

Nonlinear Estimation for Arrays of Chemical Sensors

Jason Yosinski^a and Randy Paffenroth^a

^aNumerica Corporation, 4850 Hahns Peak Drive Suite 200, Loveland, CO 80538, USA

ABSTRACT

Reliable detection of hazardous materials is a fundamental requirement of any national security program. Such materials can take a wide range of forms including metals, radioisotopes, volatile organic compounds, and biological contaminants. In particular, detection of hazardous materials in highly challenging conditions — such as in cluttered ambient environments, where complex collections of analytes are present, and with sensors lacking specificity for the analytes of interest — is an important part of a robust security infrastructure. Sophisticated single sensor systems provide good specificity for a limited set of analytes but often have cumbersome hardware and environmental requirements. On the other hand, simple, broadly responsive sensors are easily fabricated and efficiently deployed, but such sensors individually have neither the specificity nor the selectivity to address analyte differentiation in challenging environments. However, arrays of broadly responsive sensors can provide much of the sensitivity and selectivity of sophisticated sensors but without the substantial hardware overhead. Unfortunately, arrays of simple sensors are not without their challenges — the selectivity of such arrays can only be realized if the data is first distilled using highly advanced signal processing algorithms. In this paper we will demonstrate how the use of powerful estimation algorithms, based on those commonly used within the target tracking community, can be extended to the chemical detection arena. Herein our focus is on algorithms that not only provide accurate estimates of the mixture of analytes in a sample, but also provide robust measures of ambiguity, such as covariances.

Keywords: chemical detection, electronic olfaction, Raman spectrograph, maximum likelihood estimation, Levenberg-Marquardt

1. INTRODUCTION

Reliable detection of hazardous materials is a fundamental part of any national and theater security infrastructure. Such materials can take a wide range of forms including metals, radioisotopes, volatile organic compounds, and biological contaminants.¹ In particular, detection of hazardous materials in highly challenging conditions — such as cluttered ambient environments, complex collections of analytes, and sensors lacking specificity for the analytes of interest — is an important part of a robust security infrastructure.

Sophisticated single sensor systems (such as mass spectrometers and gas chromatographers) provide good specificity for a limited set of analytes but often have cumbersome hardware and environmental requirements, such as sensor size and cost and environment temperature and humidity, to name but a few. Simple, broadly responsive sensors are easily fabricated, but such sensors individually have neither the specificity nor the selectivity to address analyte differentiation in challenging environments. In an effort to combine the selectivity of sophisticated sensor systems with the ease of use of simple sensors, there has been significant work within the chemical analysis community revolving around the use of *arrays* of simple sensors. Such arrays can provide much of the sensitivity and selectivity of sophisticated sensors but without the substantial hardware overhead. For example, an area in which such sensors arrays might find use is in first responder applications, where size, weight, cost, and ease of use are paramount. Unfortunately, arrays of simple sensors are not without their challenges. In particular, the selectivity of such arrays can only be realized if the data is first distilled using highly advanced *signal processing* and *data fusion* algorithms.

There is a vast extant literature on sensor arrays, and several review articles^{2,3} give an entryway into the literature. The modern work on such sensor arrays began with sensor arrays containing broadly reactive elements that were combined with signal processing/pattern detection algorithms to create a system for analyte detection.⁴ Such ideas are often referred to using the term *electronic nose*, with more modern references referring to *electronic olfaction*, to emphasize the connection between the work in electronic sensors arrays and the mammalian olfactory

Further author information: (Send correspondence to R.P.)

J.Y.: E-mail: jason.yosinski@numerica.us, Telephone: +1 970 461 2000 x 256

R.P.: E-mail: randy.paffenroth@numerica.us, Telephone: +1 970 461 2000 x 233

system. The over 1000 olfactory genes in humans (not to mention the greater than 100 million olfactory cells in a canine) are a natural example of the power and complexity of chemical sensor arrays.³

Several mathematical approaches have been applied to processing data from sensor arrays, and reviews of the state of the art are available.⁵ The current methods range from principal component analysis, Fischer linear discriminants, quadratic discriminants clustering techniques, and other standard mathematical algorithms, to artificial intelligence approaches including neural networks and fuzzy logic. In effect, many of the current ideas are drawn from “Pattern Classification”⁶ and these techniques have been applied from a more “engineering” point of view. While these techniques are certainly applicable to a subset of the problems of interest, they are not ubiquitous in their effectiveness. We of course hope to leverage past pattern classification work to the largest extent possible, but our novel contribution has come from a somewhat different direction.

It is critical to note that chemical sensor arrays achieve their effectiveness by way of data fusion, just as distributed networks of radar, infra-red, acoustic, etc. sensors are more effective than individual sensor in target tracking problems.⁷⁻⁹ Of course, there are important differences between the two domains. For example, chemical sensor arrays have classically been colocated, though we observe that modern chemical sensing systems are moving toward spatially distributed scenarios as well. Our goal in this article is to highlight the subtle interaction between the chemical detection and target tracking domains, and demonstrate how they may be combined to the benefit of both with a particular focus on the correct handling of *ambiguity*. While the focus of this article is on colocated sensors, we note that the same approach is applicable to distributed sensors, which we will discuss in a coda at the end of the paper.

Our approach toward the multidimensional chemistry problem stems from Numerica’s experience in the world of target tracking. We discuss our initial thoughts and first steps towards this novel combination of approaches in Section 2 and offer several papers¹⁰⁻¹³ as examples of the exact type of synergy that Numerica hopes to leverage between chemical detection and target tracking. In Section 3 we will discuss the particular data for the current analysis, and in Section 4 we derive models of sensor responses which include proper treatment of ambiguity. Using those models, we demonstrate discrimination algorithms in Section 5 and Section 6 that use not only the derived estimates but also their covariances to substantially improve the performance of the detection algorithms. We close in Section 7 with a brief discussion of the issues arising for more complicated and distributed sensor modalities.

