SAW sensor's frequency shift characterization for odor recognition and concentration estimation

Olivier Hotel, Jean-Philippe Poli, Christine Mer-Calfati, Emmanuel Scorsone, Samuel Saada

SAW sensor's frequency shift characterization for odour recognition and concentration estimation

Olivier Hotel, Jean-Philippe Poli, Christine Mer-Calfati,
Emmanuel Scorsone and Samuel Saada
CEA LIST, 91191 Gif-sur-Yvette, France.

April 5, 2022

Abstract

In this paper, we propose an approach to determine the time constants and the amplitudes of the mass loading effect and of the viscoelastic contribution of SAW sensor's frequency shift. This approach consists in optimizing a function of these parameters which is independent of the concentration profile. We experimentally establish in laboratory conditions (T=22°C), on a dataset composed of seven different gases, that these features are suitable for chemical compounds identification. In particular, we obtain a higher classification rate than the traditional amplitudes of the signals during the steady state and we show that the classification success rate can be increased by using both of them in conjunction with a feature subset selection heuristic. We also propose a method based on deconvolution and kernel regression to estimate the temporal concentration profile.
Keywords: SAW sensors, odour recognition, concentration evaluation.

1 Introduction

Electronic nose based systems have been successfully used in many applications: quality control in the food and cosmetics industries, detection of disease-specific odours, detection of pollutants and toxic gases [1, 2]. An electronic nose is an instrument composed of an array of gas sensors with different selectivities and a multivariate data processing system capable of recognizing odours. In this paper, we focus on a category of gas sensors called surface acoustic waves (SAW) sensors. These sensors are based on the propagation of mechanical waves produced by piezoelectric materials along a layer composed of a substrate covered by chemically interactive materials. Volatile compounds are absorbed onto the surface of the sensitive material, changing its properties and yielding to a measurable frequency shift of the mechanical waves. It was established that the frequency shift is the superposition of two main contributions [3]: a viscoelastic contribution and of a mass loading effect. Since it has been shown that the

electro-acoustic contribution has an order of magnitude of a mHz whereas the other contributions have ones over a kHz [3, 4, 5], we neglected it in the present study. These contributions can be modelled by first order linear differential equations [3]:

$$\begin{cases} \tau_m \frac{\partial F_m}{\partial t} + F_m = K_m c \\ \tau_v \frac{\partial F_v}{\partial t} + F_v = K_v c \end{cases} \tag{1}$$

where F_m and F_v are respectively the frequency shift due to the mass loading effect and to the viscoelastic contribution, K_m and K_v are respectively their gains, τ_m and τ_v are their time constants and c is the concentration profile of the volatile compound. The total frequency shift is then given by $F = F_m + F_v$ [3]. Moreover, it was established that the mass loading effect involves a negative frequency shift i.e. $K_m < 0$ whereas the viscoelastic contribution may involve a positive or a negative frequency shift, it means that the overall frequency shift can be either positive or negative as shown in Figure 4. In this paper, we propose an approach to determine the gains and the time constants of the mass loading effect and of the viscoelastic contribution. We also show experimentally, on a dataset composed of 7 analytes, that these parameters are suitable features for chemical compound identification and concentration evaluation. This work was motivated by the fact that these features are independent of the concentration profile c and belong to a higher dimensional space than the traditional features (steady state amplitude of the response and rise time) used for compound identification and hence should carry more information about them.

2 Problem formulation

In this paper, we address the problem of determining the parameters of the sensors in a blind way, i.e. without any knowledge of the concentration profile. The discretization of the differential equations 1 using the backward difference operator

$$\frac{\partial F}{\partial t}(t) \approx \frac{F[n] - F[n-1]}{T_s}$$

gives for $i \in \{m, v\}$:

$$F_i[n] = \frac{\tau_i}{\tau_i + T_s} F_i[n-1] + K_i \frac{T_s}{\tau_i + T_s} c[n] \tag{2}$$

where T_s is the sampling rate. Such equations have two interesting properties:

