

# Discrimination and Quantification of Volatile Organic Compounds in the ppb-Range with Gas Sensitive SiC-FETs Using Multivariate Statistics

Christian Bur, Manuel Bastuk, Donatella Puglisi, Andreas Schuetze, Anita Lloyd Spetz and Mike Andersson

Linköping University Post Print



N.B.: When citing this work, cite the original article.

Original Publication:

Christian Bur, Manuel Bastuk, Donatella Puglisi, Andreas Schuetze, Anita Lloyd Spetz and Mike Andersson, Discrimination and Quantification of Volatile Organic Compounds in the ppb-Range with Gas Sensitive SiC-FETs Using Multivariate Statistics, 2015, Sensors and actuators. B, Chemical, (214), 225-233.

<http://dx.doi.org/>

Copyright: Elsevier

<http://www.elsevier.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-115560>

Discrimination and Quantification of Volatile Organic Compounds in the ppb-Range with Gas Sensitive SiC-FETs Using Multivariate Statistics

Christian Bur<sup>1,2\*</sup>, Manuel Bastuck<sup>2</sup>, Donatella Puglisi<sup>1</sup>, Andreas Schütze<sup>2</sup>, Anita Lloyd Spetz<sup>1</sup>, and Mike Andersson<sup>1</sup>

<sup>1</sup>Div. of Applied Sensor Science  
Department of Physics, Chemistry and Biology  
Linköping University  
SE-58183 Linköping, SWEDEN

<sup>2</sup>Lab for Measurement Technology  
Department of Physics and Mechatronics Engineering  
Saarland University  
D-66123 Saarbrücken, GERMANY

\* Corresponding author:

Mail: chrbu@ifm.liu.se

c.bur@lmt.uni-saarland.de

Phone: +49 681 302 3904

Fax: +49 681 302 4665

E-mail addresses of co-authors:

Manuel Bastuck: m.bastuck@lmt.uni-saarland.de

Donatella Puglisi: donpu@ifm.liu.se

Andreas Schütze: schuetze@lmt.uni-saarland.de

Anita Lloyd Spetz: spetz@ifm.liu.se

Mike Andersson: mikan@ifm.liu.se

## Abstract

Gas sensitive field effect transistors based on silicon carbide, SiC-FETs, have been studied for indoor air quality applications. The selectivity of the sensors was increased by temperature cycled operation, TCO, and data evaluation based on multivariate statistics. Discrimination of benzene, naphthalene, and formaldehyde independent of the level of background humidity is possible by using shape describing features as input for Linear Discriminant Analysis, LDA, or Partial Least Squares – Discriminant Analysis, PLS-DA. Leave-one-out cross-validation leads to a correct classification rate of 90 % for LDA, and for PLS-DA a classification rate of 83 % is achieved. Quantification of naphthalene in the relevant concentration range, i.e. 0 ppb to 40 ppb, was performed by Partial Least Squares Regression and a combination of LDA with a second order polynomial fit function. The resolution of the model based on a calibration with three concentrations was approximately 8 ppb at 40 ppb naphthalene for both algorithms.

Hence, the suggested strategy is suitable for on demand ventilation control in indoor air quality application systems.

## Keywords:

Indoor Air Quality, temperature modulation, Linear Discriminant Analysis, Partial Least Squares, trace level detection

## Highlights:

- Detection of typical VOCs in the ppb-range using SiC-FETs
- Discrimination of VOCs using Linear Discriminant and Partial Least Squares – Discriminant Analysis
- Quantification of naphthalene using Partial Least Squares Regression
- Discrimination and quantification independent of background humidity

## 1. Introduction

In today's society, people spend most of their time, up to 85 %, indoors where fresh air exchange is limited. Due to that, the quality of indoor air has become a major issue in the last years. The most commonly used indicator for indoor air quality, IAQ, is carbon dioxide, CO<sub>2</sub>, which causes fatigue at elevated concentrations around 1,000 ppm and can be measured using infrared, IR, absorption. However, the air quality is also strongly affected by volatile organic compounds, VOCs, which pose a serious health risk even at very low concentrations of a few parts-per-billion, ppb [1],[2],[3]. Lack of fresh air can lead to sick building syndrome, SBS, with symptoms like acute discomfort, headache, dizziness, difficulties in concentrating, respiratory problems like asthma, skin irritation, and hypersensitivity to odors and tastes [4], [5].

The French Agency for Food, Environmental and Occupational Health & Safety, AFSSET, suggested in 2006 the first guidelines for limiting the emission of VOCs [6]. In more recent studies, e.g., by the European project INDEX [7] and French Indoor Air Quality Observatory, OQAI, priority lists of air pollutants with an undeniable health impact were suggested. Target VOCs of high relevance are benzene, naphthalene, and formaldehyde.

Benzene is present both indoors and outdoors; however, indoor air concentrations are generally higher than outdoors [3]. Typical indoor sources are furniture, heating and cooking systems, building materials, dyes, and paints, whereas outdoors benzene is released mainly from petrochemical industry, traffic, and gas stations. The concentration of benzene is significantly higher in homes of people who smoke. It was reported that ambient air concentrations vary from sub-ppb (approx. 1 µg/m<sup>3</sup>) in rural areas to low ppb (5-20 µg/m<sup>3</sup>) in urban areas, or even to tens of ppb in source impacted areas [8], [9]. Benzene is classified as carcinogenic to humans and according to the World Health Organization, WHO, there is no safe level of exposure. However, the French decree n. 2011-1727 (December 2, 2011) has established an exposure limit for public buildings of 1.5 ppb in 2013 and 0.6 ppb by 2016 [10].

Naphthalene is also used in paints and the production of insecticides. It outgases from consumer products and other sources today are cigarette smoke and incomplete combustion processes (e.g. vehicle and air traffic, residential heating, etc.). Health effects are mostly respiratory tract lesions, but naphthalene is also classified as a possible carcinogenic [3], [11], [12]. In Germany, guidance limit levels of 0.01 mg/m<sup>3</sup>, i.e. 3.5 ppb, (guide level I), and of

0.03 mg/m<sup>3</sup>, i.e. 5.7 ppb, (guide level II) are suggested [11]. If guide level II is reached or exceeded, immediate action is required as this concentration could pose a health hazard [13].

Formaldehyde is one of the most well-known VOCs and very common in both indoor and outdoor air. Typical sources are outgassing of building materials, consumer products, and textiles. In particular, outgassing from wood-based materials (among others pressed wood products), glues, e.g., from laminate floors, and coatings including wallpaper are a major concern [14]. Exposure to formaldehyde leads to sensory irritations, especially of the eyes. The WHO has classified formaldehyde as carcinogenic for humans and recommends a short-term exposure limit of 81 ppb [3].

In order to reduce the exposure to VOCs in indoor air, heating, ventilating, and air-conditioning, HVAC, systems are usually used. However, a significant amount of energy is required to operate these systems. It was reported that approximately 40 % of primary energy is consumed for heating and air conditioning. Thus, on demand ventilation has potential to improve the quality level and efficiency of such systems considerably. In order to operate such systems effectively, sensors and sensor systems able to measure the air quality, i.e., the amount of pollutants such as VOCs, are required. It has been demonstrated that gas sensitive field effect transistors based on silicon carbide are suitable candidates for indoor air quality measurement systems [15], [16].

Gas sensitive field effect transistors, FETs, have been studied for many years [17], [18], [19]. Besides silicon, which was used in the beginning, silicon carbide, SiC, is currently used as a substrate material as it offers the possibility for high temperature application (up to at least 800 °C for SiC [17]). Due to its chemical inertness, SiC is also a suitable material for sensors operating in harsh environments, e.g., directly in the exhaust stream of combustion engines [18].

