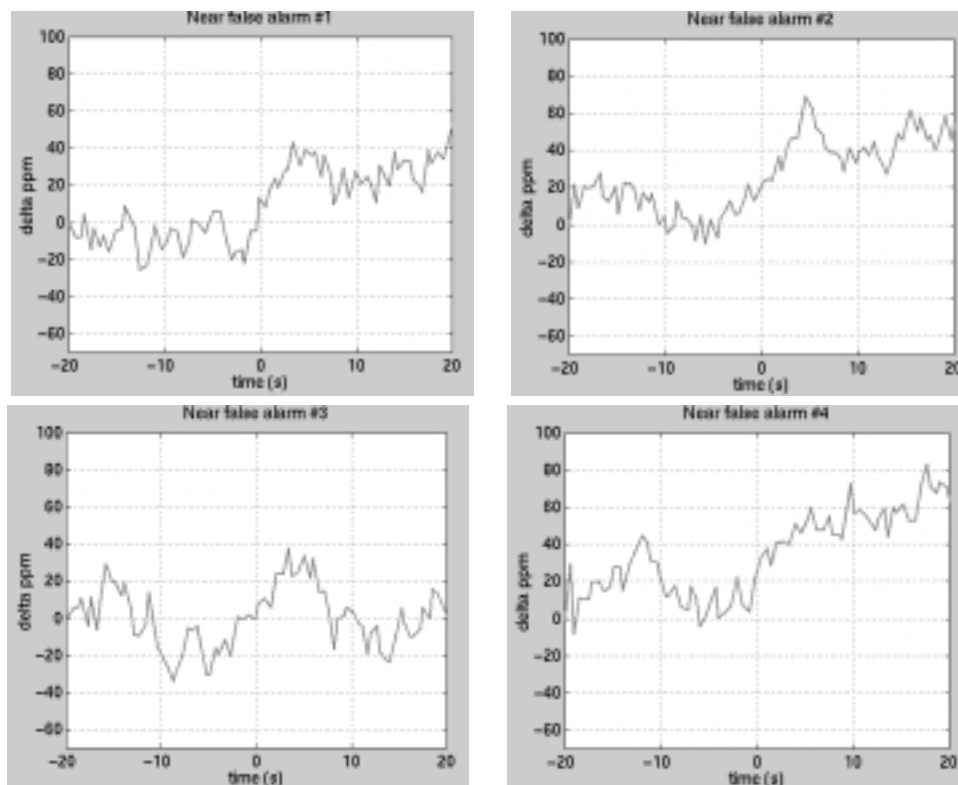
Figure 5. *Resistance (delta ppm) behavior in the four background windows that produced the largest detector output values. The maximum detector output produced over these background samples was slightly above 5, while the lowest value produced for a real DNT exposure was 7.35.*

sure), and (3) discrimination of binary mixtures (of 1-propanol and 2-propanol and of n-hexane and n-heptane) at slightly different mixing ratios.

All exposures were performed for a duration of $300s$, separated by $600s$ of flowing laboratory air. The background air contained 1.10+/- 0.15 parts per thousand of water vapor, but no active auxiliary control over the humidity or over the temperature of the bubblers or the detectors was performed during data collection. Resistance versus time measurements were recorded using an array of 20 carbon-black polymer composite films. The multi-dimensional time-series data was reduced to vector form as described in Section 2. For the H2O vs D2O experiments, 200 exposures in alternating order were recorded. The vapors were diluted to $P/P^0 = 0.050$, where P is the partial pressure of the analyte and P^0 is the vapor pressure of the analyte at room temperature. For the low concentration experiments, a series of 100 exposures each to 1-propanol and 2-propanol were performed, with exposures alternating sequentially between each analyte. All exposures were initially performed at a partial pressure P such that $P/P^0 = 0.01$ in a background of laboratory air.