1. they are linear [6] i.e. for $i \in \{m, v\}$ if

$$F_i'[n] = \frac{\tau_i}{\tau_i + T_s} F_i'[n-1] + K_i \frac{T_s}{\tau_i + T_s} c'[n]$$

and

$$F_i''[n] = \frac{\tau_i}{\tau_i + T_s} F_i''[n-1] + K_i \frac{T_s}{\tau_i + T_s} c''[n]$$

then the frequency shift for an input $c' + c''$ is

$$F_i'[n] + F_i''[n] = \frac{\tau_i}{\tau_i + T_s}(F_i'[n-1] + F_i''[n-1])$$

$$+K_i \frac{T_s}{\tau_i + T_s}(c'[n] + c''[n]); \text{ and}$$

2. they are time invariant [6] since we have for $i \in \{m, v\}$

$$F_i[n - n_0] = \frac{\tau_i}{\tau_i + T_s} F_i[n - n_0 - 1] + K_i \frac{T_s}{\tau_i + T_s} c[n - n_0].$$

If we note M_i the equation that maps c to F_i (see Eq. 2), then we have for $i \in \{m, v\}$:

$$F_i[n] = M_i\Big(\sum_{k=-\infty}^{+\infty} c[k]\delta[n-k]\Big),$$

where δ is the Dirac delta function:

$$\delta[n] = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{otherwise} \end{cases}.$$

The time invariance property yields to

$$F_i[n] = \sum_{k=-\infty}^{+\infty} c[k]h_i[n-k] = c[n] * h_i[n] \tag{3}$$

where h_i is the impulse response of the considered contribution. Hence, the mass loading effect and the viscoelastic contribution are entirely determined by their respective impulse response and so is the SAW sensor's response:

$$F = F_m + F_v = h_m * c + h_v * c = (h_m + h_v) * c. \tag{4}$$

2.1 Impulse response computation

The impulse response of a dynamic system is its output when its input is the Dirac delta function. Thus, the impulse responses are given for $i \in \{m, v\}$ by

$$h_i[n] = \frac{\tau_i}{\tau_i + T_s} h_i[n-1] + K_i \frac{T_s}{\tau_i + T_s}\delta[n], \quad i \in \{m, v\}$$

or equivalently by

$$\begin{cases} h_i[0] = K_i \frac{T_s}{\tau_i + T_s} \\ h_i[n] = \frac{\tau_i}{\tau_i + T_s} h_i[n-1] \quad \forall n \geq 0. \end{cases}$$

It is the traditional expression of a geometric series, so we have

$$h_i[n] = K_i \frac{T_s}{\tau_i + T_s}\left(\frac{\tau_i}{\tau_i + T_s}\right)^n, \quad i \in \{m, v\}.$$

Let's define the new variables for $i \in \{m, v\}$:

$$T_i = \frac{\tau_i}{\tau_i + T_s} \quad \text{or} \quad \tau_i = \frac{T_i T_s}{1 - T_i} \quad \text{and} \tag{5}$$

$$A_i = K_i(1 - T_i) \quad \text{or} \quad K_i = \frac{A_i}{1 - T_i}; \tag{6}$$

and let's substitute the impulse responses into Eq 3:

$$F[n] = \left(A_m {T_m}^n + A_v {T_v}^n\right) * c[n].$$

This relation shows that the parameters A_m and A_v can be identified up to a multiplicative constant $\alpha \neq 0$ since

$$F[n] = \left(\frac{A_m}{\alpha} {T_m}^n + \frac{A_v}{\alpha} {T_v}^n\right) * (\alpha c[n]).$$

So, without loss of generality, we can assume that the amplitude of the concentration profile c^* is unitary

$$c^* = \frac{c}{c_{SS}} \quad \text{or} \quad c = c_{SS} c^*, \tag{7}$$

where c_{SS} is the steady state value of the concentration profile. It involves

$$F[n] = \left(A_m {T_m}^n + A_v {T_v}^n\right) * c^*[n]; \quad \text{and}$$

$$\frac{A_m}{1 - T_m} + \frac{A_v}{1 - T_v} = F_{SS},$$

F_{SS} is the steady state value of F. As the presence of the convolution product in the equations makes any further development arduous, we propose to transform them using a generating function.