Excellent gas sensitivity can be achieved by using catalytically active gate materials like palladium, platinum, or iridium for SiC-FETs. The sensing properties of the FET depend mainly on the material and its structure as well as on the operating temperature of the device. For a dense, homogenous layer of, e.g., palladium or platinum, hydrogen molecules adsorbing on the surface dissociate and rapidly diffuse through the dense metal layer. At the metal-insulator, which is SiO<sub>2</sub> in this work, interface hydrogen atoms form a polarized layer of hydroxyl groups influencing the density of mobile carriers in the channel of the transistor [19], [20].

For the detection of non-hydrogen containing gases like carbon monoxide, CO, but also for ammonia, NH<sub>3</sub>, a porous layer of, e.g., platinum is required which allows direct interaction with the oxide surface like spill over from the metal and detection of dipoles formed on the oxide surface. Additionally, three phase boundaries between metal, oxide, and gas have a higher catalytic activity which is very important for the gas response [21], [22].

It was reported that the sensing mechanism of hydrogen and non-hydrogen containing gases can be explained by spill-over effects of adsorbed oxygen [23]. Reducing gases like CO would react with adsorbed oxygen and thereby lower the density of oxygen on the surface.

In order to enhance the performance of SiC-FET sensors dynamic operation, i.e., temperature cycled operation, TCO, can be used [24]. Temperature modulation gives rise to several advantages [25], among others the transient response of the sensor provides an unique response pattern or signature for each gas thereby increasing the selectivity.

In this work, the detection, discrimination, and quantification of three hazardous VOCs are presented using multivariate statistic tools, i.e., Linear Discriminant Analysis, LDA, Partial Least Squares Regression, PLSR, and Partial Least Squares – Discriminant Analysis, PLS-DA.

## 2. Experimental

In this section, the gas mixing apparatus for supplying ppb-level VOCs is briefly presented (cf. Section 2.1) followed by a description of the field effect transistor and its operating mode (cf. Section 2.2).

### 2.1. Apparatus for ppb-Level Supply of VOCs

Measuring in the low and sub-ppb ranges impose high requirements on the test equipment. For this purpose, a new gas mixing system was realized where the VOCs are supplied either by permeation ovens (VICI Dynacalibrator 150) or by a gas bottle in combination with an additional dilution stage. A detailed description of the used system and the setup can be found elsewhere [26]. In the used setup benzene and naphthalene were supplied by permeation tubes with oven temperatures for benzene and naphthalene of 30 °C and 70 °C, respectively. Formaldehyde was supplied by a gas bottle with a mixture of 50 ppm formaldehyde in nitrogen connected to a dilution stage. Dry synthetic air with a purity of 99.999 % was used as carrier gas. However, it still contains contaminations with concentrations up to ten parts per

million, ppm. Nevertheless, using the same carrier gas in all parts, i.e., for the permeation ovens and the dilution stage, these contaminations form a constant background and thus, do not affect the sensor response. The background contamination of the system was checked using gas chromatography - mass spectrometry, GC-MS, reference measurements [26].

For all experiments the continuous flow over the sensor was kept at a constant level of 200 ml/min which means that when a test gas is injected, the flow of the carrier gas is reduced accordingly. The total flow was monitored during the measurements by a mass flow meter downstream of the sensor. Humidification was realized by a temperature stabilized water bubbler through which a part of the carrier gas was passed.

## 2.2. Sensor setup and operation

For all measurements an n-channel depletion type SiC-FET supplied by SenSiC AB, Kista, Sweden was used (cf. Fig. 1). The catalytic gate metallization is porous platinum with a thickness of 25 nm. A detailed description of the sensor and the manufacturing process can be found elsewhere [19].

In order to allow precise heating of the sensor, the SiC chip is glued onto a ceramic heater (Heraeus GmbH, Germany). A Pt-100 temperature sensor was attached next to the sensor as temperature reference. The heater together with the SiC chip and the temperature sensor is spot-welded on a 16-pin TO8 header (cf. Fig. 1b). Electrical contacts to the transistor structures on the SiC chip are realized via gold wire bonding.

Sensor control and data acquisition were performed by a combined system (3S GmbH - Sensors, Signal Processing, Systems, Saarbrücken, Germany). The sensor temperature was controlled by an analog control circuit with a resolution of 1 °C. The sensor signal, i.e., the drain-source voltage, was measured by a 14 bit ADC with a measurement range of 0-6 V, resulting in a theoretical resolution of approx. 0.4 mV. The drain-source current was set with accuracy better than 1 µA. The acquisition rate for all measurements was 10 Hz.

The FET was operated in a constant current mode, i.e.  $I_D = \text{const.}$ , with a typical value of 45 µA. The resulting voltage drop,  $V_{DS}$ , was recorded as the sensor signal. The gate and the substrate contact of the transistor were grounded. However, in general the gate bias can be used to adjust the baseline and, in addition, cycling of the gate bias (GBCO, gate bias cycled operation) can be used to further improve the selectivity of the transistor [27].

The operating temperature of the sensor was modulated, i.e., run in temperature cycled operation, TCO, in order to increase the selectivity of the sensor [15], [24]. The chosen

temperature cycle consists of three temperature plateaus, i.e. 180 °C, 200 °C, and 220 °C. Each plateau is 15 s long resulting in a total cycle-duration of 45 s as shown in Fig. 2a. After the last temperature step, the cycle starts over again so that the sensor is never in steady state. In fact, most of the discriminatory information is contained in the transients. Thus, the chosen cycle is a tradeoff between the number of temperature changes and the cycle duration, here 45 s. We have reported earlier that the chosen cycle is suitable to detect and discriminate typical VOCs [15]. Nevertheless, the temperature cycle is not yet optimized for the given application but since it led to reasonably good results in earlier studies, it is kept for developing an evaluation strategy in this work. Note that the cycle optimization can be performed by comparing the achieved classification and quantification performance for different cycles.

### 3. Signal Processing

The section deals with the applied signal- and data processing which are used to discriminate and quantify the studied VOCs. After a pre-processing step, shape describing features were extracted (cf. Section 3.1) in order to use them as input data for multivariate analysis. Canonical Linear Discriminant Analysis, LDA, was mainly used (cf. Section 3.2) for discrimination and quantification of typical VOCs. However, regression algorithms like Partial Least Squares Regression, PLSR, are usually more suitable for quantification, and were therefore considered as well (cf. Section 3.3). Additionally, Partial Least Squares – Discriminant Analysis, PLS-DA, was studied for discrimination since it is simultaneously a projection algorithms and a classifier (cf. Section 3.4).

Dedicated MATLAB programs, which are mainly based on the MATLAB Statistics Toolbox, were developed for all signal processing techniques applied in this study.

#### 3.1. Pre-Processing and Feature Extraction

The aim of the signal pre-processing is to reduce the influence of sensor drift. In temperature cycled operation pre-processing, in particular normalization of the data, is done on a cycle-by-cycle basis. It was reported that setting the cycle mean to one is an effective method to reduce multiplicative drift [24]. However, it should be noted that when setting the mean of each cycle to one, information regarding discrimination of different gases which is contained in the raw data close to the mean of the cycle, is discarded.

In general, the influence of temperature on the sensor signal is much higher than the response due to gas interaction. In order to highlight the gas induced changes, the difference signal can be used instead of the raw signal. The difference signal is calculated in a way that from each cycle a reference cycle is subtracted:  $\text{difference\_cycle} = \text{cycle}(\text{test\_gas}) - \text{cycle}(\text{reference})$ . Typically, the reference cycle is an average cycle in carrier gas, i.e., pure synthetic air. In addition, the normalized difference signal is derived by first normalizing each cycle, i.e. setting the cycle mean to one, followed by the subtraction of the reference cycle. By doing so, the actual sensor response is highlighted while the influence of the changing operating temperature is suppressed. Fig. 2b exemplarily shows the (normalized) difference cycles for 100 ppb formaldehyde, 20 ppb naphthalene, and 4.5 ppb benzene, respectively, under humid conditions, i.e., 40 % relative humidity, r.h., at 20 °C. Evidently, the sensor response changes significantly for the chosen gas exposures in different parts of the temperature cycle.