	kNN	LDA	QDA	RDA	PLS	SIMCA
raw	0.125	0	0	0	0	0.015
normalized	0.370	0	0	0	0	0.005

Table 1. *Leave-one-out cross-validation error rates for H2O vs D2O.*

Similar trials were conducted at partial pressures of $0.0075, 0.005, 0.0025$ with 100 alternating exposures to each analyte. A second data set with the same number of exposures and partial pressures was collected for n-hexane vs n-heptane. For the binary mixture experiments, exposure 1 consisted of a combination of 2-propanol at $P/P^0 = 0.025$ and 1-propanol at $P/P^0 = 0.025$; exposure 2 consisted of 2-propanol at 0.027 and 1-propanol at 0.023; exposure 3 consisted of 2-propanol at 0.021 and 1-propanol at 0.029; exposure 4 consisted of 2-propanol at 0.035 and 1-propanol at 0.015, The series of exposures 1–4 was repeated 100 times, for a total of 400 exposures. A similar data set was collected for mixtures of n-hexane and n-heptane.

The vectorized response data were evaluated using the following algorithms in a leave-one-out cross-validation mode: k-Nearest Neighbor, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Regularized Discriminant Analysis (RDA) [12], Partial Least Squares (PLS), and SIMCA [13, 14]. The kNN, LDA, and QDA methods are generally well-known [15]. RDA is a method proposed by Friedman that allows a smooth transition from LDA to QDA through the use of two regularization parameters that determine how the class-conditional covariance matrices are estimated. PLS is a regression technique but can be adapted to perform binary classification by designating the target value for examples of class 1 as +1 and for examples of class 2 as -1. SIMCA is a method that is popular in the chemometrics community. It involves deriving separate principal components models to describe each class. An unknown point is then classified based on its in-space and out-of-space distance from the principal component subspace of each class.

Tests were performed with both the raw response vectors and normalized vectors. Since the detectors respond linearly to increasing concentration, the L2 normalization removes concentration information from the response vector. The normalized tests correspond to the case when discrimination is desired but the analytes are presented at unknown (and not necessarily the same) concentrations. Results are shown in Tables 1, 2, and 3.

Note that although H2O and D2O have very similar physical properties, there are externally measurable differences, including, for example, boiling point (100 vs $101.4°C$) and melting point (0 vs $3.8°C$). The detectors in the array that were most polar and hydrogen-bonding tended to respond more strongly to H2O than D2O, while the converse was true of the less polar polymers. Examination of the responses of individual detectors indicate that single detectors would perform poorly for this discrimination task, whereas, the array of broadly tuned detectors yields essentially perfect performance. Although the H2O vs D2O test case was selected to challenge the discrimination algorithms, most are able to classify all the examples with zero mistakes.

	kNN	LDA	QDA	RDA	PLS	SIMCA
0.0100	0.000 (0.010)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.025)
0.0075	0.010 (0.015)	0.005 (0.010)	0.005 (0.000)	0.005 (0.000)	0.015 (0.000)	0.020 (0.005)
0.0050	0.410 (0.495)	0.335 (0.410)	0.335 (0.410)	0.255 (0.390)	0.300 (0.495)	0.470 (0.445)
0.0025	0.385 (0.465)	0.435 (0.470)	0.435 (0.470)	0.350 (0.415)	0.380 (0.550)	0.440 (0.515)
0.0100	0.030 (0.010)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.010 (0.005)
0.0075	0.035 (0.065)	0.005 (0.010)	0.010 (0.045)	0.005 (0.010)	0.005 (0.010)	0.065 (0.110)
0.0050	0.020 (0.285)	0.005 (0.180)	0.005 (0.245)	0.005 (0.165)	0.005 (0.160)	0.010 (0.290)
0.0025	0.045 (0.410)	0.005 (0.350)	0.010 (0.320)	0.005 (0.290)	0.005 (0.305)	0.025 (0.360)

Table 2. *Leave-one-out cross-validation error rates for 1-Propanol vs 2-Propanol (top) and n-Hexane vs n-Heptane at Low Concentration. Error rates for normalized data are shown in parenthesis. Note that an error rate of 0.005 is one misclassified example out of 200 attempts (100 examples of each type).*

