Classifying Algorithms for SIFT-MS Technology and Medical Diagnosis

K.T. Moorhead[1], D. Lee[2], J.G. Chase[1], A.R. Moot[3], K.M. Ledingham[3, 4], J. Scotter[4], R.A. Allardyce[3, 4], S.T. Senthilmohan[4], Z. Endre[5]

[1]Department of Mechanical Engineering, University of Canterbury, Private Bag 4800, Christchurch, NZ
[2]Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, NZ
[3]Department of Surgery, Christchurch School of Medicine, PO Box 4710, Christchurch, NZ
[4]Syft Technologies Ltd., PO Box 28149, Christchurch, NZ
[5]Department of Medicine, Christchurch School of Medicine, PO Box 4345, Christchurch, NZ

Corresponding Author: K. Moorhead, ktm19@student.canterbury.ac.nz

Abstract

Selected Ion Flow Tube - Mass Spectrometry (SIFT-MS) is an analytical technique for real-time quantification of trace gases in air or breath samples. SIFT-MS system thus offers unique potential for early, rapid detection of disease states. Identification of Volatile Organic Compound (VOC) masses that contribute strongly towards a successful classification clearly highlights potential new biomarkers. A method utilising kernel density estimates is thus presented for classifying unknown samples. It is validated in a simple known case and a clinical setting before-after dialysis. The simple case with nitrogen in tedlar bags returned a 100% success rate, as expected. The clinical proof-of-concept with seven tests on one patient had an ROC curve area of 0.89. These results validate the method presented and illustrate the emerging clinical potential of this technology.

1 Introduction

Selected Ion Flow Tube - Mass Spectrometry (SIFT-MS) is a relatively new analytical technique for the real-time quantification of volatile organic compounds (VOCs) [1, 2]. It relies on chemical ionisation of trace gas molecules in air or breath samples introduced into a helium carrier using H_3O^+, NO^+ and/or O_2^+ precursor ions. Hence, the identity of a contaminant can be found by comparison of the mass of the product ions with an existing database. The sensitivity of the instrument is currently around five parts per billion in real time.

The SIFT-MS system can offer unique capability in the early and rapid detection of a wide variety of diseases, infectious bacteria and patient conditions. This outcome can be achieved by creating disease and normal mass scan datasets using SIFT-MS, and developing classification methods to identify an unknown patient as normal or diseased. By identifying which masses (and therefore VOCs) contribute most strongly towards a successful classification, biomarkers for a particular disease state can be discovered.

SIFT-MS is a quantitative mass spectrometric method that exploits the chemical ionisation of positively charged precursor ions that react with the VOCs in an air or breath sample. H_3O^+, NO^+ and O_2^+ precursor ions are typically used, since they do not react with the main compounds found in air or breath (N_2, O_2, CO_2 and Ar). The process steps are summarised below, and illustrated in Figure 1.

1. Precursor ions are generated by passing water through a microwave discharge;
2. A quadrupole mass filter is used to select the required precursor ion based on its mass/charge ratio;
3. The precursor ion is injected into a fast-flowing inert carrier gas (helium), which carries the precursor ion, and drawn in test sample, along the flow tube;
4. The precursor ion reacts with the VOCs from the sample to form product ions;
5. A representative proportion of the product ions then pass through a small orifice at the downstream end of the flow tube, and into a differentially pumped quadrupole mass spectrometer that filters ions according to mass;
6. The selected product ions pass to the channeltron particle multiplier/detector where they are counted.
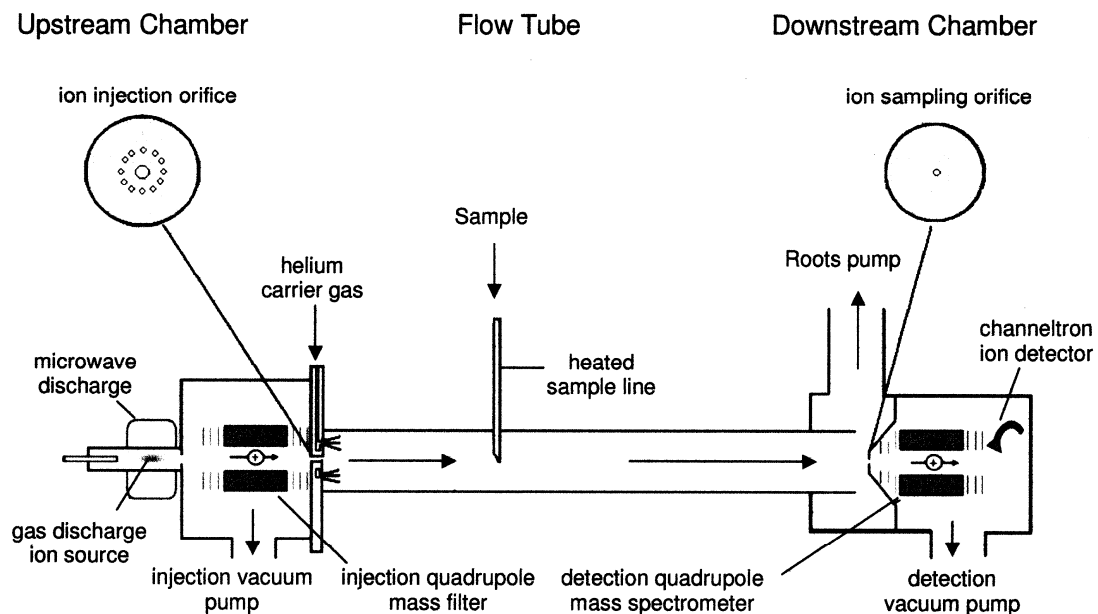

Figure 1: SIFT-MS [3]

SIFT-MS can be used in a variety of applications, ranging from environmental and agriculture sciences, such as analysing exhaust gases, polluted town air and soil emissions, to food safety and medical science, such as smoking, cancer, and infectious diseases.

This paper presents new classification methods and tests for the classification of sets of mass scan data. This process is broken into four steps:

1. Pre-processing to remove noise from the raw mass scan data;
2. Creating probability distributions for each of the 2 test classification groups;
3. Obtaining a classification and a reliability measure for that classification;
4. Identifying useful biomarkers.

Two cases are presented. The first case is a simple direct validation study that aims to differentiate 'dry' nitrogen samples from 'wet' nitrogen samples. The second case study uses the classification model in a clinical setting to determine the differences between dialysis patients before and after treatment, thus examining kidney function, which has direct application in critical care and drug dosing. The classification model is also able to determine which masses are most useful in this classification and therefore those compounds that act as biomarkers for kidney function.

2 Methodology

The study methodology is divided into three sections:

1. Experimental design for the validation study and dialysis case study;
2. Pre-processing of mass scan raw data;
3. Statistical analysis, including classification, prediction error estimation, reliability, and sensitivity/specificity analysis.

2.1 Experimental Design

2.1.1 Validation Study: Nitrogen (N_2) in Tedlar Bags

A simple test study was conducted to validate the statistical classification model developed. Samples of gas collected remotely from the SIFT-MS device are collected in a Tedlar bag composed of polyvinyl fluoride, which the

manufacturers claim to be chemically inert [4]. Before using the tedlar bags, the manufacturers recommend flushing the bags with purified air or nitrogen. Others have shown using Solid Phase Micro-extraction (SPME) that the Tedlar bags emit 15 different VOCs into samples stored in the bags for 24 hours, as detailed in Table 1. Phenol, acetone and acetic acid are persistently present even after purging the bags with purified nitrogen [5]. Hence, one outcome of this study is determining the ability to detect and account for these effects in this common form of breath sample collection.

Table 1: VOCs detected by SPME in samples of N_2 stored in Tedlar bags for 24 hours (modified from Parker et al. [5]).

VOC	No Purge	Purged bag
Acetone	+	+
Acetaldehyde	+	-
Decane	+	-
Octane	+	-
Toluene	+	-
2,3-Butadione	+	-
Ethyl Benzene	+	-
Nonane	+	-
Isoprene	+	-
Tridecane	+	-
Dodecane	+	-
Tetradecane	+	-
Acetic Acid	+	+
Pentadecane	+	-
Phenol	+	+

In this study, multiple samples of N_2 in Tedlar bags were tested by performing mass scans over a range of 10-150 atomic mass units (amu) using the precursors H_3O^+, NO^+ and O_2^+. The bags were new and all flushed at least three times prior to testing directly from the bag.

N_2 in Tedlar bags was also vented to sterile glass bottles filled with water using a sterile stainless steel needle, and a polytetrafluoroethylene (PTFE) permeable septum. The gas in the bottles was then tested by performing mass scans. Table 2 indicates the number of samples obtained using both methods.

Table 2: Number of control tests performed

Sample type	# of samples tested
N_2 in Tedlar bag (Dry Nitrogen)	25
N_2 in Tedlar bag vented to bottle (Wet Nitrogen)	25

The specific objectives of this first study are as follows:

1. Determine what VOCs are added when venting to a Tedlar bag containing purified nitrogen;
2. Determine which VOCs or their water clusters increase or diminish when passed through sterile glass bottles filled with water;
3. Differentiate between the 2 test groups;
4. Validate the classifier using this presumably 'clean' system.

2.1.2 Case Study: Dialysis

One patient with impaired kidney function underwent dialysis treatment on seven separate occasions, with breath mass scans performed one hour into the treatment and after 4 hours of treatment. These mass scans were taken over a range of 10-150 amu using the precursors H_3O^+, NO^+ and O_2^+. The aims of this study are:

1. To develop probability density profiles for the pre- and post-dialysis groups;
2. To determine if it is possible to reliably differentiate between the 2 sample groups;
3. To identify possible new biomarkers for kidney function.

The study is based on the fact that dialysis takes patient a from kidney failure to functional status (artificially). Hence, it provides a built in comparison for analysing kidney function.