2. MAXIMUM LIKELIHOOD ESTIMATION AND CHEMICAL ANALYSIS

Given an array of simple sensors, one could imagine posing the following non-linear optimization problem: from a set of measurements and library of sensor responses, often called “analyte exemplars”, one derives an *expectation maximization* problem and solves for the concentration of the various exemplars. In other words, one computes the concentrations of the exemplars that are most likely to have given rise to the observed data. Of course, posing the problem in such a way is impractical in all but the simplest of cases — one almost never has a complete library of exemplars and their mutual interactions. Even given a perfect library and a perfect understanding of the underlying expectation maximization problem, one will almost certainly have difficulty solving the full non-linear problem because of the size of the parameter space and the presence of local minima. Many of the algorithms commonly used for multi-dimensional chemical analysis^{5,6} are various forms of approximation to this underlying intractable problem.

The stochastic state estimation, signal processing, and data fusion regimes are examples of the ubiquitous nature of maximum likelihood estimation. These regimes make use of several approximations to *precisely* the same underlying problem. The approximations espouse a slightly different view of the underlying estimation problem and are promising candidates for application to multi-dimensional chemical analysis. However, there are several issues that hinder the swift application of such algorithms to the chemical detection realm. These issues and their possible solutions are precisely the topic of this paper.

2.1 Initial Impressions

As previously mentioned the chemical and biological detection problem can be posed as a maximum likelihood estimation (MLE) problem. In its most generic form, the MLE problem poses the question “Which input to a given system would most likely produce the observed output?”. In the field of target tracking, the given system is a set of tracking sensors such as radars. The input to the system is the set of target states, and the system generates output measurements. Target tracking entails inverting this state-to-measurement operator to determine the target state that most likely produced the observed measurements. In the field of chemical

detection, the system is a set of chemical sensors. Similarly to the tracking problem, the input to the system is a list of chemicals and their concentration, perhaps coupled with temperature, pressure, or other information, and the system generates output measurements. The problem is to determine which chemicals most likely produced the observed sensor data. Note that in both cases, the sensors are stochastic, i.e., a given input does not always produce the same output; rather, the output is polluted by some amount of random noise. This noise results from many phenomena including imperfect sensors and unknown parameters such as thermal effects, to name but a few.

2.1.1 Benefits of an Estimation View of Chemical Analysis

Fortunately, if the chemical problem can be posed as a maximum likelihood estimation problem, we may draw on previous research in the well developed field of estimation, which addresses the following important questions:

Given the maximum likelihood estimate for a parameter, what is the covariance of the estimate?

In the chemical case, this means we may ask our estimator not only for its estimate of the concentration of a given chemical, but also for the covariance of that estimate. The covariance provides information on the deviation of the true value from expected estimate. This feature is often sorely lacking in more traditional “pattern classification” approaches. For example an artificial neural network (ANN) is often employed to differentiate between different classes, but the ANN typically will not be able to report its confidence in addition to the answer. Conversely, an estimator that returns a mean and a covariance can provide the user statistically rigorous information for making downstream decisions.

Does representing the estimate as a mean and covariance suffice to capture the underlying probability density function? There is an underlying assumption in the above question that bears carefully analysis in the chemical detection domain. Namely, the first question assumes that a Gaussian model is appropriate for the representation of the ambiguity in the estimation problem. Even if the computed mean and covariance are the correct first two cumulants/moments of the underlying distribution, there is no guarantee that the higher order cumulants/moments are negligible.

For a given estimation problem, what is the most accurate estimator possible? The Cramér-Rao inequality specifies a lower bound on the variance of *any* estimator. It is not always possible to design an estimator that meets this bound; however, if an estimator *does* meet the Cramér-Rao bound for a given sensor and parametrized model, it is considered *efficient*. In this context, an estimator is efficient if a more accurate estimator is impossible. While space does not allow an extensive analysis of this topic in the current text, one can find such details in another paper.¹⁴

We emphasize the utility that such research could present to the field of chemical detection. Evaluating the efficiency of current estimators would shed light on whether current detection is limited by sensor hardware or software, and research could be directed appropriately. Once an estimator is designed that achieves the Cramér-Rao Bound for a given sensor, that estimator design may be concluded a success, and efforts to further improve accuracy may be focused elsewhere, for example, on the sensor hardware.¹⁴

2.1.2 Extensions to Estimation

There is a natural division of the target tracking problem along two axes of difficulty: number of sensors and number of targets. The simplest case is, of course, a single target being tracked by a single sensor. This simple case is fairly easy to solve — assuming predictable target motion and a linear sensor with white, Gaussian noise — using a standard linear least squares signal processing approach either by batch processing or recursively by way of a Kalman filter. In a chemical context, this would be a similar problem to estimating the concentration of a single analyte given a single, well understood sensor, perhaps a mass spectrometer (MS). Unfortunately, such simple scenarios are hardly commonplace. The field of target tracking then consists largely of the study of a number of techniques one may use to extend this solution to cover realistic scenarios. Furthermore, because sensors are often nonlinear, we use enhanced versions of the Kalman filter, such as the Extended Kalman filter, the Unscented Kalman filter, or the Particle filter, to solve efficiently the corresponding nonlinear least squares problem.

When multiple targets are present, but we still have but a single sensor, one may employ an N-dimensional assignment solver to determine which measurements returned by the sensor are associated with each target. Translated into a chemical context, this might correspond to a mass spectrometer tasked with detecting several analytes, of which some, all, or none may be present. In this instance an assignment solver would be required for determining which peaks (or which fractions of a peak) could be assigned to a given compound.

When tracking with multiple sensors, one must address sensor *fusion* issues as sensor bias and sensor drift (temporally dependent bias), among many others. An important issue here is that of network complexities. If many *distributed* sensors are used, the data may be too burdensome to transmit in full on the network; accordingly, advanced algorithms may be used to ensure that all sensors on the network are able to communicate effectively while using minimal bandwidth.¹⁵ In a chemical detection context, these techniques may allow improved performance when using multiple sensors that, for example, are subjected to differing environmental conditions (temperature, humidity, etc.), causing the measurements to drift independently of each other.