Δ_{mix}	kNN	LDA	QDA	RDA	PLS	SIMCA
4	0.325 (0.165)	0.030 (0.025)	0.075 (0.080)	0.030 (0.025)	0.030 (0.025)	0.205 (0.145)
8	0.120 (0.105)	0.005 (0.010)	0.015 (0.020)	0.005 (0.010)	0.005 (0.010)	0.190 (0.210)
12	0.065 (0.045)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.100 (0.050)
16	0.010 (0.010)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.010 (0.000)
20	0.005 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.005 (0.000)
28	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
4	0.420 (0.450)	0.225 (0.210)	0.005 (0.005)	0.225 (0.210)	0.230 (0.210)	0.340 (0.365)
8	0.365 (0.295)	0.025 (0.040)	0.010 (0.045)	0.025 (0.010)	0.025 (0.025)	0.085 (0.130)
12	0.300 (0.270)	0.005 (0.000)	0.005 (0.245)	0.005 (0.000)	0.005 (0.005)	0.025 (0.045)
16	0.310 (0.265)	0.005 (0.005)	0.010 (0.320)	0.005 (0.000)	0.005 (0.005)	0.035 (0.025)
20	0.215 (0.135)	0.000 (0.000)	0.010 (0.320)	0.000 (0.000)	0.000 (0.000)	0.005 (0.000)
28	0.065 (0.010)	0.000 (0.000)	0.010 (0.320)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

Table 3. *Leave-one-out cross-validation error rates for compositionally similar analyte mixtures of 1-propanol/2-propanol and n-hexane/n-heptane. The first column corresponds to the percentage difference between the two mixtures. For example, a 50-50 mixture vs a 58-42 mixture would have a Δ_{mix} equal to 8.*

For the low concentration data, LDA and RDA were the best discriminants (average cross-validation error rates of 0.079 for hexane vs heptane) with RDA offering only a slight improvement over LDA. The PLS algorithm had an average error rate of 0.089, followed by QDA. The kNN algorithm had an average error rate of 0.117, probably due to the inappropriateness of the Euclidean distance metric. SIMCA performed the worst with an average error rate of 0.13. The discriminants were more uniform in their performance on normalized data with the same relative ordering (except QDA was more competitive with LDA).

For the binary mixture data, we again found a similar relative ordering with RDA, LDA, and PLS performing best followed by QDA, SIMCA, and kNN. These results probably reflect that the covariance structure of the data is very similar for each of the pairs of analytes considered. If the true class-conditional covariance matrices are indeed the same, then QDA and LDA should perform identically in the asymptotic case in which an infinite number of training examples are avail-

able enabling the class-conditional statistics to be known exactly. However, when the statistics must be estimated from a finite number of samples, LDA may have an advantage by *assuming* the class-conditional covariance matrices are the same, whereas QDA will try to estimate these independently and take advantage of any differences it finds, even if these differences are just variance in the estimates from the small sample size.

5 Prediction of Human Percepts of Odor Quality

Perhaps the ultimate challenge for an artificial nose is to mimic faithfully the mapping of an odorant-induced detector response pattern to the quality of an odor, e.g., its "minty-ness", as perceived by a human. This task is difficult because the human olfactory system is highly nonlinear in many respects. For example, perceived odor intensity is a nonlinear function of analyte concentration [16]. In addition, qualitatively different human percepts are often produced at different concentrations of a given odorant. Cross-adaptation, masking, and other processes involved with human perception of odor mixtures [17] further complicate the signal processing involved in olfaction [18]. A further level of difficulty results because humans are variable genetically in their perception of many odorants [19]. Thus, any static, generically constructed device could at best capture some average human perceptual processes for a representative set of odorants. Of course, this does not rule out the possibility of a "trainable" device that could be tuned to match the perceptual profile of a specific individual; however, developing such a system poses yet another set of challenges.

The specific focus of this study was to investigate whether the responses of an array of vapor-sensitive polymer-based detectors could be used to predict accurately the perceived quality of an odorant as reported by human panelists. Only chemically pure single-component analytes were investigated due to the further complications described above relating to human olfactory perception of odorant mixtures. Data on human perception of odor quality were obtained from tabulations available in the literature [20]. Electronic nose responses, using a 20 polymer array, were collected for a selected subset of the same compounds. The odorants investigated included molecules that are chemically similar but perceived as being different by humans, as well as molecules that are chemically quite distinct from each other.