2.2 Pre-processing

SIFT-MS mass scans measure the concentration of products at each 0.2 of a mass unit to obtain results. However, there is a significant amount of machine and precursor noise, as those shown in Figure 2.
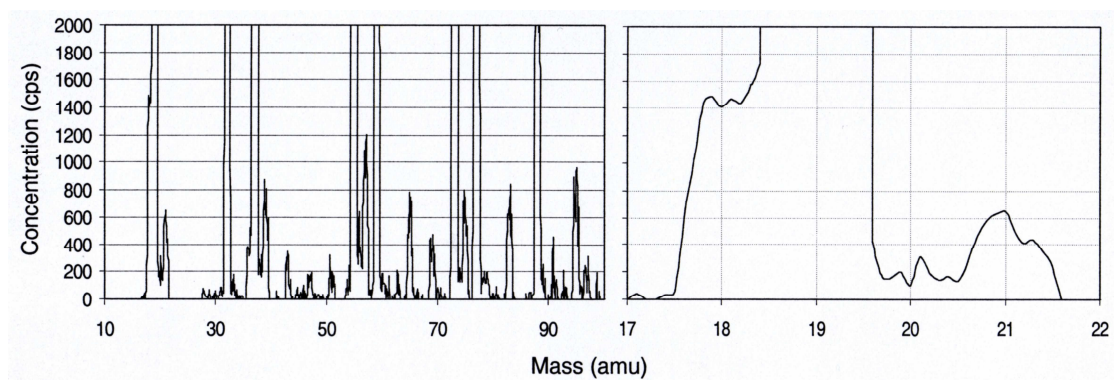

Figure 2: Typical Mass Scan over 10-100 amu (left) and focused on 17-22 amu (right)

The data contains noise, anomalous electronic signals produced by the SIFT-MS instrument, which must be filtered before analysis. This filtering process must result in an accurate reading of concentration at each whole number mass unit. A normal peak from a mass scan is usually centred around a whole mass unit. At half its height, it is usually 0.8 mass units wide, as shown in Figure 3. However, as seen in Figure 4, this definition of a standard peak is not always seen in practice, except at masses with large concentrations. The

7

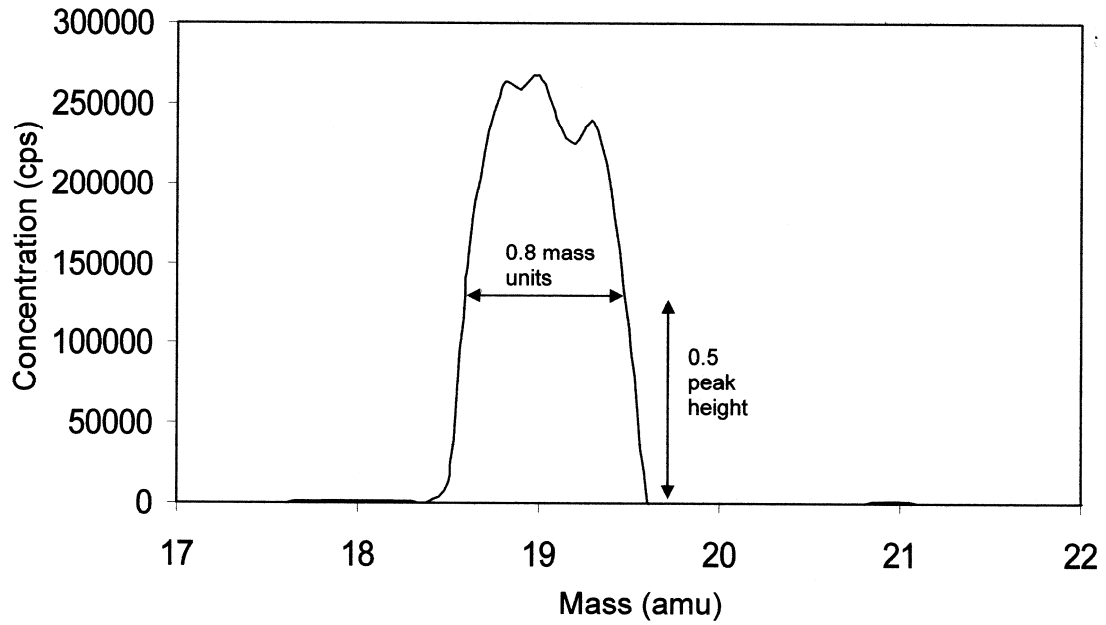actual peaks in Figure 4 are considered to be at the positions indicated by crosses.
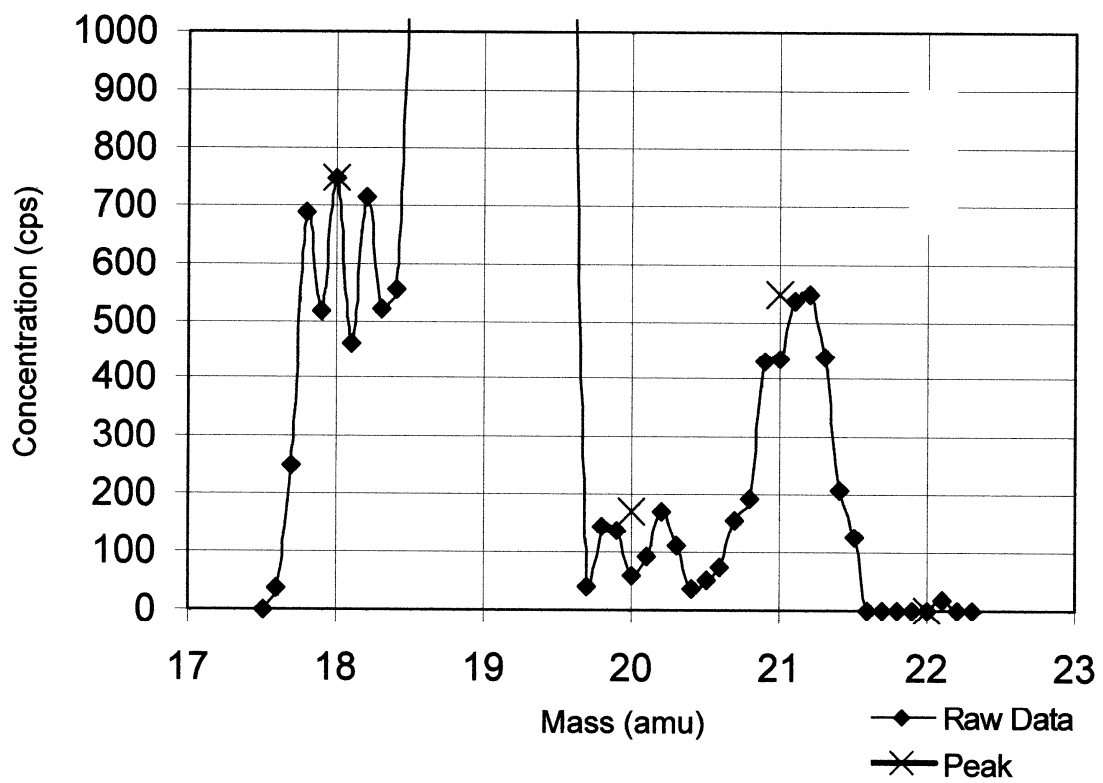

Figure 3: Typical Mass Scan Peak at 19 amu

Figure 4: Noise in the Raw Data at 18, 20 and 21 amu

8

Figure 4 is obtained using the H_3O^+ precursor. The H_3O^+ precursor has mass 19, thus accounting for the off-the-scale peak seen at this mass. The H_3O^+ precursor will also generate peaks at masses 37, 55 and 73, due to H_3O^+ forming clusters with H_2O. A large concentration of precursor ion is used in SIFT-MS, to ensure complete reaction of the VOCs in the sample. Thus the mass scan should always show maximum peaks at the precursor-related masses. A non-precursor related peak that is greater than the maximum precursor peak indicates an erroneous result because the particular VOC present in the sample would have swamped the precursor signal. Because a different amount of precursor ion is used for each reaction, mass scans are normalised to the sum of the product precursor peaks. (For the H_3O^+ precursor, this normalising factor is the sum of the peaks at masses 19, 37, 55 and 73.)

Pre-processing steps therefore involve:

- Normalising each data point to the sum of the precursor ion peaks, allowing comparisons to be drawn between samples that receive a different amount of reactant precursor ion;
- Removing large erroneous peaks greater than the precursor peaks, and data points that cause a peak to rise too sharply. This process eliminates machine errors, and samples with high concentrations of compounds that swamp the precursor signal;
- Finding the tip of the peak, by looking at concentrations before and after the whole mass unit, and determining if a suspected peak exists in reality by considering the number of non-zero readings within 1 mass unit;
- Reducing the concentration matrix to include only whole mass values for analysis.

2.3 Statistical Classification Analysis

The goal of classification is to compare the concentrations obtained at each mass in the mass scan of an unknown sample with a previously established database. The end result is a classification of the unknown sample into one of the database groups. As well as the classification, an estimate of the prediction error and a reliability measure are required. Biomarkers can be obtained by determining which masses in the mass scan were most useful in classifying the unknown sample.

2.3.1 Classification

Kernel density estimates are used for the classification of mass scan data [6]. Test datasets of known classification are used to develop probability density profiles for each of the two datasets: Groups j and k. An unknown sample is then tested against the datasets, with the result being a classification into either Group j or Group k. In a diagnostic analysis, j and k would be the non-disease and diseased states.