Evaluation of the obtained multi-dimensional sensor raw data, with each individual cycle consisting of 450 data points, is performed by pattern recognition tools. In order to reduce the dimension of the input data, feature extraction is applied. Especially, features describing the shape of the sensor response, e.g., mean values and slopes (here: line of best fit), are very effective for discrimination [24], [27], [28]. In fact, it was reported that these features outperform Fourier and Wavelet transformation coefficients which are typically extracted from periodic signals [29]. In Fig. 2b, nine intervals are marked from which features are extracted. The number of intervals was derived by looking at the difference signal, cf. Fig. 2b. In general, more features lead to better results but also an increase of the dimension of the problem again. Especially for supervised learning algorithms, like LDA, it is important to keep the dimension of the input data low, i.e. the ratio between the number of observations (here: cycles) and number of predictors (here: features) high, in order to avoid over-fitting.

### 3.2. Linear Discriminant Analysis

The main goal of pattern recognition is the reduction of dimensionality. Feature extraction reduces the dimension considerably, here 18 features compared to 450 data points per cycle, but a dimension of 18 is still rather high. In a next step, the extracted features are used as input data for pattern recognition. One of the most well-known algorithms to evaluate multidimensional data is the (canonical) Linear Discriminant Analysis, LDA, [30], [31]. LDA is a supervised learning method meaning that the correct classification is known for each object (here: cycle). LDA determines linear combinations, called discriminant functions, of the extracted features and the LDA coefficients are calculated by maximizing the ratio (also

known as the discrimination criterion) of the between-class scatter, i.e., sum of the squared distances between various classes, and the within-class scatter, i.e., sum of squared distances of the group elements from the class mean, in the feature space. This optimization task is solved by an eigenvalue problem and the corresponding eigenvectors are the coefficients in the LDA model, i.e., the coefficients in the linear combinations of the extracted features.

The dimension of the space spanned by the discriminant functions, DFs, is  $K = \min\{n - 1; m\}$  with  $n$  number of groups and  $m$  number of features. Usually,  $m$  is much larger than  $n$  so that  $n-1$  discriminant functions are determined achieving a dimensionality reduction. Typically, the first two discriminant functions represent more than 90 % of the total information contained in the data set. Thus, the higher dimensions can be neglected.

However, (canonical) LDA just projects the data into a new space but does not classify them. Therefore, a separate classifier needs to be applied beyond discrimination. In this work, the Mahalanobis distance classifier is used which assigns new data points to the group with the shortest Mahalanobis distance [31].

Compared to the Euclidean distance, the Mahalanobis distance also takes the direction of the largest variance of the group into account. Thus, the classification is usually better and more stable compared to the simple Euclidean distance.

Since LDA is a supervised method it has a tendency to over-fit the data, i.e. suggesting better group separability. Therefore, the discrimination, i.e., result of the LDA, needs to be validated. Cross-validation is a suitable way to proof the obtained results. There are several different algorithms to validate the results [30]. One of the most well-known validation algorithms is leave-one-out cross validation, LOOCV. The basic idea is to take one data point (here: cycle) out of the data set, calculate the LDA with the rest of the data, and then classify the held-out point based on the newly calculated LDA model. The Mahalanobis distance was again used for classification.

In order to study the robustness, i.e., stability, of the determined model, it can be validated by using a second data set. The procedure is similar to cross-validation but instead of taking data points out of the training set, a data set from a second measurement is used. The obtained LDA coefficients from the training set are then used to project the “unknown” data from the second set, i.e., evaluation data, into the LDA space with the training data. The projected points are then classified and the correct classification rate is a measure of stability / robustness.

### 3.3. Partial Least Squares Regression

Contrary to LDA as a discrimination algorithm, Partial Least Squares Regression, PLSR, builds a linear model for quantification. Hence, this algorithm is usually better suited when it comes to determining the concentration of a known gas.

PLSR was developed by Wold et al. [32] and makes heavy use of the concept of latent variables. This is a sparse set of variables which cannot be observed directly and does not necessarily have a physical meaning, but is assumed to determine the physical process in question. The PLS algorithm constructs such variables as linear combinations of the measured variables, i.e., the features in this case, in a way that they explain most of the variance of the features and, at the same time, have maximum covariance to the response. Multiple Linear Regression, MLR, is then used to regress the latent variables against the response. In this case, the response is the concentration of a given gas, and thus, the resulting linear model will have features as input and return an estimate for the gas concentration.

The number of latent variables, i.e., components, is crucial to the quality of the model. Similarly to LDA, each component captures a certain amount of information of the original data. Too few components will result in an under-fitted model which does not provide a good estimate for the concentration. However, too many components produce over-fitted models, which fit the training data almost perfectly, but have very poor general predicting ability, very similar to fitting data with a polynomial of too high order. Thus, the number of components must be chosen carefully. This is usually done by cross-validating the model and determining the Prediction Error Sum of Squares, PRESS, which is kind of the Mean Squares Error, MSE, of the projected, “unknown” data in each iteration of the cross-validation. The number of components for which the PRESS has a minimum will give the best model [33].

### 3.4. Partial Least Squares – Discriminant Analysis

Partial Least Squares – Discriminant Analysis, PLS-DA, is a discrimination and classification algorithm based on PLSR [34]. Instead of a numerical vector, the response is a categorical matrix. It has as many rows as the number of data points (here: cycles) and as many columns as the number of classes, e.g., different gases. In each column, “1” indicates class membership of the data point, “0” non-class-membership. If the response is multivariate, PLS will also construct latent variables in exactly the same way as described above in Section 3.3, and then regress both sets of latent variables against each other.

This process gives as a result a number of sets of regression coefficients equal to the number of classes. Each set will project the data in a way that one class will be located near 1, while

all other classes will be located near 0. Hence, a discrimination threshold can be set around 0.5, above which unknown data will be classified as belonging to a specific class. However, this threshold can be set independently. In this work, normal distribution fits of the target class “1” and of all other classes “0” are done and the threshold value is set based on their intersection.

#### 4. Results and Discussion

First, results regarding the discrimination of VOCs when using Linear Discriminant Analysis are presented and discussed (cf. Section 4.1). In a next step, the same data are evaluated using Partial Least Squares – Discriminant Analysis (cf. Section 4.2). After discrimination, the detected VOC needs to be quantified. Different concentrations of naphthalene are discriminated by using Linear Discriminant Analysis (cf. Section 4.3) and quantified by Partial Least Squares Regression (cf. Section 4.4).

##### 4.1. Discrimination of VOCs using Linear Discriminant Analysis

For on demand ventilation control which takes the kind of air pollution into account, it is important to detect the pollutant, i.e., to discriminate between the main VOCs of interest. It was reported earlier that discrimination is possible if the LDA model is trained just with one concentration per gas [15]. However, in real applications the sensor response is affected by changing humidity and also single concentrations are not feasible. In order to enhance the robustness of the discrimination, it is important to include as much information as possible in the training phase. One way is to combine measurements at different humidity levels to make the discrimination nearly independent of background humidity [35]. For the discrimination of several gases, the training data set should also contain various concentrations of the target gas. The LDA treats these different concentrations as one group, and thus, the stability of the discrimination is increased although, the quality of the separation decreases [35]. Fig. 3a shows the LDA scatter plot discriminating benzene, naphthalene, and formaldehyde as well as a group which represents “clean” air. Each test gas group consists of three different concentrations, all of them above or close to the legal threshold. Since the aim is to control on demand ventilation, the system should be able to distinguish between the two cases “no ventilation needed”, i.e., the concentrations of the pollutants are below their thresholds, and

“ventilation needed”, i.e., the concentration of at least one of the VOCs exceeds the threshold. Thus, the fourth group should not just contain pure synthetic air but rather concentrations of the target gases which are below the threshold. In Fig. 3a the “below threshold” group contains besides pure synthetic air also 50 ppb formaldehyde, 2.5 ppb naphthalene, and 1.5 ppb benzene at two humidity levels, i.e., 20 % and 40 % r.h. at 20 °C. Although, the scatter seems high and the groups overlap a bit, a correct cross-validation rate of almost 90 % was achieved when using LOOCV with a Mahalanobis distance classifier. Again, the shown discrimination is very robust against changing humidity and also varying concentration of the target VOCs. Since (canonical) LDA does only projection, a separate classifier is needed. Here, the Mahalanobis distance was used and the resulting territorial plot showing how “unknown” points would be classified is given in Fig. 3b.