Finally, when we combine multiple sensor with multiple targets, we must integrate all of the solutions presented above. The combination of bias mitigation and network constraints with multi-frame assignment often entails both improved performance as well as additional complexity.

We hope that we have convinced the reader of at least some measure of connection between chemical detection problems and target tracking algorithms. In the rest of this article we expound on that connection and demonstrate how the combination of the two fields leads to substantially improved chemical detection performance.

2.2 Signal Processing

Several obstacles await those who would implement a maximum likelihood estimator for chemical detection. Perhaps the most difficult step in implementing an MLE chemical detector is characterizing the sensor, that is, defining a probability density function (pdf) on the sensor’s output space for each point in its input space.

In the target tracking world, many sensors are well characterized by a Gaussian pdf in their measurement space. While it is certainly the case that the actual performance of the sensor is far more complicated, a Gaussian assumption in the measurement space of the sensor is a good approximation for many of the sensors used in target tracking. This is to say that a radar can be thought of as a machine that performs two operations on its input: it first transforms the input point from input space into output space, and then it adds approximately zero-mean Gaussian noise to the transformed point in output space. As an example a 3D radar might transform a target’s true location in (X, Y, Z) coordinates into a measurement in (range, bearing, azimuth) coordinates and then add Gaussian noise to the reported (range, bearing, azimuth) measurement. This mechanism can be written as follows:

$$\mathbf{z} = h(\mathbf{x}) + \mathbf{v} \tag{1}$$

where \mathbf{z} is the noisy measurement, \mathbf{x} is the point in input space, h is the (possibly nonlinear) mapping function from input space to output space, and \mathbf{v} is the Gaussian distributed noise with some covariance \mathbf{R} :

$$\mathbf{v} \sim N(\mathbf{0}, \mathbf{R}) \tag{2}$$

For some chemical sensors, this characterization may be straightforward, but for others, the characterization is difficult. We present the mass spectrometer as an example of a sensor which may be easy to characterize. For a given input analyte to a mass spectrometer, we consider the output to be a vector containing the entire mass spectrum. We typically expect that for a pure sample, this vector will contain a relatively small number of large values; these large values correspond to the peaks for the charge-to-mass ratios that were most prevalent. Data collected over many runs would show slight variations in peak height and peak location, and it is likely that these variations could be well characterized by Equation 1 and Equation 2.

In such circumstances the sensor characterization is complete once we can come up with an appropriate mapping function h and measurement covariance \mathbf{R} for each point in the input space. For some sensors this mapping can be derived through first principles (e.g., deriving the relative frequencies of differing charge-to-mass ratios in a mass spectrometer), while other sensors exhibit behavior too complex to be modeled by first principles (e.g., stochastically distributed tin oxide nanowire resistors).

Equation 1, for some appropriate h , will be suitable to describe any type of sensor; however, the h may be difficult to obtain. On the other hand, Equation 2 may *not* be a good representation for the noise of a given sensor, in which case another noise model must be used. Clearly many of the algorithms derived herein require representations for h and \mathbf{v} for particular sensors of interest. Accordingly a key focus of our work has been on deriving models for which Equation 2 at least roughly holds and that work can be found in Section 4. A somewhat more subtle point is determining whether a particular h well models a sensor of interest. Clearly an improperly defined h has profound effect on any algorithms that use such an h . One of the most important facets of our work has been the detection of sensors for which a particular h and \mathbf{v} are a good model of the sensor’s performance and for which it is not.

3. ELECTRONIC OLFACTION DATA

There are a vast number of different chemical sensors on which one could imagine applying the techniques from Section 2. Examples of such sensors include tandem mass spectrometers, chemresistors, ion mobility spectrometers (IMS), differential ion mobility spectrometers (DMS), semi-conductor metal oxide sensors, and surface acoustic wave sensors.

For this paper we focus our attention on a data set produced from an array of 512 chemresistors. Our particular set of sensors is challenging both because of their slow response time as well as their particularly noisy nature, as we see in Section 5. These data sets covers a variety of environmental changes over a 616 minute test sequence and is particularly interesting to us for several reasons. First, it displays striking non-linearities that are very reminiscent of some previous work of ours having to do with non-linear maximum likelihood estimators.¹⁴ Second, the time dependence of the measurement morphology, sampled over a long period of time, appears to be something of which we can take advantage. A canonical time sequence from this data set can be seen in Figure 1. Of particular interest is the fact that two regions that have same underlying environment can have substantially different morphology, depending on the initial state when the environment changes. This dichotomy between environment and state is the focus of Section 4.

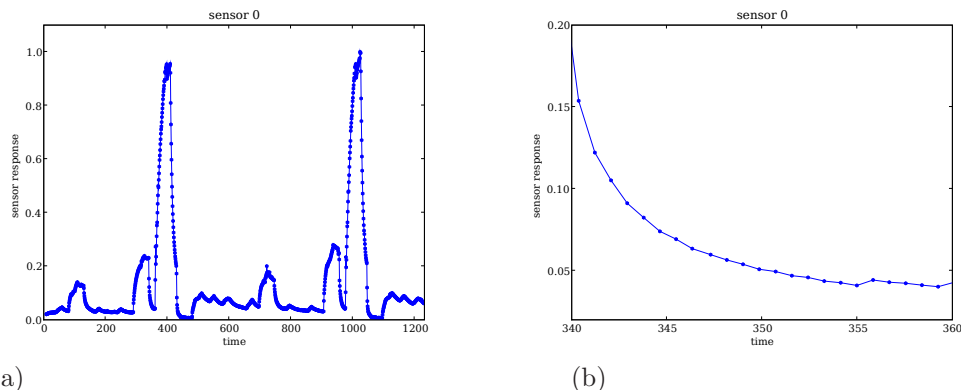


Figure 1. The left figure show a canonical time sequence from the 512-sensor data set. This is a plot from a single sensor over the time period covering the first 1232 minutes of data collection. Note that the data from between zero and 616 minutes and the data between 616 minutes and 1232 minutes look very similar; this is because the data was gathered twice in identical manners. The right figure shows an example of the exponential decay in the measurements. For a region in which the environment is not changing, the measurements are decaying to some fixed value related to the contents of the environment, plus noise. Accordingly, an exponentially decaying model appears appropriate.