After pre-processing the raw mass scan data, each sample is left with a vector of concentration values at each whole mass unit. With mass scans performed over a range of 10 -150 amu, the resulting vector has length 141. For each mass, a mixed distribution made up of a kernel density and a dirac delta function is used to develop a density profile from each group using each sample's concentration value at that mass. Stronger, generally smoother density profiles are obtained with greater numbers of different mass scan samples. Typical density profiles fitted to two datasets are shown in Figure 5.
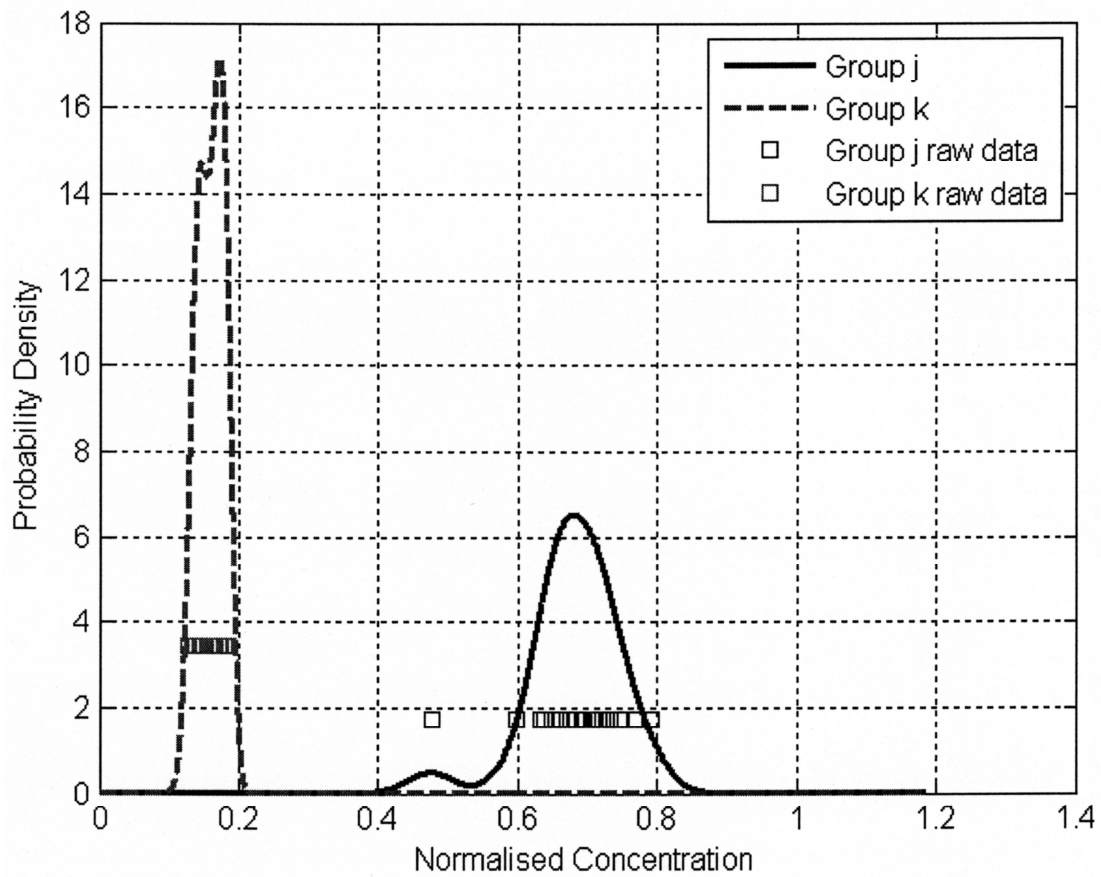
Figure 5: Fitting Probability density profiles to raw data

When a mass scan from an unknown sample is obtained, pre-processing is used to create a vector of concentration values at each of the 141 masses. For each mass, the probability densities for groups j and k at the concentration of the unknown sample are compared. Let x_0 denote the concentration of the unknown sample at the given mass, and $f_j(x_0)$ and $f_k(x_0)$ denote the probability densities of groups j and k respectively at the concentration of the unknown sample. This situation is illustrated in Figure 6, where the plot shown would be useful for classification as given by the minimal overlap of distributions.
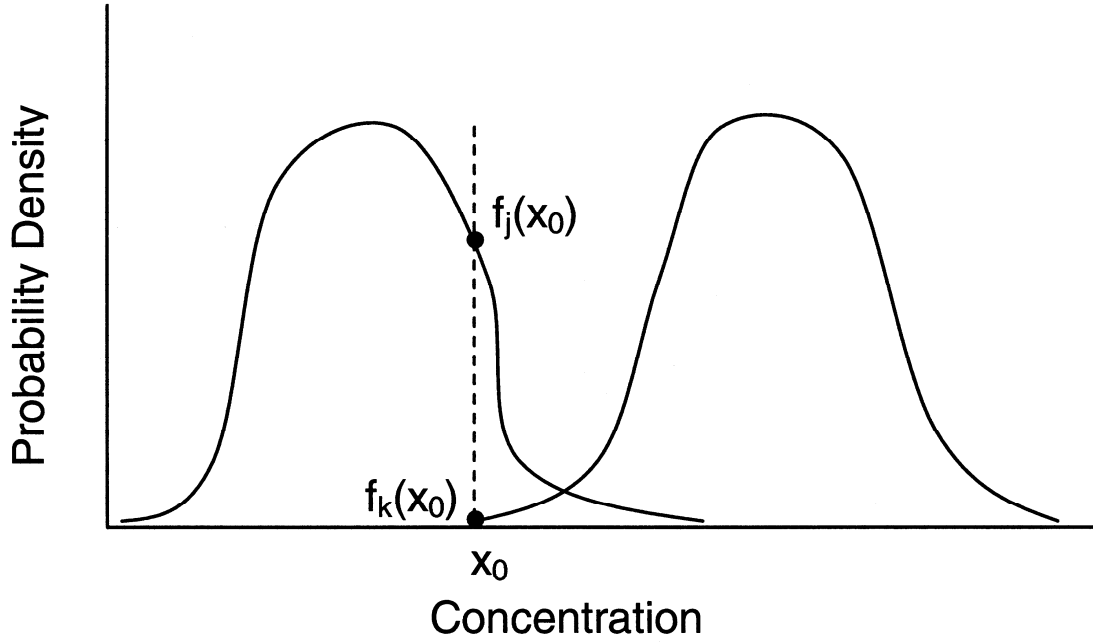
Figure 6: Probability densities for a given mass

For each mass, Equation (1) gives the probability of the given sample being from Group j, given the concentration value obtained at that mass is x_0.

$$\hat{Pr}(j \mid x_0) = \frac{\hat{\pi}_j \hat{f}_j(x_0)}{\hat{\pi}_k \hat{f}_k(x_0) + \hat{\pi}_j \hat{f}_j(x_0)} \qquad \text{Eqn (1)}$$

where $\hat{\pi}$ is the prior probability (prior) of the sample being in that group. When no information is known about the data, the priors are set to 0.5. If the ratio in Equation (1) is greater than a specified threshold, q, then the sample is classified as being in the numerator group (Group j), otherwise it is classified as being in Group k.

The log-odds ratio is defined as the natural log of the probability of a group j classification divided a group k classification.

$$LogoddsRatio = \ln(\frac{\hat{Pr}(j \mid x_0)}{\hat{Pr}(k \mid x_0)}) \qquad \text{Eqn (2)}$$

Equation (2) can be broken down into Equation (3), whereby the log-odds ratios are effectively summed over all masses to result in a final log-odds ratio over all masses.

$$\ln\frac{\hat{\Pr}(j\mid x_0)}{\hat{\Pr}(k\mid x_0)} = \ln\frac{\hat{\pi}_j}{\hat{\pi}_k} + \sum_{m=1}^{M}\ln\frac{\hat{f}_j(x_0., m)}{\hat{f}_k(x_0., m)} \qquad \text{Eqn (3)}$$

A final log-odds ratio greater than $\ln[q/(1-q)]$, where q is the threshold described above, indicates the sample is in the numerator group (Group j); otherwise it is in the denominator group (Group k).

2.3.2 Bootstrap Method for Estimating Prediction Error

Once density profiles have been created for the two datasets, the stratified bootstrap method is used to estimate the prediction error of the classification model. Bootstrap samples are created by choosing with replacement from the original sample until a bootstrap sample is created that is the same size as the original sample [6]. For example, if Group j contained 20 sets of mass scans, and Group k contained 25 sets of mass scans, each bootstrap sample would contain 20 Group j scans and 25 Group k scans, with some scans included more than once, and some completely left out. This process is repeated B times, producing B bootstrap datasets, where B is sufficiently large to ensure that all patients are left out of at least 1 bootstrap dataset.

The bootstrap estimate of the classification error is defined:

$$\hat{E}^{(1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{|C^{-i}|} \sum_{b \in C^{-i}} I(\hat{y}^b(x_i) \neq y_i) \qquad \text{Eqn (4)}$$

where N is the total number of samples; $|C^{-i}|$ is the number of bootstrap samples that do not contain sample i; $\hat{y}^b$ is the classifier trained on bootstrap sample b; $I(\hat{y}^b(x_i) \neq y_i)$ equals 1 if sample i is classified incorrectly and 0 otherwise.

The kernel classifier is trained with the bootstrap datasets that do not contain sample i as the training set, and then use sample i as the test set. This process is repeated for all bootstrap samples that do not contain sample i. The total of incorrect classifications is summed and divided by the number of bootstrap samples that did not contain sample i. This overall process is repeated for each sample. It concludes by averaging the number of incorrect classifications over all samples, giving $\hat{E}^{(1)}$.

The bootstrap estimate is biased upward as an estimate of the true classification error, and this limitation is alleviated using the 0.632 estimator [6], defined:

$$\hat{E}^{(0.632)} = 0.368 \times \bar{e} + 0.632 \times \hat{E}^{(1)} \qquad \text{Eqn (5)}$$

where $\bar{e}$ is the biased error, calculated using all the data as the training set, and testing each sample against this set. This result is biased downward of the true error because the test data is also found in the training set.