#### 4.2. Discrimination of VOCs using Partial Least Squares – Discriminant Analysis

Besides Linear Discriminant Analysis which achieves good results regarding the discrimination, Partial Least Squares – Discriminant Analysis was additionally studied since the algorithm itself does classification. Fig. 4 shows the result of the PLS-DA when 10 components are used for the same data set which was already used for the discrimination shown in Fig. 3a. The number of components was chosen based on the minimum misclassification rate of the different groups. PLS-DA performs one-vs-all classification and since the given problem contains four groups / classes, the result of PLS-DA is four different scatter plots. On the top left panel in Fig. 4 formaldehyde is separated from the other classes. It should be noted that the other classes are treated as one single class by the algorithm and are only colored in order to show how well they are separated from the target class. Also in this case, each class contains several concentrations at two different levels of relative humidity. A correct classification rate of 76 % was achieved for formaldehyde, i.e., 76 % of all the formaldehyde training data was correctly classified. Naphthalene is classified correctly to 99 % and benzene up to 90 %. The overall-classification rate is 83.4 % which is lower than that obtained using LDA with a Mahalanobis distance classifier. However, as mentioned in Section 3.4, the classification threshold in PLS-DA can be set independently so that the level of misclassification can be chosen. For on demand ventilation, this advantage could be used to adjust the decision process for concentrations of a target gas close to the threshold. This kind of adjustment of the classifier is not as easily possible for LDA.

#### 4.3. Quantification of Naphthalene by Using Linear Discriminant Analysis

After identification of a target VOC, quantification is required as basis for determining the ventilation strategy. Fig. 5 shows the discrimination of various concentrations of naphthalene when data sets at two different humidity levels are combined. As shown in the figure, discrimination is easily possible and the determination of the concentration can be done using the value of the first discriminant function (see also the histogram below the scatter plot) which holds almost 99 % of all the discriminatory information.

In a next step, the LDA model was trained with only three concentrations, i.e., 0 ppb, 5 ppb, and 40 ppb naphthalene, cf. Fig. 6a. The LDA coefficients obtained from the training were then used to project “unknown” data points in the same plot, i.e., 2.5 ppb, 10 ppb, and 20 ppb naphthalene (open symbols). As can be seen in the figure, the projected concentrations fit perfectly to the trained LDA model. In Fig. 6b the dependency of the concentration on the value of the first discriminant function, DF1, is highlighted by plotting the concentration versus their DF1 value. The centroids of the training groups are fitted by a second order polynomial and the “unknown” groups are projected into the same plot. They are located close to the fit-curve so that a prediction of unknown concentrations using the given equation is possible. In principle, the uncertainty of the quantification is represented by the scatter of the groups along DF1 together with the distance of the centroids to the fit-curve.

A threshold for control of a ventilation system could be set by using the 5 ppb training group which represents nearly the legal limit, together with its scatter. In the shown example, the threshold is set to a DF1 value of 2. Projected data points with a smaller DF1 value would be classified as “no ventilation needed”, and for points with a larger DF1 value the ventilation should be switched on.

#### 4.4. Quantification of Naphthalene using Partial Least Squares Regression

Linear Discriminant Analysis is typically an algorithm for discrimination of different gases. Hence, quantification can only be done by using discrete concentrations as training and unknown concentrations can only be determined by a fit-function as described in Section 4.3.

For quantification purposes usually regression algorithms, e.g., Partial Least Squares Regression, PLSR, are used which allow a continuous determination of the concentration.

Since nine intervals for feature extraction were defined (cf. Fig. 2) in which the slope and the mean value are calculated, 18 components can be used in the PLSR model. However, the dimension should be kept as low as possible in order to limit the required computational power and to avoid over-fitting. Thus, in a first step the PLSR model is built with all 18 components and the Prediction Error Sum of Squares, PRESS, is calculated. Then the number of components is reduced step-by-step to determine where the PRESS curve has its minimum. In the given example, only ten components are sufficient to build the model. Over-fitting is

not an issue in this example since the number of cycles is much larger than the number of features. Fig. 7 shows the result of PLSR when three concentrations of naphthalene, i.e., 0 ppb, 5 ppb, and 40 ppb at 20 % and 40 % r.h., are used for building the model (blue solid symbols). The standard deviation ( $3\sigma_{\max}$ ) for 5 ppb and 40 ppb is 7.2 ppb and 9.0 ppb, respectively. After that “unknown” concentrations, i.e., 2.5 ppb, 10 ppb, and 20 ppb (red open symbols), are projected using the built regression model. In general, the predicted response, i.e., concentration, is higher than the actual response. Nevertheless, the quality is similar to the one which was achieved using LDA with a second order polynomial fit. The accuracy of the model can of course be enhanced when more concentrations are used for building the model as well as when the spanned concentration range is reduced. Alternatively, a hierarchical approach with two different regression models can be used. For on demand ventilation systems it might be of interest to determine only the concentration of target VOCs near the threshold. However, with every additional training-concentration, the effort for preparing the training set increases considerably. For a rough estimation of the VOC concentration which is sufficient for ventilation control, the suggested model based on three concentrations for the training is sufficient.

## 5. Conclusion and Outlook

In this work, gas sensitive platinum gate silicon carbide field effect transistors, SiC-FETs, in temperature cycled operation, TCO, were studied as suitable candidates for indoor air quality applications. It was demonstrated that discrimination of three hazardous VOCs of high interest, i.e., formaldehyde, benzene, and naphthalene, independent of the humidity level is possible. When using Linear Discriminant Analysis, LDA, a correct cross-validation rate of almost 90 % was achieved. The classification rate when using Partial Least Squares – Discriminant Analysis, PLS-DA, was 83 %. Although, the classification rate is lower than that obtained with LDA, PLS-DA offers the advantage to adjust the classifier to the given problem, and to the individual gases to be detected. Hence, PLS-DA is a suitable algorithm for evaluating multidimensional data for IAQ applications. Quantification of naphthalene was done by using a combination of LDA and a second order polynomial fit. The accuracy for the training was approximately 2.5 ppb and 8 ppb for 5 ppb and 40 ppb naphthalene, respectively. However, for a continuous prediction of the concentration regression algorithms like Partial Least Squares Regression, PLSR, are more suitable. For the given problem, the accuracy ( $3\sigma_{\max}$ ) was for all training groups between 6.5 ppb and 9 ppb.

Temperature cycled operation is a suitable approach to increase the selectivity of SiC-FETs. The robustness of the discrimination was enhanced considerably by using extended training sets, i.e., combining different humidity levels and concentrations.