4. MODELING AND PARAMETRIZATION

4.1 Derivation

This chapter combines several ideas from Section 2 within the context of the 512-sensor data set discussed in Section 3. In particular the maximum likelihood estimation ideas from Section 2, representations of the covariances of the estimates, and standard dimensional reduction concepts have been combined to provided an analysis framework for the 512-sensor data set.

The first component required for the proposed analysis is a model for the non-stochastic portion of the measurement dynamics. For this model we have assumed that, in a time range of fixed environmental composition, the measurements will converge exponentially to a particular fixed value, plus noise. Figure 1 shows an example time series which is indicative of the reason for making this assumption for the current analysis. Of course, the assumption of exponential decay may very well be incorrect for some, if not many, of the 512 sensors. Perhaps for a given sensor it is even correct for some analytes and incorrect for others. One of the novel parts of our approach is that the model mismatch is *detected and accounted for in the final fusion algorithm*.

Note that we do not assume that during any interval the measurements reach their final equilibrium state. For that reason we do not estimate the *current position* of a measurement, but instead endeavor to estimate the *most likely asymptotic final value* of the set of measurements. As shown in Figure 2, the sensor value returned

for a given environment depends on the sensor’s past environments, whereas the asymptotic value will not. This is akin to a joint state/parameter estimation problem, quite common in target tracking.

To be precise, we endeavor to find a curve of the form

$$y = f(a, b, c, x) = ae^{bx} + c \quad (3)$$

that best fits the data in a least-squares sense. As the dependence of the errors in the fit are a non-linear function of the input parameters $[a, b, c]$, we solve this problems using a Levenberg-Marquardt method originally appearing in.^{16,17} Levenberg-Marquardt couples a gradient descent method with a Gauss-Newton algorithm to combine the robustness of the gradient descent with the fast convergence of Gauss-Newton approach.

One interesting facet of the above approach is that once a non-linear optimization problem is proposed, there is great freedom in the choice of *which* non-linear problem is actually solved. In particular, the parametrization chosen in Equation 3 in terms of $[a, b, c]$, while rather simple, is not the most desirable for our actual numerical calculations. Other parametrization one might consider are

$$y = f(y_0, \dot{y}_0, y_\infty, x) \quad (4)$$

where y_∞ represents the asymptotic value of y , or

$$y = f(y_0, \dot{y}_0, \ddot{y}_0, x). \quad (5)$$

The parametrization in terms of $[a, b, c]$ can be determined from the other parametrization (4 and 5) by solving the following equations:

$$y_0 = ae^{bx_0} + c \quad (6)$$

$$\dot{y}_0 = abe^{bx_0} \quad (7)$$

$$\ddot{y}_0 = ab^2e^{bx_0} \quad (8)$$

$$y_\infty = c \quad (9)$$

Note, all of these parametrizations are *equivalent*. We say two parametrization $[a_0, b_0, c_0]$ and $[a_1, b_1, c_1]$ are equivalent if there exists a diffeomorphism g such that

$$[a_0, b_0, c_0] = g(a_1, b_1, c_1). \quad (10)$$

In other words, a solution to the problem of interest in one parametrization can be smoothly mapped to a solution in the other parametrization.

Of course, two equivalent parametrization may not have the same performance within a numerical scheme such as Levenberg-Marquardt in terms of their robustness and convergence rates, and indeed that is the case here. Our experimentation reveals that the parametrization in Equation 5 is superior to that in both Equation 4 and Equation 3.

4.2 Results

Given the problem derivation in Section 4.1, we study the 512-sensor data set discussed in Section 3. We endeavor to categorize the various environments given in the 512-sensor data set by the c parameter we compute using the non-linear estimator in the previous section. The c parameter is the asymptotic value we predict as x goes to infinity, and it should be independent of the previous environment as we move along the time sequence.

Our first set of calculations is intended to be a proof of concept, and makes use of the fact that we know over which regions the environment is constant. Note that we do not make any assumptions about the environment composition in each segment, though we do know when the environment changes. Of course, such an assumption cannot be made for a real system, but one can consider these calculations to be *training data* which we use as an input to some recursive method. They identify the characteristics of the data (such as the moments of their probability density function) that we may endeavor to detect using a recursive estimator.

Our first result is shown in Figure 2. In this plot we have used the batch estimator above to compute the asymptotic c value of the exponential decay. In Figure 2 we show three different environments labeled AIR, WET, and WET+IDLH. Note that even though the measurement sequences are quite different, both the AIR

and the WET regions get similar markings, both in the mean and covariance of the estimates. On the other hand, while the WET+IDLH area is visually similar to the adjoining WET areas, the both the mean and covariance of the estimate are substantially different.

In Section 5 we analyze the performance of sensors based upon several characteristics, not the least of which is the variances they produce. The variance derivation in Section 4.1 plays a similarly substantial role, now just viewed from a different perspective.