The overall prediction error should be compared with random classification, for which the overall error rate, *P**, is given by:

$$P^* = p_j(1 - \pi_j) + (1 - p_j)\pi_j$$

Eqn (6)

With a prior probability set at 0.5, the overall error rate for a random classifier is 50%, regardless of group proportions (p_j and p_k) in the sample.

2.3.3 Reliability

Density profiles of the log-odds ratio obtained from the bootstrap method can be plotted for each group, and a reliability curve can be generated, as shown in Equation (7) and Figure 7.

$$\widehat{\Pr}(S = j \mid \lambda) = \frac{\widehat{f}(\lambda \mid S = j)\widehat{\Pr}(S = j)}{\widehat{f}(\lambda \mid S = j)\widehat{\Pr}(S = j) + \widehat{f}(\lambda \mid S = k)\widehat{\Pr}(S = k)}$$

Eqn (7)

where $\widehat{\Pr}(S = j \mid \lambda)$ is the probability of the sample, *S*, being in Group *j* given the log-odds ratio, *λ*; $\widehat{f}(\lambda \mid S = j)$ is the probability of the log-odds ratio for Group *j*; $\widehat{\Pr}(S = j)$ is the probability that the sample is in group *j*.

This situation is illustrated in Figure 7. A good dataset will have minimal overlap between the two density profiles because where a significant difference is detected between sample groups, a large log-odds ratio should always be obtained. As shown in Figure 7, if an unknown sample is classified with a log-odds ratio, *λ*, of +200, one can be approximately 100% certain that classification is correct. This result occurs because at a log-odds ratio of +200, the probability that the sample is in group *k* is approximately 0. Therefore by Equation (7), the reliability reduces to unity, represented here as

100% on the Reliability scale. However, the log-odds ratio obtained where group j and k profiles overlap (at approximately -5 in Figure 7), has a reliability of 50%, indicating an equally likely probability of correct as incorrect classification.
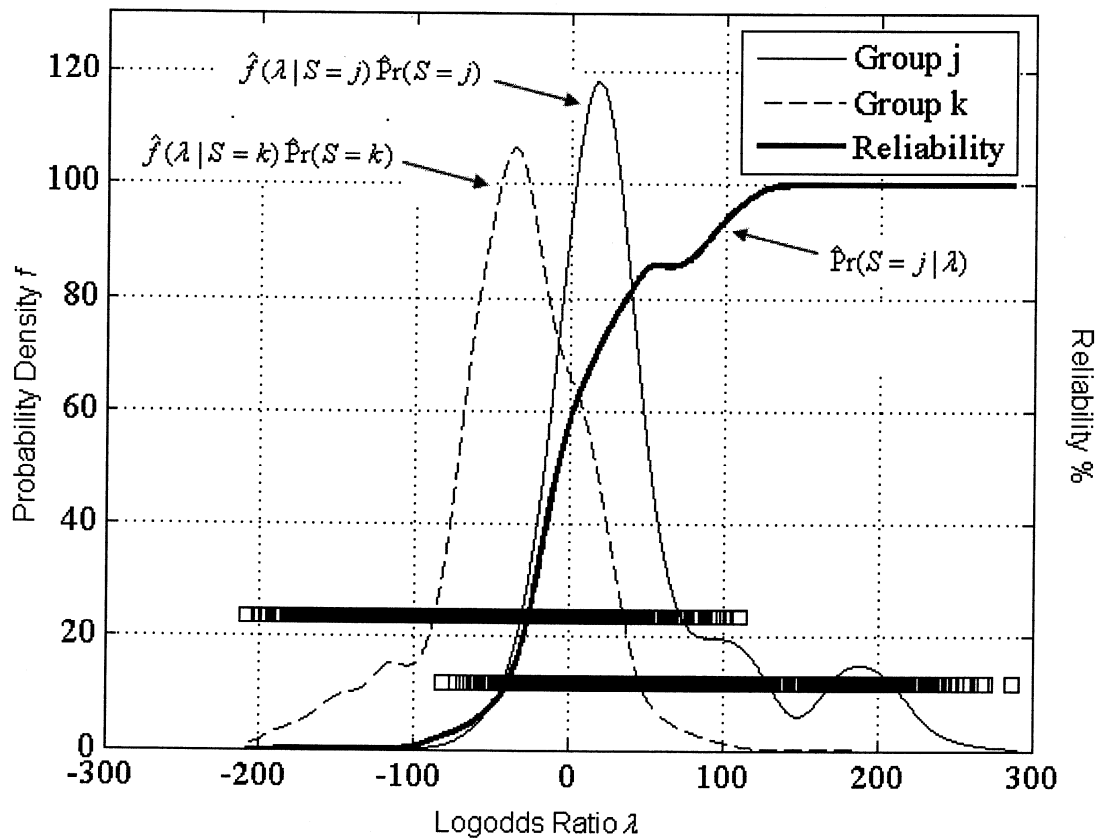

Figure 7: Reliability Curve

2.3.4 Biomarkers

Any sample to be classified, is done so by summing the log-odds ratios at each mass, to give an overall log-odds ratio. With a classification threshold of 0 for Equation (3), a positive log-odds ratio indicates the sample is in Group j, and a negative log-odds ratio indicates that the sample is in Group k. The two examples shown in Figure 8 display relatively large log-odds ratios, with each mass's ratio generally consistently displaying the same sign (positive or

16

negative). Therefore, one can be confident of the final classification. Those masses that contribute large log-odds ratios towards the final classification may be considered as useful biomarkers.
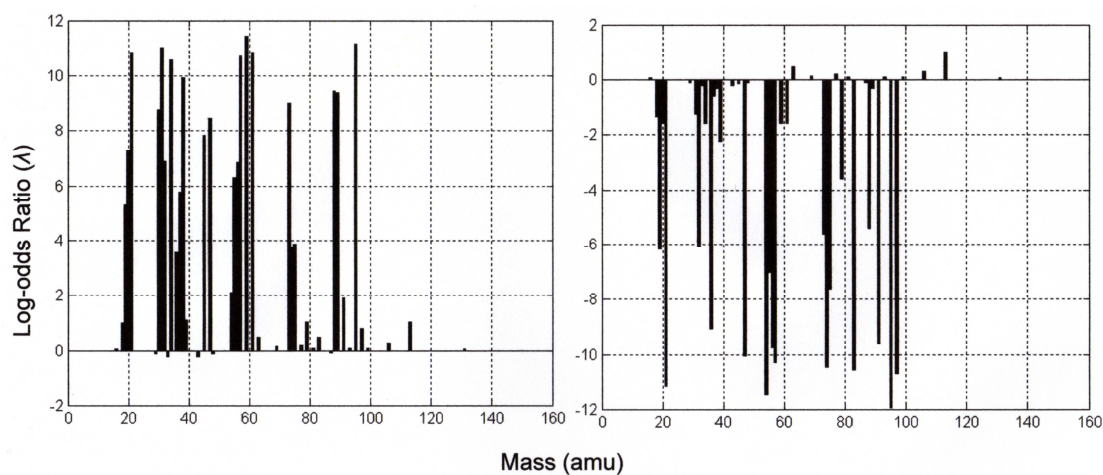

Figure 8: Individual classification

Biomarkers are found by determining which masses have log-odds density profiles with minimal overlap. Density profiles are created as described in Section 2.3.3 for each mass, *individually*. The area of overlap is shaded in the schematic of Figure 9 and is calculated using the Trapezoid Rule. The masses with the smallest areas of overlap are the best biomarkers.
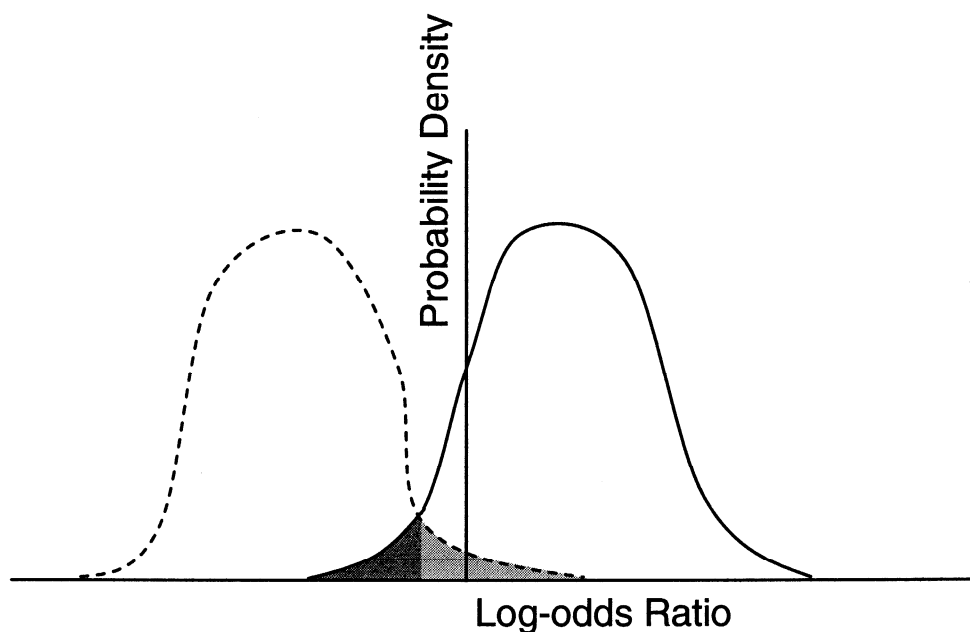
Figure 9: Biomarkers schematic with Log-odds ratio

3 Results and Discussion

3.1 Validation Study: Nitrogen in Tedlar Bags

Using the 0.632 bootstrap estimator method to estimate prediction error, with

B = 500 bootstrap samples, there was 0% classification error over all

precursors, indicating excellent differentiation between sample groups. All

results for this validation study are shown for the H_3O^+ precursor only. In

addition, concentrations shown are normalised to the sum of the precursor

value as described in section 2.2.