Future work will address the optimization of the used temperature cycle and the pre-processing in order to further boost the selectivity. Additionally, we will focus on linearization methods (as suggested in [36]) to enhance the prediction performance of the quantification, especially using PLSR. Besides the detection of individual VOCs in pure synthetic air, future work will address discrimination and quantification of VOCs in a high background of interfering gases, e.g. alcohols or other VOCs, as well as mixtures of the target VOCs. Very preliminary results indicate that detection of VOCs in the low ppb range in a background of ethanol in the sub-ppm and low ppm range, i.e., up to three orders of magnitude higher than the target gas concentration, is possible but requires a hierarchical approach [35]. The selective detection of VOCs in a mixture of other VOCs is challenging. Thus, future work within the project will also address the development of new sensing materials and the use of selective pre-concentration in addition to advanced signal processing. The final sensor systems might consist of a small array of a few sensors but all of them run in TCO.

## Acknowledgment

The authors would like to thank SenSiC AB, Kista, Sweden, for providing the sensors, and 3S - Sensors, Signal Processing, Systems GmbH, Saarbrücken, Germany, for providing the hardware for sensor operation and read-out.

The research leading to these results has received funding from the European Seventh Framework Programme ([FP7/2007-2013]) under grant agreement number 604311 within the EU project SENSIndoor.

## Biographies

**Christian Bur** received his diploma in Mechatronics, specifically Microtechnology and sensor science, from Saarland University, Saarbrücken, Germany, in 2011. He is currently enrolled in the European doctoral program DocMASE presuming a double Ph.D. between Saarland University and Linköping University, Linköping, Sweden. His main research interests are in the field of field effect based gas sensors run in dynamic operation, e.g. temperature modulation and gate bias modulation, and appropriate signal processing based on multivariate statistics.

**Manuel Bastuck** received his M.Sc. in Microtechnology and Nanostructures from Saarland University, Saarbrücken, Germany, in 2014. He is currently a PhD student at the Lab for Measurement Technology at Saarland University. His main research interests are in the fields of dynamic gas sensor operation, especially of gas-sensitive field effect transistors, and subsequent signal processing based on multivariate statistics.

**Donatella Puglisi** graduated in Physics at University of Catania, Italy, in 2005, and pursued her Ph.D. degree in Physics from the same Institution in 2009. She worked for three years as a postdoctoral research fellow at Politecnico di Milano, Como Campus, Italy, and for two years as a post-doc at Linköping University, Sweden, in the research division Applied Sensor Science. At present, she is an assistant professor at Applied Sensor Science. Her main research activities include sensor processing, characterization, and development of gas

sensitive field effect transistors based on silicon carbide for detection of trace amounts of hazardous indoor air pollutants.

**Andreas Schütze** studied Physics and received his doctorate in Applied Physics from Justus-Liebig-Universität in Gießen in 1994 with a thesis on micro gas sensor systems. After some years in industry, he joined the University of Applied Sciences in Krefeld, Germany, as professor for Microsystems Technology from 1998 to 2000. Since 2000 he is a full professor for measurement science and technology in the Department of Mechatronics at Saarland University, Saarbrücken, Germany. His research interests include microsensors and microsystems, especially advanced chemical sensor systems, both for gas and liquid phase, for security and control applications.

**Anita Lloyd Spetz**, is Professor in Applied Sensor Science at Linköping University and Finland Distinguished Professor, FiDiPro, at University of Oulu, Finland (50 % 2011–2015). She is Director of the VINN Excellence centre, FunMat at Linköping University and vice chair for the COST network EuNetAir, TD1105. She is member of the board of SenSiC AB. Her research involves SiC-FET high temperature gas sensors with MAX material contacts, transducers for biosensors, sensor systems for small and medium scale bioheaters, resonators, PMand graphene sensors and at University of Oulu portable nanoparticle detectors and an electrical method to detect influence of particles on cells

**Mike Andersson** received his M.Sc. in Engineering from Linköping University, Sweden, in 2002, his Ph.D. in Applied Physics, specifically Sensor Science, in 2007 and was appointed Associate Professor in 2014 at Linköping University. Following two Post Doc years in a project with SKF, Nieuwegein, The Netherlands, he is currently employed as Research Scientist at Linköping University (50 %), University of Oulu, Finland (25 %), and as CTO of SenSiC AB (25 %). His main research interests are in the field of semiconductor based gas sensors for high temperature applications.

## References

- [1] D.A. Sarigiannis, S.P. Karakitsios, A. Gotti, I.L. Liakos, and A. Katsoyiannis, “Exposure to major volatile organic compounds and carbonyls in European indoor environments and associated health risk,” *Environ. Int.*, vol. 37, no. 4, pp. 743–765, 2011. DOI: 10.1016/j.envint.2011.01.005
- [2] D.H. Do, H. Van Langenhove, C. Walgraeve, S.F. Hayleeyesus, P. de Wispelaere, J. Dewulf, K. Demeestere, “Volatile organic compounds in an urban environment: A comparison among Belgium, Vietnam and Ethiopia,” *Int. J. Environ. Anal. Chem.*, vol. 93, no. 3, pp. 298–314, 2013. DOI: 10.1080/03067319.2011.62070
- [3] WHO Guidelines for Indoor Air Quality: Selected Pollutants, World Health Organization, Geneva, Switzerland, 2010.
- [4] W. Kirch, Ed., *Encyclopedia of Public Health*. Berlin, Germany: Springer-Verlag, 2008, p. 1303.
- [5] H. Salonen, A-L. Pasanen, S. Lappalainen, H. Riuttala, T. Tuomi, P. Pasanen, B. Bäck, K. Reijula, “Volatile organic compounds and formaldehyde as explaining factors for sensory irritation in office environments,” *J Occup Environ Hyg*, vol. 6, pp. 239-247, 2009. DOI: 10.1080/15459620902735892
- [6] *Indoor Air Quality Guideline Value Proposals*, AFSSET, Paris, France, 2006.
- [7] K. Koistinen, D. Kotzias, S. Kephelopoulos, C. Schlitt, P. Carrer, M. Jantunen, S. Kirchner, J. McLaughlin, L. Mølhav, E.O. Fernandes, B. Seifert, “The INDEX project: Executive summary of a European Union project on indoor air pollutants,” *Allergy*, vol. 63, no. 7, pp. 810–819, 2008. DOI: 10.1111/j.1398-9995.2008.01740.x
- [8] WHO Air Quality Guidelines for Europe, WHO, Geneva, Switzerland, 2000.
- [9] C.P. Weisel, “Benzene exposure: An overview of monitoring methods and their findings,” *Chem.-Biol. Interact.*, vol. 184, nos. 1–2, pp. 58–66, 2010. DOI: 10.1016/j.cbi.2009.12.030
- [10] *Guideline Values for Indoor Air for Formaldehyde and Benzene*, French Standard 2011-1727, Dec. 2011.
- [11] German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety. (2014, Oct. 28). Indoor air guide values for naphthalene and naphthalene-like compounds. Ad-Hoc Work. Group Indoor Air Guide Values [Online]. Available:

<http://www.umweltbundesamt.de/themen/gesundheit/kommissionen-arbeitsgruppen/ad-hoc-arbeitsgruppe-innenraumrichtwerte>  
DOI: 0.1007/s00103-013-1836-9