Twenty-one odorants were investigated in this work. All chemicals were obtained form Aldrich Chemical Corp. and were used as received. Sets of chemically homologous odorants (for example, a series of straight-chain alcohols, a series of aliphatic esters, a series of straight-chain aliphatic acids, a series of benzene derivatives, etc.) were chosen such that the odors were associated with common, but not identical, human odor descriptors within a set and between sets of odorants. Odorant exposures consisted of $300s$ of pre-exposure to clean air, followed by $300s$ of odorant exposure, followed by $300s$ of clean air. The resulting multivariate time-series of resistance measurements was reduced to vector form as described earlier.

Perceptual odor quality values for humans were obtained from Dravnieks' *Atlas of Odor Character Profiles* [20]. In Dravnieks' study, over 100 people of both

sexes, spanning a wide range of ages, and including smokers as well as non-smokers, were evaluated. The rationale for using such a diverse group was to insure that the reported percepts would be consistent with those of the population at large. Each participant was asked to smell a collection of odorants and was asked to assign a score from 0 through 5 to each of 146 different descriptors (adjectives) that are used in the English language to describe odors. For example, a panelist could give an odorant a score of 3 in the "etherish, anesthetic" category, a 4 in the "minty" category, and 0 in each of the remaining categories. Scores are intended to reflect the degree to which the panelist believes that a descriptor is appropriate for a given odorant, with a value of 0 meaning not appropriate and 5 meaning very appropriate. Of the 146 descriptors used by Dravnieks, we focused on just 17 based on the frequency and extent to which they were used by the panelists to describe our selected set of test odorants.

For the purposes of our study, a limitation with Dravnieks' data is that only averages across the entire group of panelists are provided, so scores for individual participants are not available. Also, the variance (or the distribution) of scores given to a particular odorant-descriptor pair was not reported. Instead, the available data for each odorant-descriptor pair consist of two quantities: percentage of usage (percentage of people who gave a non-zero score level) and average score-level. The goal then is to assess the degree to which the response of the electronic nose can quantitatively predict the usage or score level that would be provided on average by a group of human panelists for each of the seventeen descriptors.

We explored a variety of prediction algorithms including: standard linear regression, nearest-neighbor prediction, ridge regression, principal components regression, partial least squares regression, and regression based on feature subsets. Tests were performed using a cross-validation paradigm in which one chemical was withheld for testing and the other $N-1$ were used for training. This process was repeated with each chemical serving a turn as the holdout. Predictions for each of the holdouts were then compared with the human data yielding a cross-validated correlation score. For specific descriptors, some models provided cross-validated predictions that correlated well with the human data (above the 0.60 level), but none of the models could accurately predict the human values for more than a few descriptors.

The models based on feature subset selection were especially intriguing. For most of the descriptors, relatively small subsets of detectors (e.g., 2-6 detectors) existed that enabled accurate prediction of the human data. However, it was not possible with the amount of data available to identify these subsets through a rigorous model selection procedure in which a separate cross-validation was used to select the feature subset and estimate the regression parameters, which would then be applied to predict a completely separate hold out example. Thus, at this point, we have concluded that human perception of odor quality cannot be reliably predicted from the electronic nose signals with the methods considered. There are several possible explanations for this result. First, the number and diversity of receptors in the artificial nose was significantly (1-2 orders of magnitude) more limited than in humans. Second, linear and affine regression models may be too simplistic to enable human percepts of even single component organic vapor odorants to be predicted

well from conducting polymer composite electronic nose signals. This may reflect the fact that these data models probably have no physically-based relationship to the neuronal connections and signal processing involved in olfactory perception. A third possibility is that the form of the data models could be adequate for the limited task considered, but the parameters of these models could not be estimated reliably enough from the amount of data available. A fourth (related) possibility is that the feature subset selection procedure did not have enough data to identify feature subsets that would provide good generalization ability.

In any case, it is clear that the situation would be significantly worse for odorants that are mixtures of pure compounds, because human odor perception of mixtures is often not linearly related to the mole fraction of the individual components of the mixture. Similarly, for mixtures of odorants that contain aroma-active compounds, which are detectable at very low concentration levels in the human nose, but which would not even produce signals above the noise level of the current conducting polymer detectors, no correlation would be expected between the electronic nose responses and human perception.