2.2 Generating functions

The generating function of a sequence $a_0, a_1, ..., a_n$ is

$$G(a[n], x) = \sum_{n \geq 0} a[n] x^n. \tag{8}$$

The main advantage of generating functions is their property to transform a convolution product into a scalar product [7]:

$$G(a[n] * b[n], x) = G(a[n], x) G(b[n], x).$$

Generating functions are defined only for the x where the sum (Eq. 8) converges. The region of convergence (ROC) of a generating function is the set

$$\text{ROC} = \left\{ x \ : \ \sum_{n \geq 0} a[n] x^n \text{ converges} \right\}.$$

2.2.1 Generating function associated with c^*

The generating function associated with the concentration profile can not be computed since we have no information about c^*. However, to determine the region of convergence of $G(c^*[n], x)$ one can use the fact that $c^*[n]$ is a bounded quantity: $0 \leq c^*[n] \leq 1$. Without any other information about the concentration profile c^*, it is not possible to determine the ROC of its associated generating function. However, since

$$|c^*[n]x^n| \leq |c^*_{max}x^n|$$

and the series

$$\sum_{n\geq 0} c^*_{max}x^n = c^*_{max}\sum_{n\geq 0} x^n$$

is absolutely convergent on the set $\{x : |x| < 1\}$ the direct comparison test theorem allows us to state that

$$\{x : |x| < 1\} \subset ROC(G(c^*[n], x)). \tag{9}$$

2.2.2 Generating function associated with h

The generating function associated with the impulse response is

$$\begin{aligned} G(h[n], x) &= \sum_{n\geq 0}(A_m T_m^n + A_v T_v^n)x^n \\ &= A_m \sum_{n\geq 0}(T_m x)^n + A_v \sum_{n\geq 0}(T_v x)^n \\ &= \frac{A_m}{1 - T_m x} + \frac{A_v}{1 - T_v x} \end{aligned} \tag{10}$$

and its region of convergence is

$$ROC(G(h[n], x)) = \left\{ x : |x| < min\left(\frac{1}{T_m}, \frac{1}{T_v}\right)\right\}. \tag{11}$$

2.2.3 Generating function associated with F

The generating function associated with the sensor's response F is

$$G(F[n], x) = G(c^*[n] * h[n], x) = G(c^*[n], x)G(h[n], x)$$

and its region of convergence is

$$ROC(G(F[n], x)) = ROC(G(c^*[n], x)) \cap ROC(G(h[n], x)).$$

Eq. 5, Eq. 9 and Eq. 11 allow us to conclude that

$$\{x : |x| < 1\} \subset ROC(G(F[n], x)). \tag{12}$$

In practice, the generating function of the digitalized signal F of length N can be approximated with high precision using Eq. 8 for $|x| < 10^{\frac{\epsilon}{N}}$ where ϵ is the machine precision if x belongs to the ROC of F. Figure 1 shows an example of a SAW sensor's response and of the consecutive differences between two successive terms of its associated generating function:

$$G(h[n], x) - G(h[n-1], x).$$

It illustrates that for $|x| < 10^{\frac{\epsilon}{N}}$ the successive differences tend to 0 (the estimate of $G(F[n], x)$ is accurate) while they are important when x is over this bound.

2.3 Optimization problem formulation

Since SAW sensor based electronic noses are composed of an array of N_s sensors, they can be modelled as a single input multiple output system. The e-nose is driven by a single input sequence $c[n]$ and yields to N_s output sequences $F_i[n]$. Computing the associated generating functions yields to the set of equations $i = 1 \dots N_s$:

$$G(F_i[n], x) = \left(\frac{A_{i,m}}{1 - T_{i,m}x} + \frac{A_{i,v}}{1 - T_{i,v}x} \right) G(c^*[n], x). \tag{13}$$

One can determine $G(c^*[n], x)$ using the equation associated with the first sensor $(i = 1)$

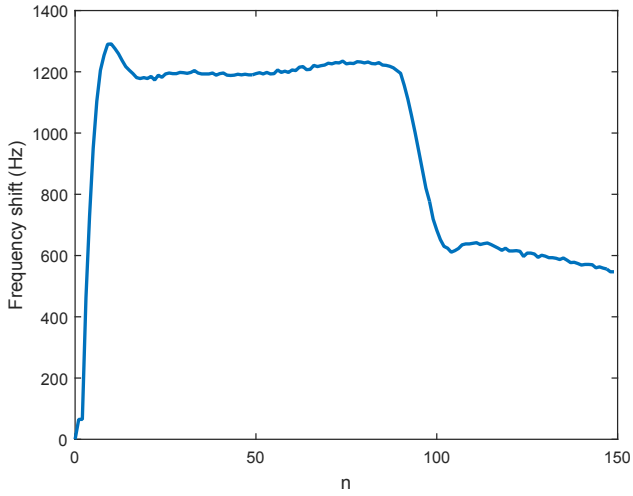
$$G(c^*[n], x) = \frac{G(F_1[n], x)}{\frac{A_{1,m}}{1-T_{1,m}x} + \frac{A_{1,v}}{1-T_{1,v}x}}$$