3.1.1 Probability Density Profiles

Density profiles including log-odds ratios are shown in Figure 10 for a

selection of the biomarkers noted in Section 3.1.2. Excellent separation is

evident between density profiles for the two groups. Note that the squares

indicate the concentrations of the raw data, and the corresponding curves are

the probability density profiles fitted to that raw data. It is clear in the figure

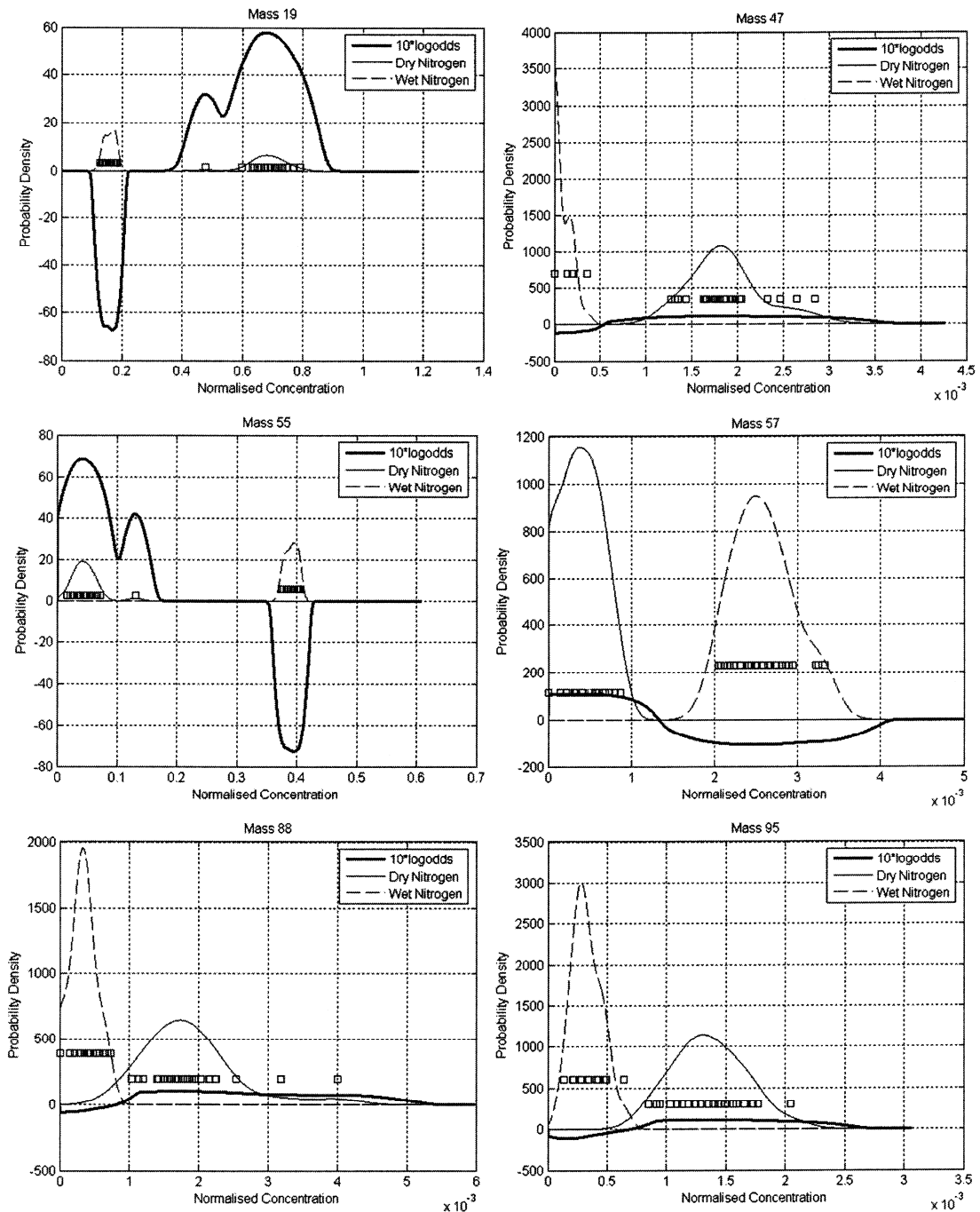that the raw data is well separated between the two groups being compared.

Figure 10: Selected Probability Density Profiles using H₃O⁺ Precursor

3.1.2 Biomarkers

The classification model found that the masses indicated in Table 3 were most useful in aiding towards a correct classification, and they are ranked as described in section 2.3.5 and Figure 9, with the best biomarker having the smallest overlap between density profiles. Their log-odds density profiles are shown in Figure 11.

Table 3: Biomarkers for H_3O^+ classification in validation study

| Product ions	Mass	Explanation	Rank
H_3O^+ and its water clusters	19, 55, 73	The wet nitrogen group showed much higher concentrations at masses 55 (and 73), corresponding to the water clusters of H_3O^+.	3, 4, 5
Isotope of H_3O^+ and its water clusters	57	The wet nitrogen group showed much higher concentrations at mass 57, corresponding to the water clusters of H_3O^+, (and lower concentrations at mass 21, corresponding to the mass of the H_3O^+ isotope (D_3O^+) with no water cluster).	2
$C_4H_9NO.H^+$ (product of N,N-dimethyl acetamide)	88	Due to the solubility of $C_4H_9NO.H^+$ in water, by venting the nitrogen through the water bottle, the concentration at mass 88 decreased dramatically, (and increased at mass 106 – its water cluster).	7
$C_6H_6O.H^+$ (product of phenol)	95	Due to the solubility of $C_6H_6O.H^+$ in water, by venting the nitrogen through the water bottle, the concentration at mass 95 decreased dramatically.	6
$N_2H^+.H_2O$	47	Water cluster of N_2H more prevalent when N_2 passes through water.	1
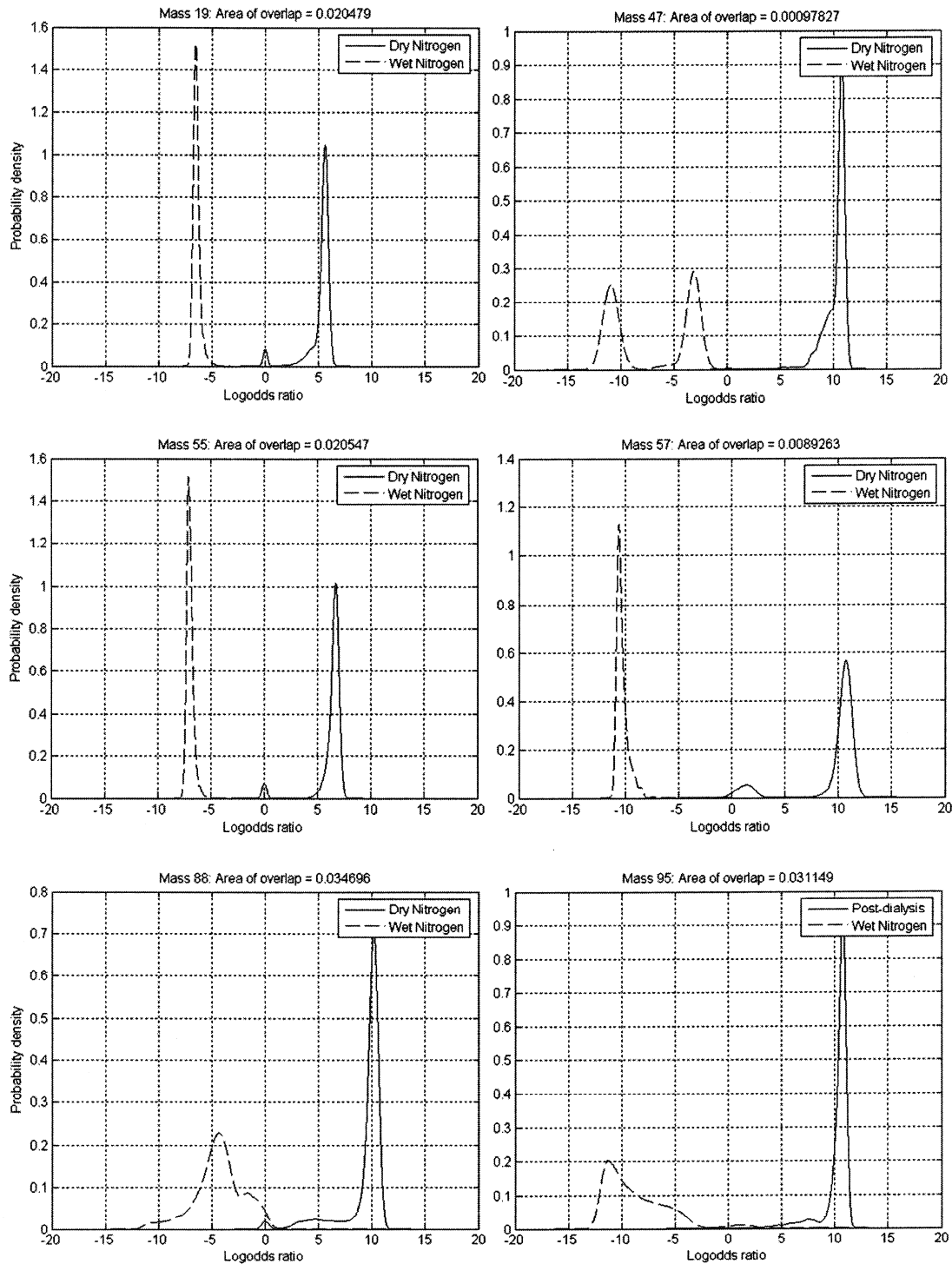
Figure 11: Log-odds Density Profiles for Biomarker Selection in validation study