6   Classification of Analytes According to Functional Group

Another interesting study which we have recently undertaken (only partially completed) involves classifying analytes according to their functional group based on electronic nose responses. Chemists have grouped analytes into a number of "families" or functional groups based on similarities in physical structure or externally-observable properties. The goal in this study was to determine whether a previously unseen (unsmelled?) compound could be classified into the proper functional group based on the responses of an electronic nose. Five chemical families (alcohols, alkyl halides, aromatics, hydrocarbons, and esters) were selected. Within each class, 15 members were chosen for a total of 75 different chemicals. These were presented to an electronic nose containing 40 polymer-based sensors (two copies of 20 different polymers). The analytes were presented to the nose in groups of eight because the physical setup of the gas dispensation system has eight bubblers. A total of 80 sniffs (ten of each analyte) in randomized order was obtained from each set of eight and then a new set of eight analytes was swapped in. Each group of eight analytes contained two members of four classes so that temporal effects would not bias the results (e.g., if all alcohols were sniffed in the morning and the temperature was cooler then, it might introduce an artificial bias into the separability of alcohols from the other classes).

Each sniff produced a multivariate time series which was converted to vector form. Responses of polymer "twins" (duplicates of the same polymer type) were averaged to produce 20-dimensional vectors for each sniff. Classification into functional groups was performed using a 1-nearest neighbor algorithm. Euclidean distances of the test example from all members of the reference library were computed. The functional class label of the closest member of the reference library was taken to be the class label of the test example. *For a given test example, all other*

	alcohol	alkyl halide	aromatic	hydrocarbon	ester
alcohol	1.000	0.000	0.000	0.000	0.000
alkyl halide	0.000	0.831	0.069	0.006	0.094
aromatic	0.000	0.267	0.527	0.200	0.007
hydrocarbon	0.037	0.019	0.213	0.688	0.044
ester	0.047	0.147	0.000	0.018	0.788

Table 4. *Confusion matrix for 1NN classification of compounds into functional classes. Each functional class contained fifteen compounds with 10 or in a few cases 20 sniffs each. For a given test example, all other examples of the same compound were withheld from the reference library.*

sniffs of the same compound were sequestered from the reference library. In other words, we wanted to determine if a sniff of methanol could be used to classify it as an alcohol *without* having previously smelled methanol, but perhaps having smelled ethanol, butanol, cyclopentanol, etc. A confusion matrix showing the results of this experiment is given in Table 4. The first row shows that all members of the alcohol family were correctly classified as alcohols. The second row shows that 83% of the members of the alkyl halide family were correctly classified, with 6.9% of the members confused as aromatics, 0.6% confused as hydrocarbons, and 9.4% confused as esters. Overall, the average correct classification percentage is 77%.

The fact that the performance was not closer to 100% was at first disappointing. However, after consulting with the domain experts, we realized that many of the compounds are not strictly members of one and only one functional group. For example, 1-chlorobenzene can be considered to be a member of both the halide group and the aromatic group. We are now considering ways in which this non-exclusive class information can be incorporated into the evaluation (and possibly training depending on the method used).

We are also looking into stronger methods for performing the functional classification including support vector machines (SVM) [21, 22] and pairwise Fisher linear discriminants. Gaussian mixture models also seem to be a natural choice, but the dimensionality of the data relative to the number of sniffs of each compound (typically only 10) probably forces a dimensionality reduction step (e.g., PCA) to be performed first to put the data into a manageable space. Support vector machines, by virtue of their maximum margin properties, may be able to generalize directly from the original (sensor-space) data.

7   Conclusion and Future Work

The responses of arrays of polymer composite-based detectors have provided a rich set of possibilities for mining useful information from raw data. In addition, to the four studies presented here, we are conducting a number of additional investigations. It is known that the polymer sensors exhibit aging effects over time leading to reduced responsiveness. Data analysis studies to characterize and perhaps compensate for these changes are underway. As one example, we are exploring ways to

adapt classification rules or decision boundaries learned with data taken with fresh sensors to the case where the sensors have degraded over time. We are also looking to understand the shorter-term environmentally induced effects such as those exhibited in the DNT study of Section 3. The dataset collected to study supervised classification of compounds into functional groups also offers significant possibilities for identification of specific compounds and unsupervised clustering studies to determine whether the classes induced are the same as those used by chemists (e.g., alcohols, aromatics).