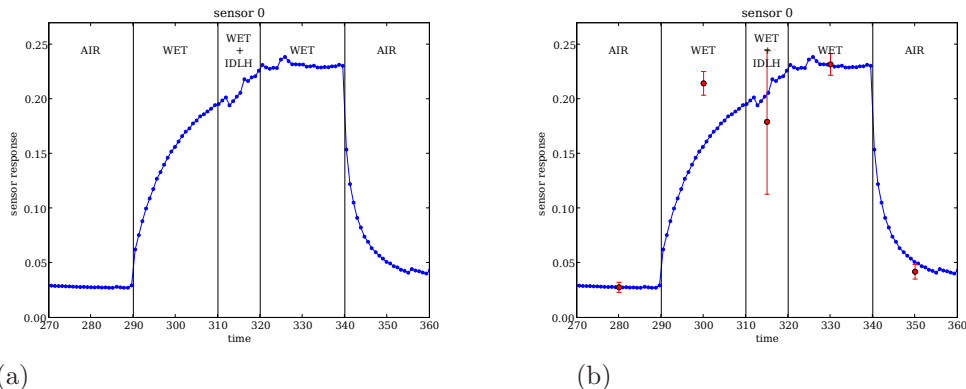


Figure 2. Plot (a) shows that while the measurements appear to decay to a fixed value, plus noise, based upon the current state of the environment, their exact values depend not only upon the final state they are attracted to, but also upon their initial state. For example, the two regions marked AIR should both identified as such even though the morphology of the two measurement sequences are quite different. Plot (b) shows an example calculation. The red dots denote the mean of the c estimate using the estimator from Section 4.1, while the error bars denote the variance. Roughly speaking, the WET and AIR regions are well differentiated because the distance between their means, normalized by their variances, is large. On the other hand, the smaller amount of data and large variance in the WET+IDLH region makes the differentiation of the WET and WET+IDLH regions less apparent. Accordingly, in Section 5 we study those sensors that have variances which are ubiquitously small even in the presence of regions with small amounts of data.

5. SENSOR SELECTION

As opposed to Section 4 where the focus is on computations involving single sensors, here we change our focus to sensor fusion. In particular we show in Section 6 that a naïve Fisher Linear Discriminants (FLD) implementation gives rise to erroneous results. In this chapter we examine the 512-sensor data set and propose preprocessing techniques that are applicable to more robust FLS analysis. One feature of particular note in this section is a discussion of the subtle roles that both errors and variances play in the analysis of sensor quality.

5.1 Data Analysis and Normalization

For an FLD-like method to be maximally effective, it is required that the input data satisfy certain constraints. For example, in Figure 3 one may observe that the maximum value of a particular sensor’s response varies wildly over the set of 512 sensors, by as much as four orders of magnitude. As is well known, these large variations have a tendency to emphasize the sensors with large values over the sensors with small values in an FLD analysis, even when the sensors with large values may not be particularly sensitive to the analytes of interest or are particularly noisy. Accordingly, we take the obvious step of normalizing each sensor so that they all vary over a fixed range, in this case $[0, 1]$.

We leverage the results in Section 4 and compute the c values for both the unnormalized and normalized sensors. As one might expect, and is demonstrated in Figure 3, the c values for the unnormalized data vary even more wildly than the original data, while the c values for the normalized data are much more well behaved. Modulo a small number of outliers, the maximum c value for each sensor varies over only one order of magnitude and presents a much more promising scenario for an FLD analysis. One might be tempted to remove the few outliers and proceed with an analysis, but the derivation of the c values in Section 4 gives rise to a more powerful and rigorous approach.

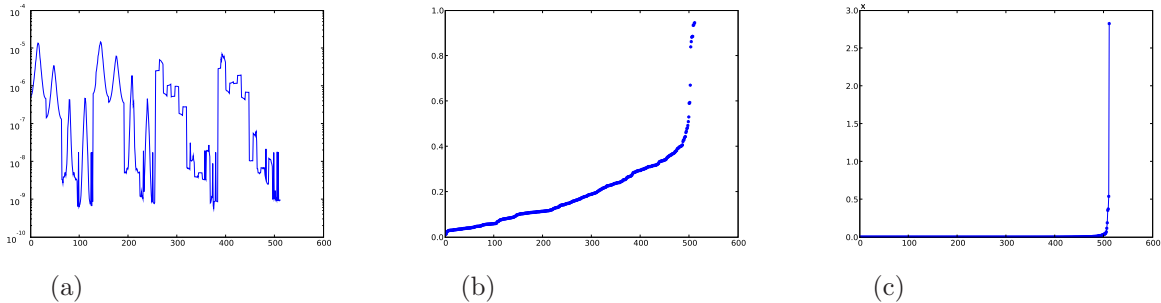


Figure 3. Plot (a) shows the maximum amplitude achieved by each sensor in the 512-sensor data set over the entire run. The maximum sensor c values vary by more than four orders of magnitude, which, in many discrimination schemes, will cause the smaller sensor responses to be essentially ignored. To allow all sensors a fair chance at selecting for the analytes of interest, we first normalize the sensor data by scaling it such that it falls in the $[0, 1]$ interval. Plot (b) shows the normalized maximum RSS error for each sensor, after normalization, sorted by value. This offers a quick glimpse of the relative frequency of sensors with a given level of accuracy. In particular the worst 5% or so of the sensors exhibit a much poorer fit to the exponential model than do the majority of sensors; when we select sensors to be used for discrimination, we will prohibit these sensors from being used. Plot (c) shows the maximum standard deviation reported by the Levenberg-Marquardt minimizer. These values show a dramatic variation and will also be used to select which sensors are used for classification.

In particular, the calculations in Section 4 provide not only values of c for each environmental region across the 512 sensors, but also estimates of the *variances* (the square of the standard deviation) of c and the *error* of the exponential decay fit. Figure 3 shows plots of the maximum standard deviation and error for both the normalized data. As one may observe, the maximum standard deviations and errors range over several orders of magnitude. There are several possible reasons for the wide variation, including an insufficient number of iterations of the Levenberg-Marquardt, variations in the noise characteristics of different sensors, or, perhaps most importantly, a sensor whose response is not correctly captured by the exponential decay model in Section 4.