3.1.3 Reliability

Density profiles were created from the log-odds values obtained from each

bootstrap sample entry for each of the test groups, with Dry Nitrogen

represented as Group j and Wet Nitrogen represented as Group k. It is

21

observed in Figure 12 that there is excellent separation between the two

profiles, with consistently large log-odds values obtained. The relatively flat

gradient of the reliability curve indicates that if an unknown sample were

classified with a small log-odds ratio, the reliability of correct classification

would be relatively low. As indicated by the ROC curve in Figure 13, the

model is highly sensitive and highly specific, with the area under the ROC

curve approaching 1. It should be noted that these ideal results are the

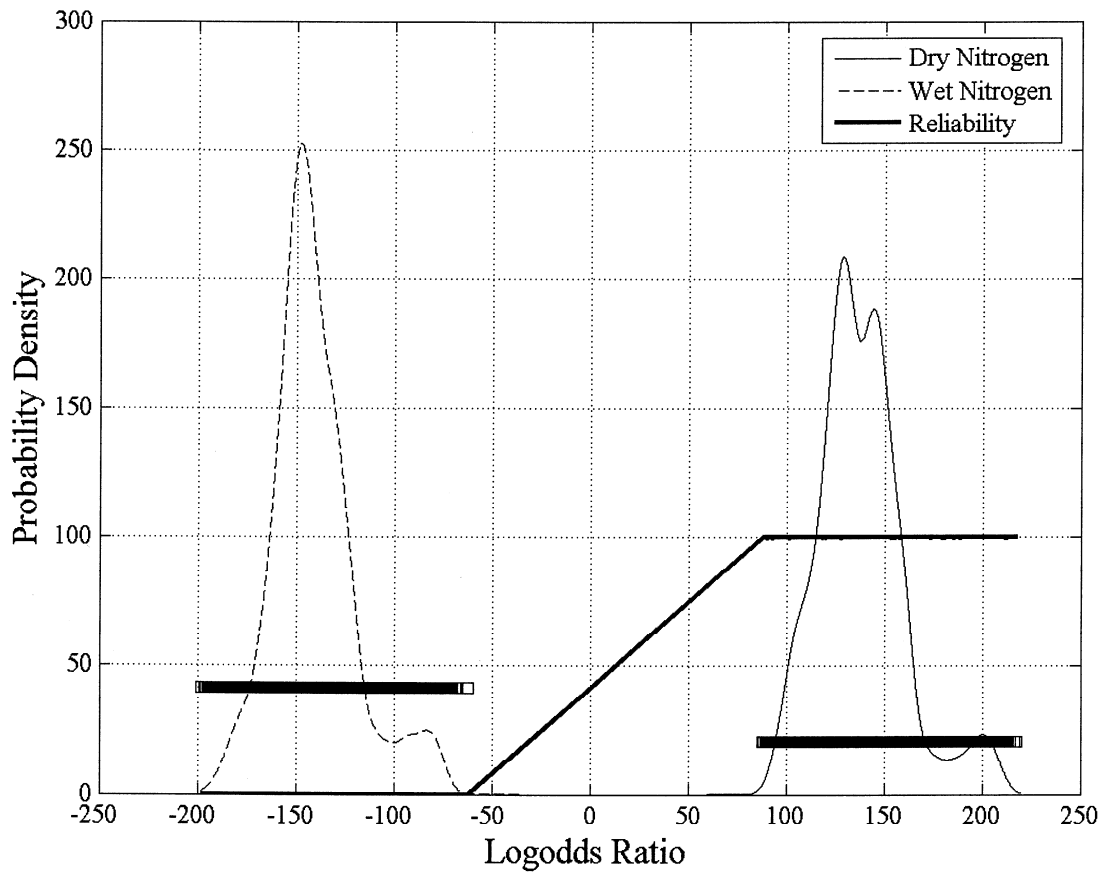expectation for this model classification validation study.

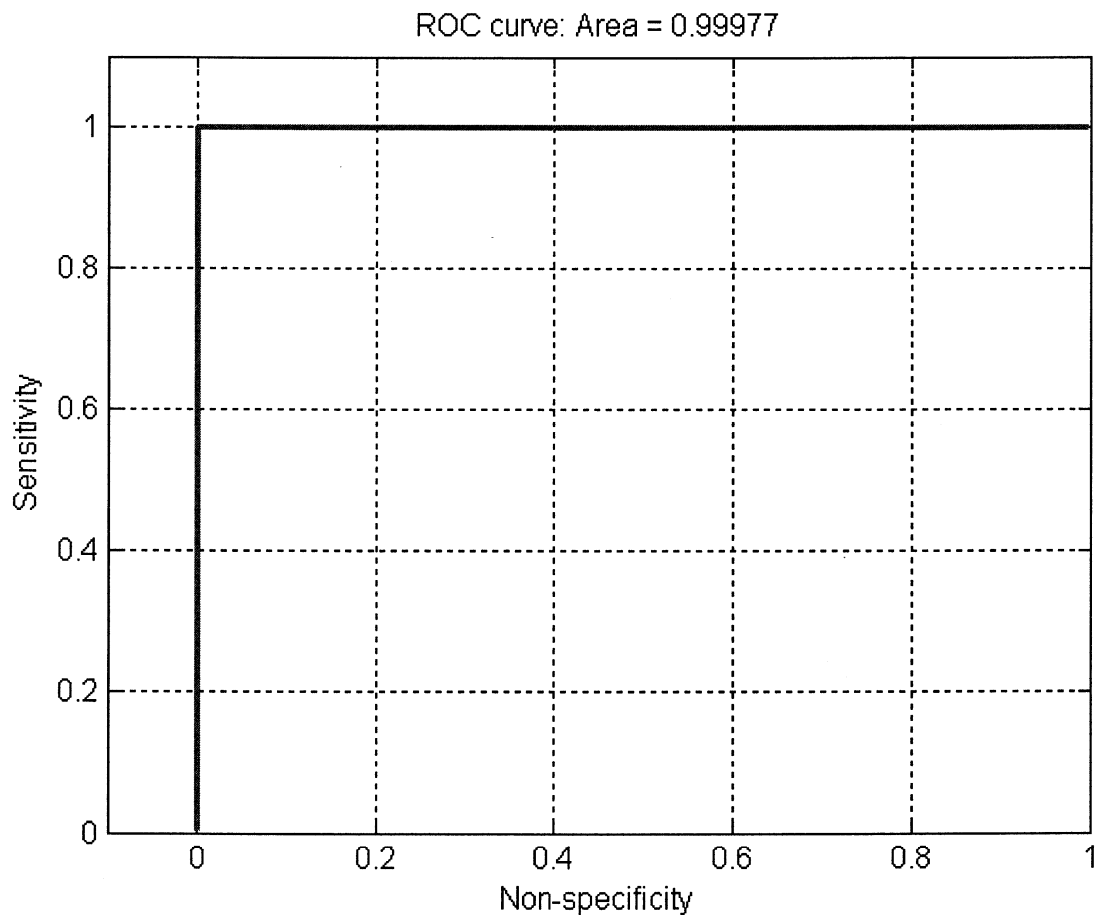**Figure 12: Reliability Curve for Validation Study Classification Model**

Figure 13: ROC Curve for Validation Study Classification Model

3.2 Case Study: Dialysis

Seven repeat mass scans were taken of one dialysis patient over the course of 6 months. These mass scans were taken at time t = 1 hour and t = 4 hours into dialysis treatment. Mass scans were analysed for key biomarkers that aid in a classification between the pre and post -dialysis data sets. A sample size of seven was much smaller than desired, but sufficient to test the classifier concept.

With B = 1000 using the H_3O^+ precursor, the overall prediction error was estimated at 11.2%, with error in classifying pre-dialysis readings contributing 16.7%, error in classifying post-dialysis readings contributing 18.8%, and

biased error contributing 0%. Due to the equal sample sizes in the pre and post dialysis groups and the zero biased error, the overall estimated prediction error in Eqn (5) reduces to 0.632 x average of the pre and post dialysis errors.

Using the NO^+ precursor, the overall prediction error was 26.0%, with pre-dialysis, post-dialysis, and biased errors contributing 19.3%, 46.4%, and 14.3%, respectively.

Using the O_2^+ precursor, the overall prediction error was 18.2%, with pre-dialysis, post-dialysis, and biased errors contributing 19.8%, 37.8%, and 0%, respectively.

3.2.1 Probability Density Profiles

Density profiles including log-odds ratios are shown in Figure 14 for a selection of the biomarkers indicated in Section 3.2.2. It can be seen that the density profiles are relatively strong for the 4-hour dialysis group (Group *j*), because after dialysis treatment, levels of ammonia (masses 18, 36 and 54) and other VOCs fall to normal levels. Depending on factors such as diet and the length of time since the last dialysis treatment, VOC levels can vary dramatically prior to treatment, as seen by the spread of data in the 1-hour dialysis group (Group *k*). Note that the density profiles of the 1 hour data at mass 89 and the 4 hour data at mass 35 are undistinguishable in the figure. This is because all repeat mass scans had the same zero concentration reading for these masses at those time points, and the density profiles are therefore spikes on the y axis.