and substitute it to the other ones $(i = 2..N_s)$

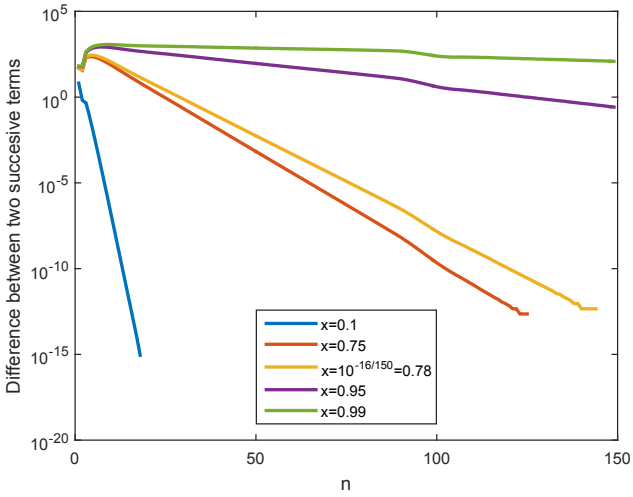
$$G(F_i[n], x) = \left(\frac{A_{i,m}}{1 - T_{i,m}x} + \frac{A_{i,v}}{1 - T_{i,v}x} \right) \frac{G(F_1[n], x)}{\frac{A_{1,m}}{1-T_{1,m}x} + \frac{A_{1,v}}{1-T_{1,v}x}}. \tag{14}$$

So, by construction, the parameters A and T can be estimated by solving the following optimization problem:

$$\begin{cases} \text{argmin} & \sum_{x \in X} \left\| GH(x) \frac{GF(F_1,x)}{\frac{A_{m,1}}{1-T_{m,1}x} + \frac{A_{v,1}}{1-T_{v,1}x}} - F(x) \right\|_2 \\ \text{subject to :} & \\ & \forall i \in [1..N_s] \quad T_{i,m} \in]0..1[\text{ and } T_{i,v} \in]0..1[\\ & \forall i \in [1..N_s] \quad A_{i,m} < 0 \\ & \forall i \in [1..N_s] \quad \frac{A_{m,i}}{1-T_{m,i}} + \frac{A_{v,i}}{1-T_{v,i}} = SS_i \end{cases}$$



(a) Sensor's response.



(b) Convergence of its generating function.

Figure 1: Error on the estimated generating function for different values of x.

where $GH(x) = \begin{pmatrix} \frac{A_{m,2}}{1-T_{m,2}x} + \frac{A_{v,2}}{1-T_{v,2}x} \\ \vdots \\ \frac{A_{m,N_s}}{1-T_{m,N_s}x} + \frac{A_{v,N_s}}{1-T_{v,N_s}x} \end{pmatrix}$,

$F(x) = \begin{pmatrix} GF(F_2, x) \\ \vdots \\ GF(F_{N_s}, x) \end{pmatrix}$ and X is a finite subset of $[-10^{\frac{\epsilon}{N}}, 10^{\frac{\epsilon}{N}}]$. Any distance

function may be used in the objective function. In particular, and without any loss of generality, we used the Euclidean distance. As the previously described optimization problem can not be easily solved analytically and is not convex, we propose to use meta-heuristic algorithms to solve it. In particular, in the experimental section (section 4), we compare the performances of simulated annealing, particle swarm optimization and $(\lambda + \mu)$ evolution strategy.

3 Odour recognition and concentration evaluation

In this section, we assume that we have an estimate of the different parameters of the SAW sensor's frequency shift. We describe a method to identify the chemical compounds and to estimate their concentration.