Acknowledgments

Bibliography

[1] D. Hau and E. Coiera, "Learning Qualitative Properties of Dynamic Systems", *Machine Learning*, vol. 26, no. 2-3, pp. 177-211, (1997).

[2] S. Muggleton, "Scientific Knowledge Discovery Using Inductive Logic Programming", *Communications of the ACM*, vol. 42, no. 11, pp. 42-46, (1999).

[3] R. Valdes-Perez, "Discovery Tools for Scientific Applications", *Communications of the ACM*, vol 42, no. 11, pp. 37-41, (1999).

[4] E.J. Severin, B.J. Doleman, N.S. Lewis, "An investigation of the concentration dependence and response to analyte mixtures of carbon black-insulating organic polymer composite vapor detectors", *Anal. Chem.*, 72, 658-668 (2000).

[5] V. George, T.F. Jenkins, D.C. Leggett, J.H. Cragin, J.M. Phelan, J.C. Oxley, J. Pennington, "Progress on Determining the Vapor Signature of a Buried Landmine", Proc. of SPIE Aerosense Conf. Session on Detection of Mines and Mine-like Targets, (2000)

[6] S. Briglin, M.C. Burl, M.S. Freund, N.S. Lewis, D.N. Ortiz, and P. Tokumaru, "Progress in Use of Carbon Black-Polymer Composite Vapor Detector Arrays for Land Mine Detection", Proc. of SPIE Aerosense Conf. Session on Detection of Mines and Mine-like Targets, (2000)

[7] T.P. Vaid, M.C. Burl, N.S. Lewis, "Comparison of the Performance of Different Discriminant Algorithms in Analyte Discrimination Tasks Using an Array of Carbon Black-Polymer Composite Vapor Detectors", *Analytical Chemistry*, 73(2), pp. 321-331, (2001).

[8] M.C. Burl, B.J. Doleman, A. Schaffer, and N.S. Lewis, "Assessing the Ability to Predict Human Percepts of Odor Quality from the Detector Responses of a Conducting Polymer Composite-Based Electronic Nose", *Sensors and Actuators B - Chemical*, (72)2, pp. 149-159, (2001).

[9] Proc. of SPIE Aerosense Conf. Session on Detection of Mines and Mine-like Targets, Orlando, FL, (2000).

[10] B.J. Doleman, M.C. Lonergan, E.J. Severin, T.P. Vaid, N.S. Lewis, *Anal. Chem.*, 70, 4177 (1998).

[11] T.P. Vaid, N.S. Lewis, unpublished results (2000).

[12] J.H. Friedman, *J. Am. Stat. Assoc.*, 84, 165, (1989).

[13] S. Wold, *Pattern Recognition*, 8, 127 (1976).

[14] S. Wold, M. Sjostrom, In *Chemometrics: Theory and Application*, p. 243, (1977).

[15] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, Inc. (1973).

[16] F.T. Schiet, W.S. Cain, "Odor intensity of mixed and unmixed stimuli under environmentally realistic conditions", *Perception* 19, 123-132 (1990).

[17] D.G. Laing, H. Panhuber, M.E. Wilcox, E.A. Pittman, "Quality and intensity of binary odor mixtures", *Physiol. Behav* 33, 309-319 (1984)

[18] J.D. Pierce Jr., X.N. Zeng, E.V. Aronov, G. Preti, C.J. Wysocki, "Cross-adaptation of sweaty-smelling 3-methyl-2-hexenoic acid by a structurally similar, pleasant-smelling odorant" *Chem. Senses* 20, 401-411, (1995).

[19] S. Ayabe-Kanamura, I. Schicker, M. Laska, R. Hudson, H. Distel, T. Kobayakawa, S. Saito, "Differences in perception of everyday odoers: a Japanese-German cross-cultural study", *Chem. Senses* 23, 31-38, (1998).

[20] A. Dravnieks, *Atlas of Odor Character Profiles*, American Society for Testing and Materials, Baltimore, MD, (1985).

[21] V.N. Vapnik, *The Nature of Statistical Learning*, Springer, New York, 1995.

[22] C.J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Knowledge Discovery and Data Mining*, 2(2), (1998).