Regardless of the reason, the statistical information provided by the Levenberg-Marquardt algorithm provides a powerful tool for separating those sensors that are well modeled by the exponential decay assumptions in Section 4 from those that are not. At this point the attentive reader might wonder why we examine both the variances and the errors, since the errors are just draws from the probability density function defined by the variance. Unfortunately, there are subtleties of which one must take into account for a proper evaluation of the sensors. First, a Levenberg-Marquardt algorithm returns a variance which assumes the underlying distribution of the errors is Gaussian. If this is not the case, then the variance can be overly optimistic and the actual errors can be much larger than one would expect. Second, if one has small amounts of data, it might very well be that the errors of the fit are unrealistically small even though the variance of the estimated c parameter, in a Cramer-Rao sense, is large. In other words if we desire to accurately describe a sensor’s motion with an exponential decay c parameter, we must ensure that two conditions are met: the exponential model must fit fairly well with the observed data, and the variance of the computed c parameter must not be too large. Accordingly, we use *both* variance and error as measures of the quality of the sensors.

Currently, our criterion for choosing the cut-off for the errors and standard deviations is rather ad hoc; we choose the cut-offs so that approximately $\frac{1}{4}$ of the sensors would remain after down selection. Of course, more rigorous criteria are certainly desirable, but even these rather rough cut-offs lead to a set of sensors which has given rise to the substantially improved results in our FLD analysis in Section 6.

6. DISCRIMINATION RESULTS

6.1 Introduction

In this chapter we apply the results from Section 2, Section 4, and Section 5 to standard methods in classification (Fisher linear discriminants). We compare and contrast classification results along two dimensions of interest. Along the first dimension, we evaluate the potential performance gains one may achieve by fusing a subset consisting of the best (in a variance and error sense) eleven out of 512 sensors versus all 512. Along the second

dimension, we examine performance changes when one classifies the c value estimates rather than the raw sensor data (Section 4).

We demonstrate an application of FLD to the data from Section 3 in Figure 4. In keeping with industry standard practices, we show Receiver Operating Characteristic (ROC) curves for each of the four classifiers. If we assume that the two classes (positive and negative for a given analyte) have identical covariances and that a false positive is just as undesirable as a false negative, then the decision threshold for FLD falls at zero. This threshold is plotted as a red circle in Figure 4. As one can see, the application of Fisher discriminants to the original data does not provide an accurate classification of the measurements which have an IDLH concentration of our analyte versus the those measurements which do not have an IDLH concentration of our analyte. Similarly, using either the computed invariant descriptor c values or the low error/low variance sensors by themselves* also leads to inaccurate classification. It is only the use of the computed invariant descriptor c values in tandem with the low error/low variance sensors that leads to a completely correct classification.

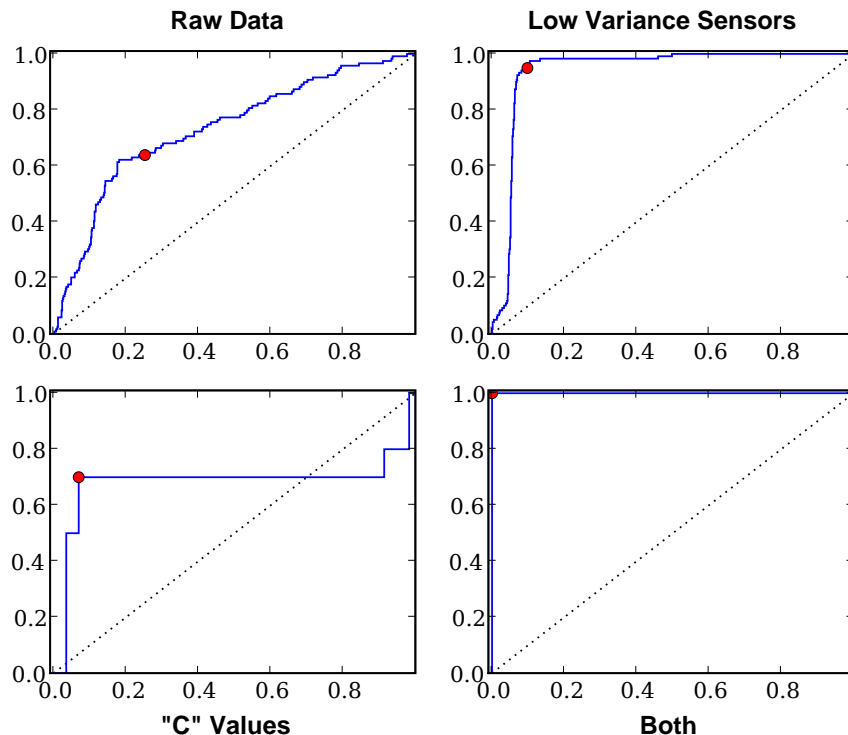


Figure 4. Four examples of Fisher linear discriminants using various preprocessing techniques. Within each plot the y-axis is used to differentiate training data from testing data as well as measurements with IDLH concentration of our analyte from those without. The top left plot uses the raw data for all the sensors, the top right plot uses the raw data from with low error/low variance c values, the bottom left plot uses the invariant descriptor c values from all the sensors, and the bottom right plot uses the invariant descriptor c values from the low error/low variance sensors. The best (in this case, complete) classification of the testing data occurs for the combination of invariant descriptor c values and low error/low variance sensors (bottom right). Note that *both the modeling and covariance calculations* in Section 4 are required to achieve 100% correct classification. Without proper covariance calculations, so common in target tracking but quite uncommon in chemical detection, correct classification is not achievable in this context.

*For this particular plot we use the c values to select the sensors, but use the raw sensor data for the selected sensors in the FLD analysis.

7. CODA: EXTENSION TO OTHER SENSOR MODALITIES

As a brief coda, we note that these results can be extended to other sensor modalities and fusion scenarios. In particular, the authors have extended the methods used for electronic olfaction to more advanced sensors such as Raman spectrographs. Herein we merely foreshadow that work, which is presented in detail in another paper.¹⁸

One might imagine that advanced, data rich sensors, such as Raman spectrographs, do not suffer from the same ambiguities as simple electronic olfaction sensors. Unfortunately, this is not the case. For example, even very simple calculations reveal that the underlying statistics of such sensors can be rather complicated, as shown in Figure 5.