- [12] R. Preuss, J. Angerer, H. Drexler, “Naphthalene—An environmental and occupational toxicant,” *Int. Arch. Occupat. Environ. Health*, vol. 76, no. 8, pp. 556–576, 2003. DOI: 10.1007/s00420-003-0458-1
- [13] German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety. (2014, Jan. 7). Guide values for indoor air. Ad-Hoc Work. Group Indoor Air Guide Values [Online].  
<http://www.umweltbundesamt.de/en/topics/health/commissions-working-groups/ad-hoc-working-group-for-indoor-air-guide-values>
- [14] T. Salthammer, S. Mentese, and R. Marutzky, “Formaldehyde in the indoor environment,” *Chem. Rev.*, vol. 110, no. 4, pp. 2536–2572, 2010. DOI: 10.1021/cr800399g
- [15] C. Bur, M. Andersson, A. Lloyd Spetz, A. Schütze, “Detecting volatile organic compounds in the ppb range with gas sensitive platinum gate SiC-field effect transistors,” *IEEE Sensors Journal*, vol. 14, no. 9, pp. 3221–3228, 2014. DOI: 10.1109/JSEN.2014.2326693
- [16] D. Puglisi, J. Eriksson, C. Bur, A. Schütze, A. Lloyd Spetz, and M. Andersson, “Silicon carbide field effect transistors for detection of ultra-low concentrations of hazardous volatile organic compounds,” *Mat. Sci. Forum*, vol. 779–780, pp. 1067–1070, 2014. DOI: 10.4028/www.scientific.net/MSF.778-780.1067
- [17] I. Lundström, H. Sundgren, F. Winqvist, M. Eriksson, C. Krantz-Rülcker, A. LloydSpetz, “Twenty-five years of field effect gas sensor research in Linköping,” *Sens. Actuators B*, vol. 121, pp. 247–262, 2007. DOI: 10.1016/j.snb.2006.09.046.
- [18] A. Lloyd Spetz, S. Svage, *Advances in SiC field effect gas sensors*, in: W.J.Choyke, H. Matsunami, G. Pensl (Eds.), *Silicon Carbide—Recent Major Advances*, Springer Verlag, Berlin, Heidelberg, New York, 2003, ISBN 3540404589.
- [19] M. Andersson, R. Pearce, A. Lloyd Spetz, “New generation SiC based field effect transistor gas sensors,” *Sens. Actuators B*, vol. 179, pp. 95–106, 2013. DOI: 10.1016/j.snb.2012.12.059.
- [20] M. Eriksson, L.-G. Ekedahl, “Hydrogen adsorption states at the Pd/SiO<sub>2</sub> interface and simulation of the response of a Pd metal-oxide-semiconductor hydrogen sensor,” *J. Appl. Phys.*, vol. 83, no. 8, pp. 3947–3951, 1998, DOI: 10.1063/1.367150.

- [21] S. Nakagomi, P. Tobias, A. Baranzahi, I. Lundström, "Influence of carbon monoxide, water and oxygen on high temperature catalytic metal-oxide-siliconcarbide structures," *Sens. Actuators B*, vol. 45, no. 3, pp. 183–191, 1997. DOI: 10.1016/S0925-4005(97)00292-X.
- [22] M. Wallin, H. Grönbeck, A. Lloyd Spetz and M. Skoglundh, "Vibrational study of ammonia adsorption on Pt/SiO<sub>2</sub>," *Applied Surface Science*, vol. 235, no. 4 pp. 487-500, 2004. DOI: 10.1016/j.apsusc.2004.03.225
- [23] J. Schalwig, P. Kreisl, S. Ahlers, G. Müller, "Response mechanism of SiC-based MOS field-effect gas sensors," *IEEE Sens. J.*, vol. 2, no. 5, pp. 394–402, 2002. DOI: 10.1109/JSEN.2002.806214
- [24] C. Bur, P. Reimann, M. Andersson, A. Schütze, A. Lloyd Spetz, "Increasing the selectivity of Pt-Gate SiC field effect gas sensors by dynamic temperature modulation," *IEEE Sens. J.*, vol. 12, no. 6, pp. 1906–1913, 2012. DOI: 10.1109/JSEN.2011.2179645
- [25] A.P. Lee, B.J. Reedy, "Temperature modulation in semiconductor gas sensing," *Sens. Actuators B*, vol. 60, no. 1, pp. 35–42, 1999. DOI: 10.1016/S0925-4005(99)00241-5
- [26] N. Helwig, M. Schüler, C. Bur, A. Schütze, T. Sauerwald, "Gas mixing apparatus for automated gas sensor characterization," *Meas. Sci. Technol.*, vol. 25, no. 5, 2014. DOI:10.1088/0957-0233/25/5/055903
- [27] C. Bur, M. Bastuck, A. Lloyd-Spez, M. Andersson, and A. Schütze, "Selectivity enhancement of SiC-FET gas sensors by combining temperature and gate bias cycled operation using multivariate statistics," *Sens. Actuators B, Chem.*, vol. 193, pp. 931–940, 2014. DOI: 10.1016/j.snb.2013.12.030
- [28] A. Schütze, A. Gramm, T. Rühl, "Identification of organic solvents by a virtual multisensor system with hierarchical classification," *IEEE Sens. J.*, vol. 4, no. 6 pp. 857–886, 2004. DOI: 10.1109/JSEN.2004.833514
- [29] A. Gramm, A. Schütze, "High performance solvent vapor identification with a two sensor array using temperature cycling and pattern recognition," *Sens. Actuators B*, vol. 95, pp. 58–65, 2003. DOI: 10.1016/S0925-4005(03)00404-0
- [30] R. Gutierrez-Osuna, "Pattern analysis for machine olfaction: a review," *IEEE Sens. J.*, vol. 2, no. 3, pp. 189–202, 2002. DOI: 10.1109/JSEN.2002.800688
- [31] W.R. Klecka, "Discriminant analysis Quantitative Applications in the Social Sciences," SAGE University Paper, 1980, ISBN 0803914911.

- [32] S. Wold, M. Sjostrom, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemom. Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, 2001. DOI: 10.1016/S0169-7439(01)00155-1
- [33] S. Wold, "Cross-validatory estimation of the number of components in factor and principal components," *Technometrics*, vol. 20, no. 4, pp. 397–405, 1978. DOI: 10.2307/1267639
- [34] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemom.*, vol. 17, no. 3, pp. 166–173, Mar. 2003. DOI: 10.1002/cem.785
- [35] C. Bur, M. Bastuck, D. Puglisi, A. Schütze, A. Lloyd Spetz, M. Andersson, „Discrimination and quantification of volatile organic compounds in the ppb-range with gas sensitive SiC-field effect transistors," *Proc. Eng.*, vol. 87, pp. 604-607, 2014. DOI: 10.1016/j.proeng.2014.11.561
- [36] G. Horner and C. Hierold, "Gas analysis by partial model building," *Sens. Actuators B*, vol. 2, pp. 173-184, 1990.

## Figure Captions

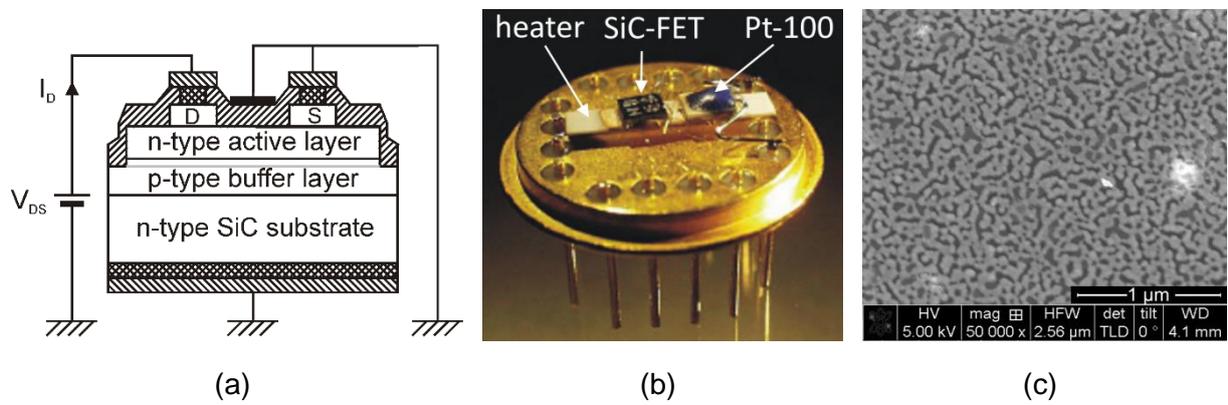


Fig. 1: (a) Schematic cross-sectional view of the used depletion type SiC-FET, (b) photo of the SiC-FET glued on a ceramic heater together with a Pt-100 temperature sensor and mounted on a 16 pin TO-8 header, and (c) SEM picture of the catalytic platinum gate.