It is observed that the result is considerably less 'clean' than the Nitrogen validation study. Note that all density profiles are shown for the H_3O^+ precursor only.
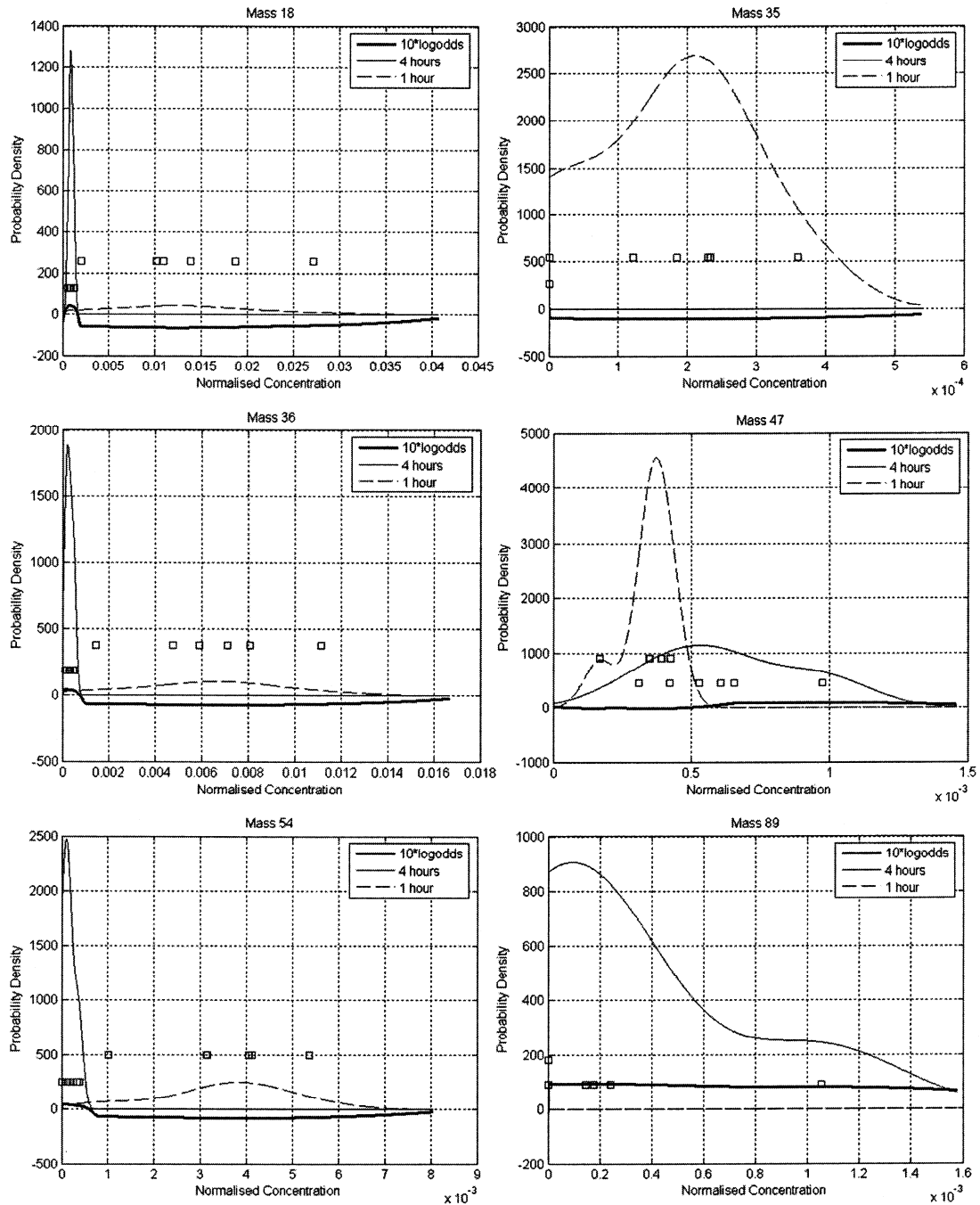

Figure 14: Probability Densities using H_3O^+ precursor

3.2.2 Biomarkers

The classification model found that the masses indicated in Table 4 were most useful in aiding towards a correct classification, and they are ranked as described in section 2.3.4. A selection of their log-odds density profiles obtained using the H_3O^+ precursor are shown in Figure 15. Note that smoother profiles would be obtained with a greater sample size. It is observed that masses relating to ammonia and its water clusters are most useful in distinguishing between pre- and post- dialysis datasets. This result is expected since it is well documented that ammonia concentration is seen to decrease during dialysis treatment [7], as further evidenced by the fact that it is the urea reduction ratio (URR) that is used to measure dialysis efficacy.

Table 4: Biomarkers for classification in Dialysis study

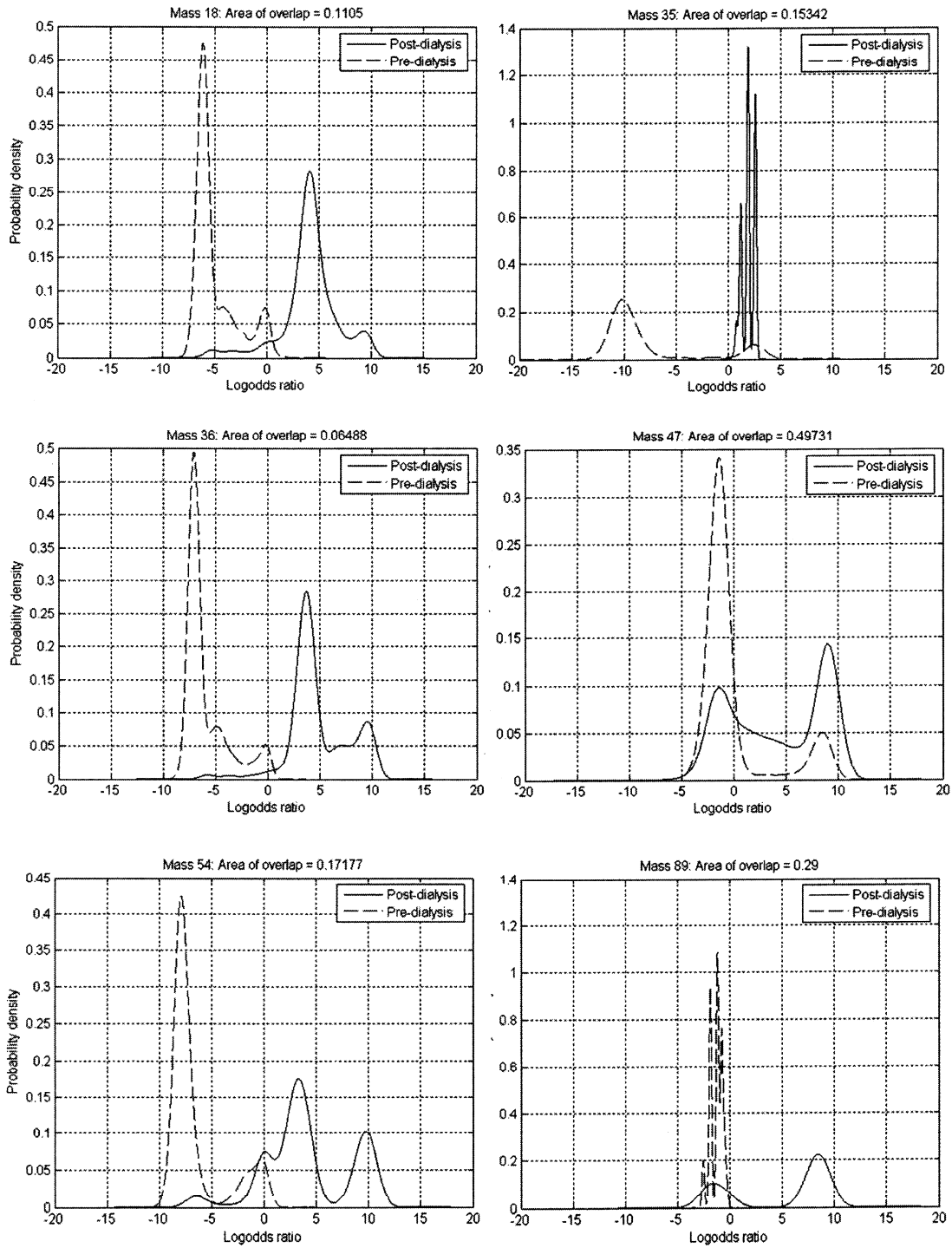| Precursor	Mass	Explanation	Rank
H_3O^+	18, 36, 54	Ammonia and its water clusters	2,1,5
	35, 17	Unknown	3, 4
	89	Acetaldehyde	6
	47	Ethanol	7
NO^+	18, 36	Ammonia	2,3
	47	Unknown	1
O_2^+	53	Ammonia, isoprene	1
	17, 18, 35, 36, 53, 54	Ammonia and clusters	4,3,6,5,1,2
	77, 58	Acetone	7,8

Figure 15: Log-odds Density Profiles for Biomarker Selection in the Dialysis study