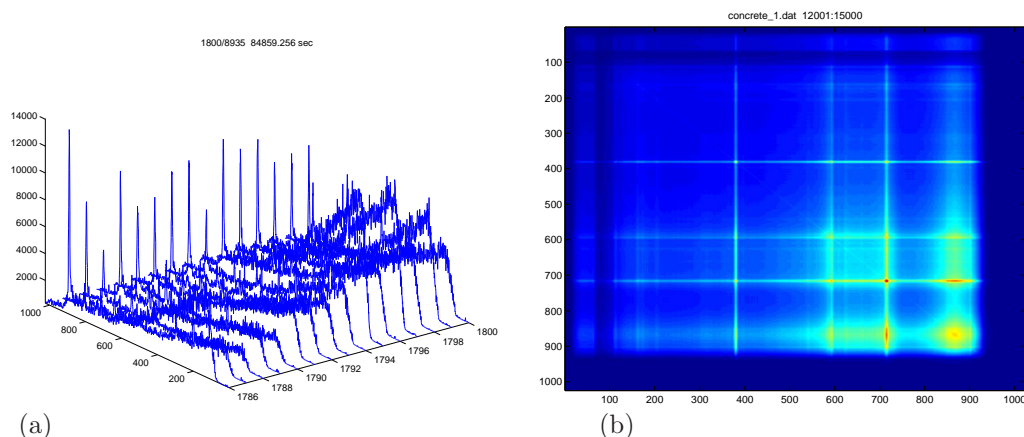


Figure 5. Plot (a) is a collection of measurements from a Raman spectrometer stacked in time. Given sufficient data one can attempt to compute the distribution of the 1024 wave numbers as a 1024 dimensional mean and 1024 x 1024 covariance matrix. We have plotted such a covariance matrix by mapping each element of the matrix to pixels in image (b). Blue values correspond to small entries and red values correspond to large entries.

Given a covariance matrix that represents a “normal” or “background” response, as in Figure 5, one might attempt to detect chemicals of interest by looking for differences between these estimates of the background and spectra measured from the environment. In particular, given a Gaussian representation for the background one might use a Mahalanobis distance to detect whether a new spectra should be considered from the “background” or not.¹⁸

Unfortunately, as discussed elsewhere,¹⁸ such calculations can be rather deceiving. For example, consider Figure 6 where we plot histograms for several different wave numbers along with two confidence intervals, one computed by way of a Gaussian fit and one computed directly from the data. As can be seen, the tails of the distributions are too large, causing the confidence intervals to differ drastically. Accordingly, similar to the results in this paper, a Gaussian assumption must be used judiciously even for sensors which are much more advanced than the simple chemresistors analyzed here.

As an example of the poor results which can stem from inaccurately modeled pdfs, consider Figure 7. Here we compare the system end results of interest to the warfighter — in terms of false alarm frequency — when the background noise is accurately modeled versus the case where the background distribution is close to, but not quite, Gaussian. As one can see in Figure 7 the accurate estimate of the pdfs of interest cannot be under emphasized.

8. CONCLUSIONS

One of the major highlights of our work is deriving a batch parameter estimator to solve a non-linear least squares problem based on the Levenberg-Marquardt algorithm. We have examined the entire 512-sensor data set, and created a preprocessing method to determine the sensors that provide more important information for the purpose of detecting the analytes. It has been our observation that the effectiveness of the provided set of sensors varies wildly and we therefore endeavored to separate out sensors which satisfy a particular set of statistical measures. Here we have another of the core results of this work. From first principles, and *without a priori recourse to information on particular sensor morphologies*, we detected which sensors are well modeled

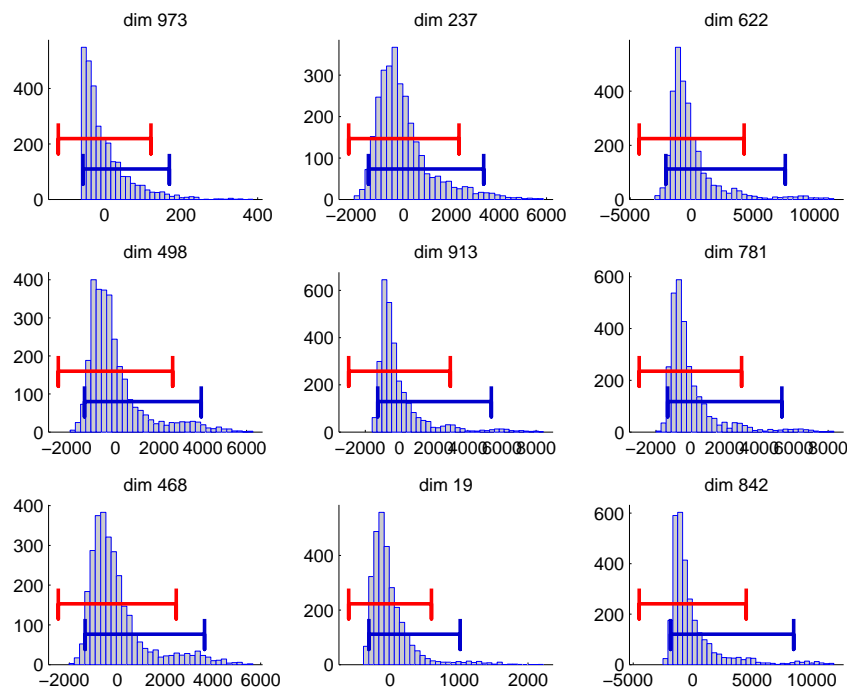


Figure 6. Histograms from some random dimensions of a Raman data set. The two intervals are 95% confidence intervals. The top interval is that predicted by the Gaussian fit, and the bottom interval is derived by counting the actual data. While many of the distributions are fairly Gaussian, they are not exactly so, a fact which is evidenced by the disparate confidence intervals.

by the work in Section 4. Using these methods we are able to select the best sensors, removing those that may pollute future calculations, but retaining those that provide useful information. Section 6 wraps up our study on discrimination by applying Fisher’s Linear Discriminants to a carefully chosen subset of the 512 sensors in the data set. We demonstrate improvement in data classification using both exponential modeling or using sensor set pruning.