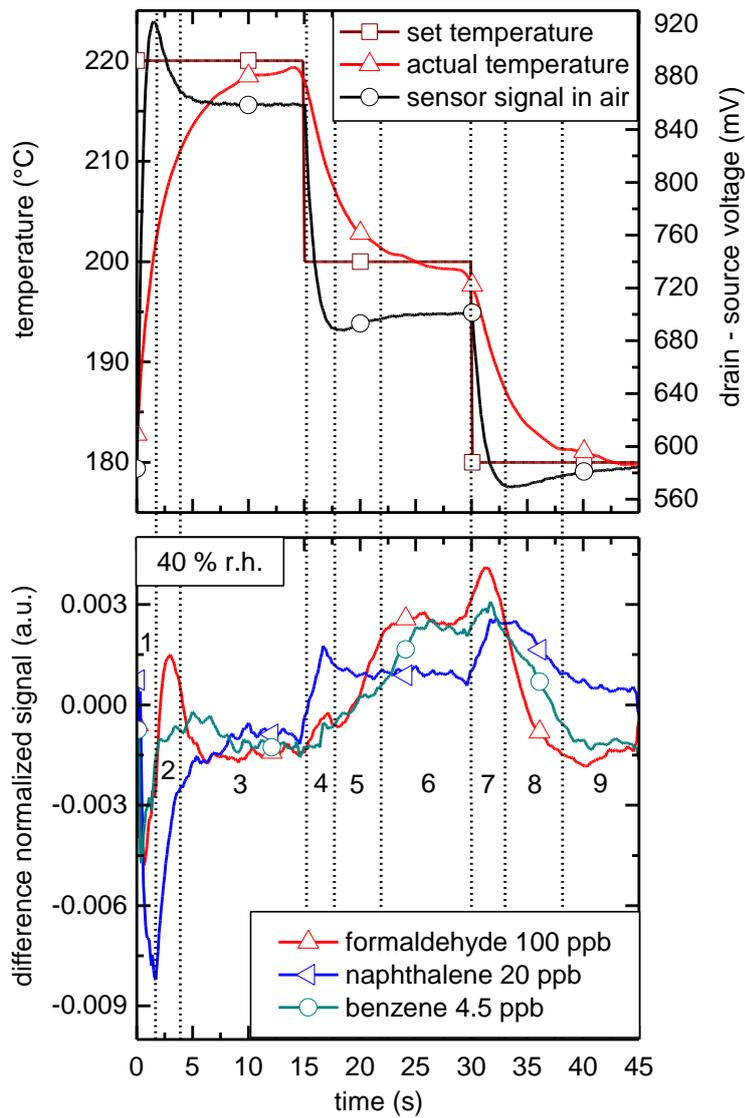


Fig. 2: (a) Applied temperature cycle, actual temperature measured with the additional Pt-100 temperature sensor, and sensor signal in pure air, (b) normalized difference signal of the sensor response for 100 ppb formaldehyde, 20 ppb naphthalene, and 4.5 ppb benzene with 40 % r.h. at 20 °C. Nine intervals for feature extraction are marked.

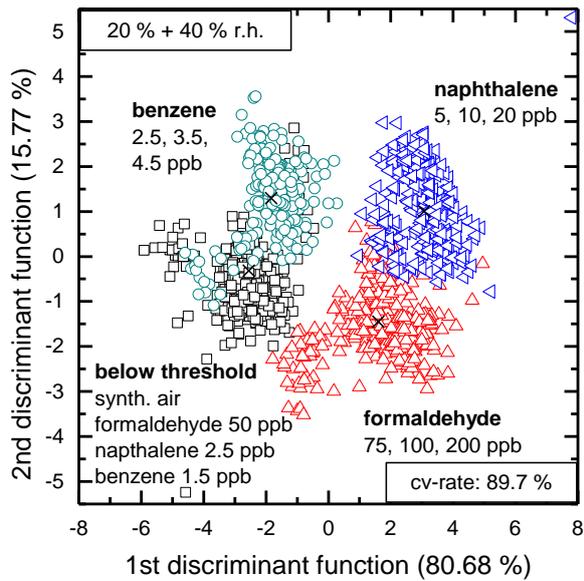


Fig. 3a: LDA scatter plot showing the discrimination of benzene, naphthalene, and formaldehyde, each group contains three different concentrations above the corresponding legal threshold value, and a fourth group containing pure synthetic air as well as one concentration of each gas which is below the legal threshold. All groups contain data at two different humidity levels, i.e. 20 % and 40 % r.h. at 20 °C. The leave-one-out cross-validation rate, cv-rate, using a Mahalanobis distance classifier is almost 90 %.

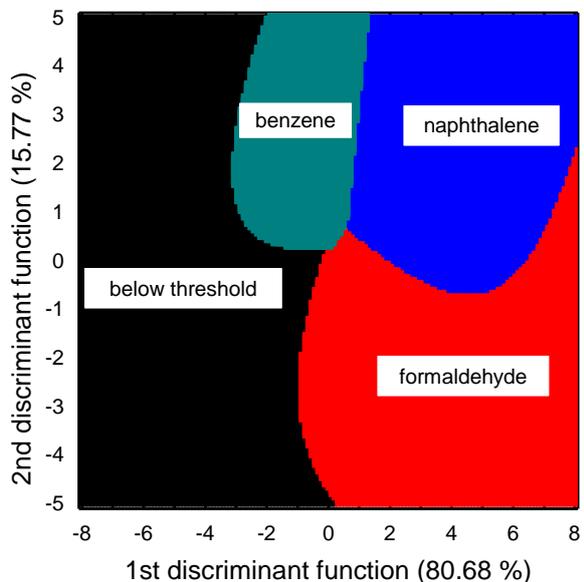


Fig. 3b: Territorial plot of the discrimination using a Mahalanobis distance classifier.