3.2.3 Reliability

Density profiles were created from the log-odds values obtained from each

bootstrap sample entry for each of the test groups. Post-dialysis (4 hours) is

represented as Group j and Pre-dialysis (1 hour) is represented as Group k. It

is observed in Figure 16 that there is much greater overlap in the two density profiles compared with the Nitrogen validation study, and an absolute log-odds value of approximately 50 must be obtained to classify with 90% certainty in Figure 16. As indicated by the ROC curve in Figure 17, the model classifies significantly better than random, with an ROC area of 0.89.
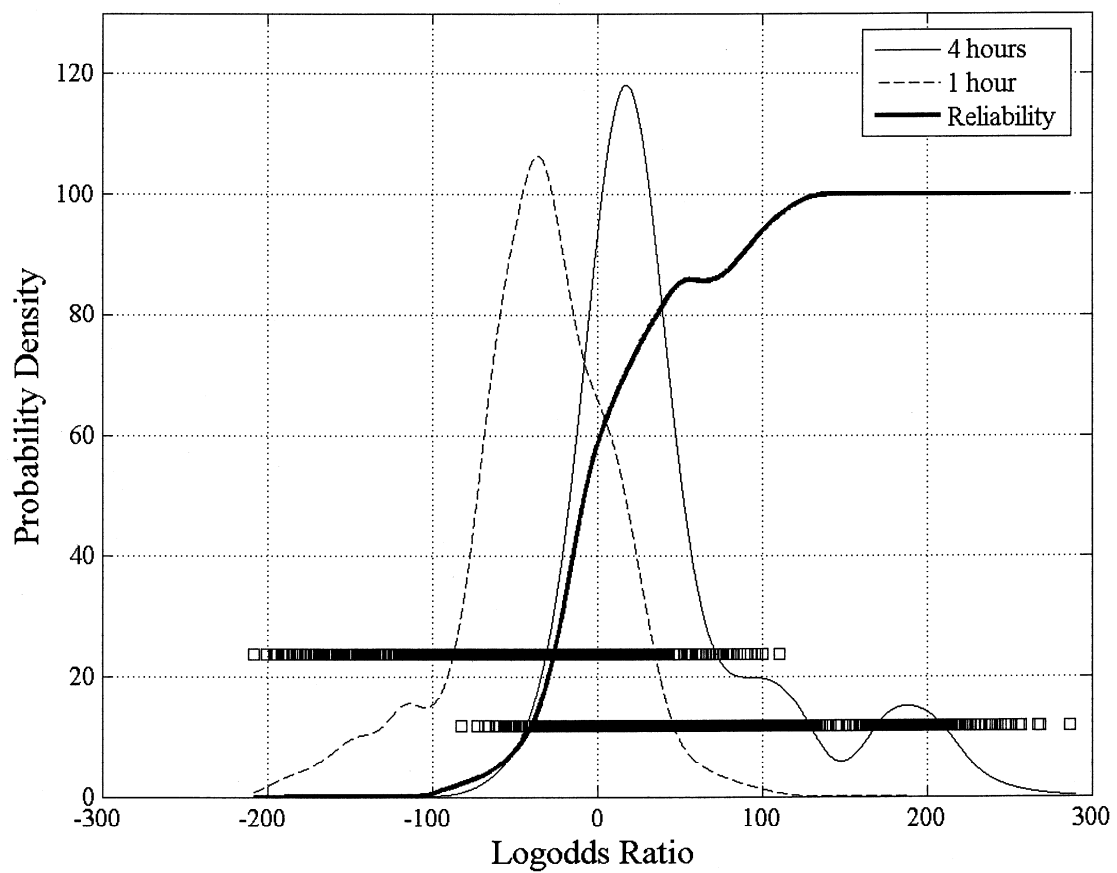

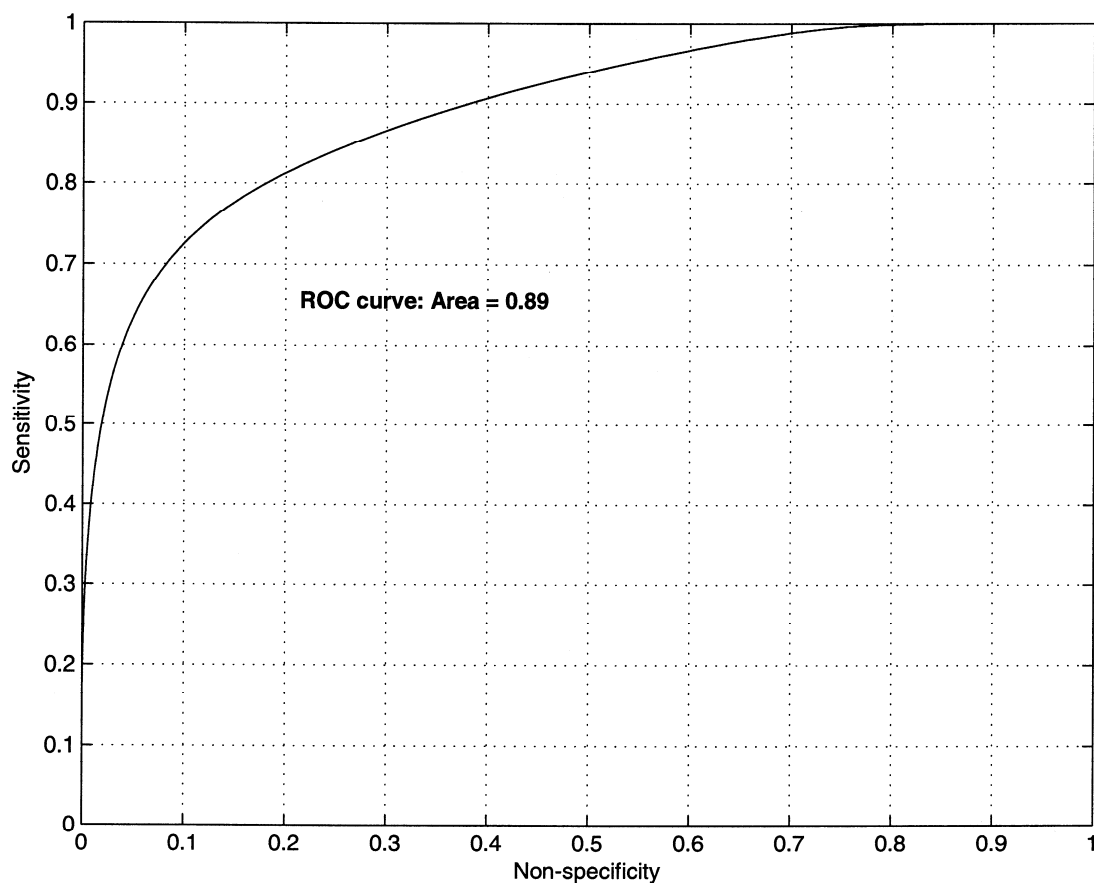**Figure 16: Reliability Curve for Dialysis Study Classification Model**

Figure 17: ROC Curve for Dialysis Study Classification Model

4. Conclusions

The SIFT-MS system can potentially offer a unique capability in the early and rapid detection of a wide variety of diseases, infectious bacteria and patient conditions. A method is presented for the classification of unknown samples. This method is initially validated in a simple study in which saturated nitrogen in tedlar bags is differentiated from dry nitrogen in tedlar bags. There was an estimated 0% error in classification using the bootstrap error prediction method with 500 bootstrap samples. All expected biomarkers were identified, with the most reliable being $N_2H^+.H_2O$, and isotopes and water clusters of H_3O^+. In addition, phenol and N,N-dimethyl acetamide products were found to be present in the tedlar bags despite flushing, although levels were reduced in

saturated nitrogen samples, presumably related to the solubility of these compounds in water.

Because of the lack of data at low log-odds ratios due to the very distinct sample groups, the reliability curve showed that log-odds ratios of approximately 50 and 70 are required for 90% reliability in classifying saturated and dry nitrogen respectively. However, with 500 bootstrap samples, the minimum log-odds ratios obtained were 70 and 85 for saturated and dry nitrogen respectively. Hence, the high reliability is observed. The area under the ROC curve was found to be an expected ideal of 1.00, showing that the classification model is highly sensitive and highly specific.

After initial validation, the classification model was employed to differentiate patient breath samples after one and four hours of dialysis treatment, thus estimating or measuring kidney function. The sample size of seven was much smaller than desired, but sufficient for this first proof-of-concept. Using 1000 bootstrap samples, estimated prediction errors were found to be 11.2%, 26.0% and 18.2% for the H_3O^+, NO^+ and O_2^+ precursors, respectively. Density profiles are relatively strong for the 4-hour dialysis group because after dialysis treatment, levels of ammonia and other VOCs fall to normal levels. Depending on factors such as diet and the length of time since the last dialysis treatment, VOC levels can vary dramatically prior to treatment, as observed by the spread of data in the 1-hour dialysis group. Biomarkers for classification were ammonia, acetaldehyde, ethanol, isoprene and acetone. An absolute log-odds value of approximately 50 must be obtained to classify

with 90% certainty. As indicated by the ROC curve, the model classifies significantly better than random, with an ROC area of 0.89.

A method is presented and validated for breath classification using SIFT-MS mass scan data. As well as classification, this method enables the identification of biomarkers relevant to the condition being investigated. SIFT-MS mass scans used in conjunction with the current method could potentially allow non-invasive real-time diagnosis of various patient conditions, as well as giving insight into the metabolic reactions that occur in certain disease states.

Future work will involve model validation, which will require much greater sample sizes. In addition, as more dialysis mass scans become available, the model will be modified to exploit the paired nature of this type of data. Unlike the pure classification problem described here, with paired data of this type, sample points in each group are intrinsically linked, and this information can be used to further inform the model.

5. References

[1] Smith, D., Spanel, P. (1996). "Application of ion chemistry and the SIFT technique to the quantitative analysis of trace gases in air and on breath". International Reviews in Physical Chemistry, 15(1), 231-271.

[2] Freeman, C.G., McEwan, M.J. (2002). "Rapid Analysis of Trace Gases in Complex Mixtures Using Selected Ion Flow Tube - Mass Spectrometry". Australian Journal of Chemistry, 55, 491-494.

[3] Smith, D. and Spanel, P. (2005). "Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) For On-Line Trace Gas Analysis". Mass Spectrometry Reviews, 24, 661-700.

[4] SKC inc. Gas Sample Bags. http://www.skcinc.com/prod/tedlar.asp#tedlar

[5] Parker, DB, Rhoades, MB, Koziel, J, Spinhirne, J. "Background odours in Tedlar Bags used for CAFO odour sampling". Conference proceedings, ASAE Annual International Meeting, Las Vegas, July 2003. Paper No. 034144.

[6] Hastie, T., Tibshirani, R., and Friedman, J.H. "Elements of Statistical Leaning: data mining, inference and prediction". New York: Springer, 2001.

[7] L. R. Narasimhan, William Goodman, and C. Kumar N. Patel (2001). "Correlation of breath ammonia with blood urea nitrogen and creatinine during hemodialysis". Proc Natl Acad Sci U S A. Apr 10;98(8):4617-21