ACKNOWLEDGMENTS

The authors acknowledge DHS contract NBCHC080063 and MDA contract HQ0006-06-C-7330 with DTRA MIPR for their support of this work. We also want thank Nick Coult, Allison Floyd, Scott Lundberg, Aubrey Poore, and Louis Scharf for their guidance and support of this work.

REFERENCES

- [1] Ho, C. K., Robinson, A., Miller, D. R., and Davis, M. J., “Overview of sensors and needs for environmental monitoring,” *Sensors* **5**, 4–37 (February 2005).
- [2] James, D., Scott, S. M., Ali, Z., and O’Hare, W. T., “Chemical sensors for electronic nose systems,” *Microchimica Acta* (149), 1–17 (2005).
- [3] Albert, K. J., Lewis, N. S., Schauer, C. L., Sotzing, G. A., Stitzel, S. E., Vaid, T. P., and Walt, D. R., “Cross-reactive chemical sensor arrays,” *Chem. Rev.* **100**, 2595–2626 (2000).
- [4] Persaud, K. and Dodd, G., “Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose,” *Nature* **299**, 352–355 (September 1982).
- [5] Scott, S. M., James, D., and Ali, Z., “Data analysis for electronic nose systems,” *Microchimica Acta* (156), 183–207 (2007).
- [6] Duda, R. O., Hart, P. E., and Stork, D. G., [*Pattern Classification*], John Wiley (2001).
- [7] Blackman, S. and Popoli, R., [*Design and Analysis of Modern Tracking Systems*], Artech House, Norwood, MA (1999).

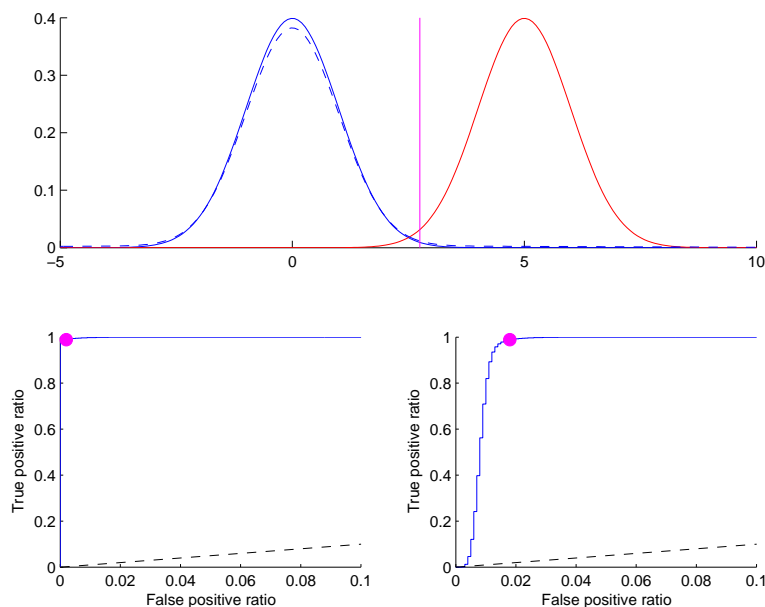


Figure 7. The effects of poorly modeled pdfs. The upper figure shows three pdfs: the two left pdfs (in blue) represent the background distribution, and the right pdf (in red) is the analyte of interest. Assuming that the chosen chemical classifier can separate the classes to a distance of five standard deviations — a reasonable assumption — we then choose a threshold at 2.75 standard deviations from the mean of the background, such that the false positive rate will be one in 500. The bottom left ROC curve shows the performance tradeoffs using this model when the background is actually Gaussian (solid blue line). However, if the assumption of Gaussian background noise is only slightly incorrect — depicted as the dotted blue pdf with slightly fatter tails — use of the same threshold generates false positives one in every 30 times, or more than 16 times more often. The ROC curve for this near-Gaussian distribution is shown in the bottom right.

- [8] Bar-Shalom, Y., Li, X. R., and Kirubarajan, T., [*Estimation with Applications to Tracking and Navigation*], John Wiley & Sons, Inc., New York (2001).
- [9] Bar-Shalom, Y., ed., [*Multitarget-Multisensor Tracking: Advanced Applications*], Artech House, Dedham, MA (1990).
- [10] Vanslyke, S. J. and Wentzell, P. D., “Real-time principal component analysis using parallel kalman filter networks for peak purity analysis,” *Analytical Chemistry* **63**, 2512–2519 (1991).
- [11] Wentzell, P. D. and Vanslyke, S. J., “Parallel kalman filter networks for kinetic methods of analysis,” *Analytical Chimica Acta* **257**, 172–181 (February 1992).
- [12] Wentzell, P. D. and Vanslyke, S. J., “Parallel kalman filters for peak purity analysis - extensions to nonideal detector response,” *Analytical Chimica Acta* **307**, 459–470 (May 1995).
- [13] Chen, J. C. and Rutan, S. C., “Identification and quantification of overlapped peaks in liquid chromatography with uv diode array detection using an adaptive kalman filter,” *Analytical Chimica Acta* **335**, 1–10 (December 1996).
- [14] Yosinski, J., Coult, N., and Paffenroth, R., “Network-centric angle only tracking,” in [*Signal and Data Processing of Small Targets*], SPIE (2009).
- [15] Yosinski, J. and Paffenroth, R., “A distributed database view of network tracking systems,” in [*Signal and Data Processing of Small Targets*], SPIE (2008).
- [16] Levenberg, K., “A method for the solution of certain non-linear problems in least squares,” *The Quarterly of Applied Mathematics* (2), 164–168 (1944).
- [17] Marquardt, D., “Algorithm for least-squares estimation of nonlinear problems,” *SIAM Journal on Applied Mathematics* (11), 431–441 (1963).
- [18] Yosinski, J. and Paffenroth, R., “Estimation algorithms for raman chemical detection,” (2009). in preparation.