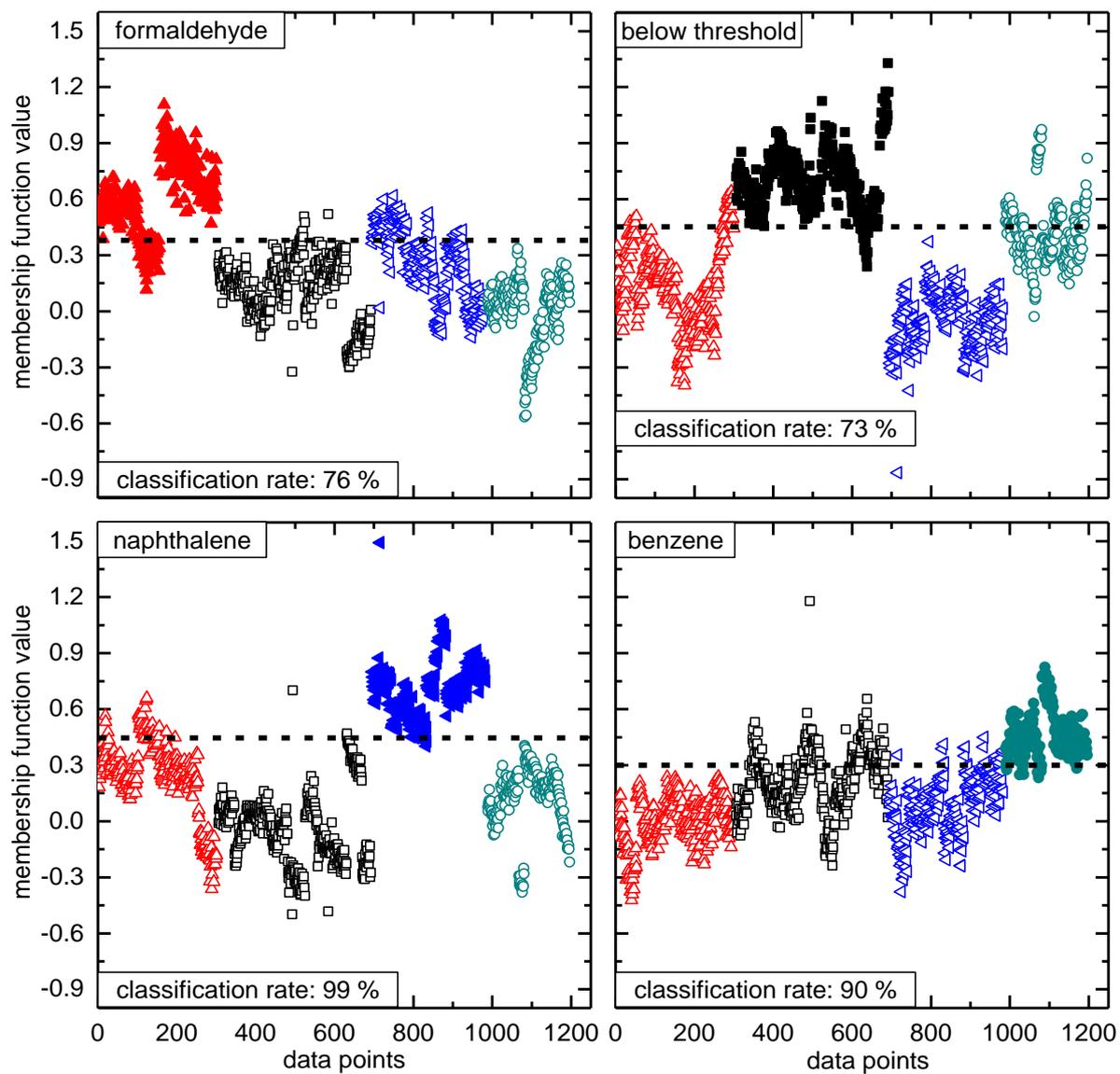


Fig. 4: Scatter plots using Partial Least Squares – Discriminant Analysis, PLS-DA, of the same data as in Fig. 3a. The class which is to be classified is marked with solid symbols. The classes marked with open-symbols are treated as one single class by the algorithm.

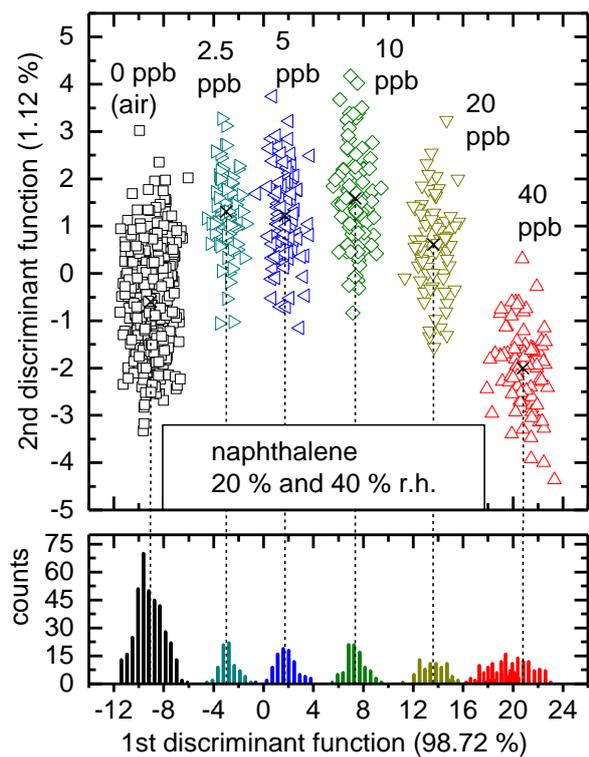


Fig. 5: LDA scatter plot showing the discrimination of several naphthalene concentrations at two different humidity levels, i.e. 20 % and 40 % r.h. at 20 °C.

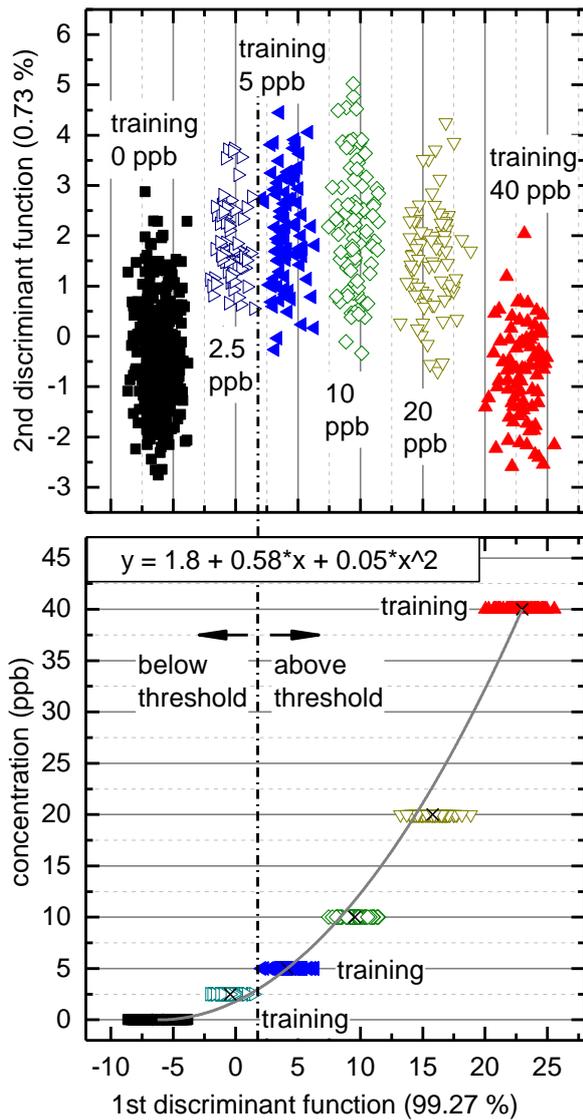


Fig. 6: (a) LDA scatter plot showing the discrimination of 0 ppb, 5 ppb, and 40 ppb naphthalene at two humidity levels, i.e. the training data, as well as the projection of 2.5 ppb, 10 ppb, and 20 ppb, i.e. the evaluation data. (b) Plot concentration vs. value of first discriminant function with second order fit of the training data. The projected data is marked by open-symbols. A cutting line is given which can be used as a threshold e.g. for ventilation control.

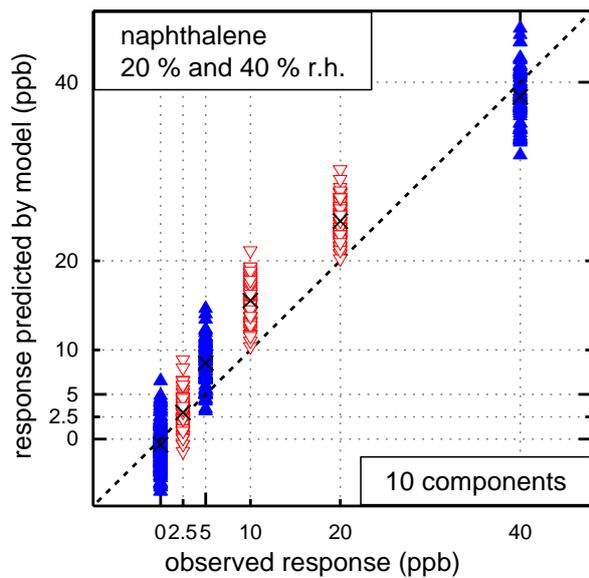


Fig. 7: Result of Partial Least Squares Regression, PLSR. Data which were used to build the model are marked by blue solid symbols while projected data are marked by red open symbols.