3.1 Chemical compound identification

Chemical compound identification is a typical supervised learning problem for which we have a training set consisting of n labelled observations also called labelled samples or labelled examples

$$\{(x_1, y_1)...(x_i, y_i)...(x_n, y_n)\}.$$

The x_i are called descriptors and the y_i are categorical variables. The objective is to build, from these learning samples, a model which allows to predict the output y associated with a new sample x. Most of the time, the model is built by choosing a parametrized function and by determining the parameters which minimize an error criterion on the training set [8]. Many models have been investigated to perform odour recognition: SVM [9], neural network [10], k-nearest neighbours [11] to name a few. Since the design of new models is not the primary topic of this article and because a previous unpublished study showed its efficiency, we only investigated the large margin nearest neighbour (LMNN) [12] in the experimental section (section 4). LMNN is a variant of the well known k-nearest neighbours supervised classification algorithm. It was designed to overcome the fact that the traditional Euclidean distance ignores any statistical regularities of the labelled examples. The authors propose to compute the Euclidean distances after performing a linear transformation:

$$D(x, y) = ||\mathbf{L}(x - y)||_2.$$

They propose to determine $\mathbf{L}$ by solving the following convex optimisation problem using specially designed gradient descent algorithm.

$$\mathbf{L} = argmin\ (1 - \mu) \sum_{i \to j} ||\mathbf{L}(x_i - x_j)||_2^2$$

$$+\mu \sum_{i,j \to i} \sum_{l} (1 - y_{i,l})(1 + ||\mathbf{L}(x_i - x_j)||_2^2 + ||\mathbf{L}(x_i - x_l)||_2^2)_+$$

where, $y_{i,l} = 1$ iff $y_i = y_j$, and $y_{i,l} = 0$ otherwise, $(x)_+ = max(0, x)$. The coefficient $\mu \in [0;\ 1]$ is a parameter which balances the two terms of the objective function; when $\mu = 0$, only the large distances between two samples of the same label whereas are penalized, when $\mu = 1$ only the small distances between differently labelled samples are penalized. The notation $j \rightarrow i$ indicates that x_j is a target neighbour of x_i. The target neighbours of x_j are those that we desire to be closer to x_j. This objective function is composed of two terms:

- the first term penalizes large distances between each sampled of the same label; and
- the second term penalizes small distances between differently labelled samples.

3.1.1 Hasse diagram feature subset selection

A common technique in machine learning consists in concatenating many feature sets and selecting the most appropriate subset. Many feature selection heuristic have been proposed in the literature, among them the forward approach, based on a Hasse diagram, have gain a major interest [13]. A Hasse diagram is a directed acyclic graph representing the structure of a partially ordered set. Each element of the set is represented as a node and the vertices correspond to the ordering relation: there exists a vertex between the node A and the node B if A > B. In the context of features subset selection, the considered ordered set is the power set of the features and the ordering relation is the inclusion relation i.e. there exists a vertex between the node A and the node B if B $\subset$ A. Figure 2 shows the Hasse diagram associated with the features $V_1, \ldots, V_4$. The proposed heuristic consists in evaluating the classification performances on

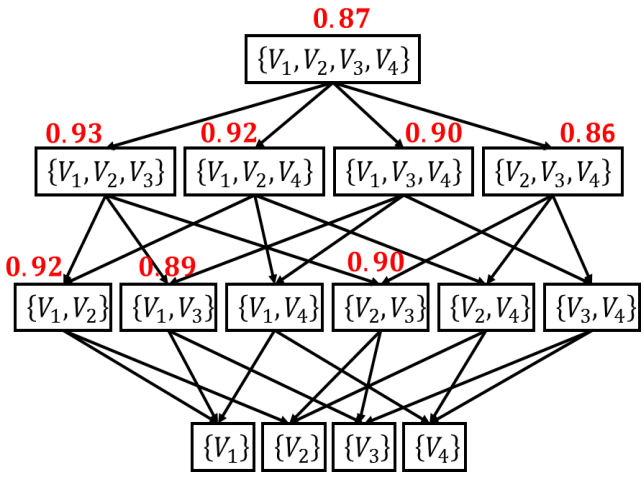


Figure 2: Hasse diagram features subset selection.

the root node and recursively evaluating the classification performances on the

children of the node yielding to the best result. The process is stopped when no children yield to better performances than its parents. In the example of Figure 2, the performance of the first node and of its children is evaluated (the classification performances are in red); since the subset $\{V_1, V_2, V_3\}$ yields to better performances than its parent, its children are recursively evaluated. As none of its children yield to better performance, the algorithm stopped and the subset $\{V_1, V_2, V_3\}$ is returned. While this heuristic has no guaranty to return the optimal subset, the number of node evaluation is drastically lower than the total number of nodes. For instance, for a set of dimension N, in the worst case the Hasse diagram heuristic requires the evaluation of $\frac{N(N+1)}{2}$ nodes while the total number of nodes is $\sum_{k=1}^{N} \frac{N!}{(N-k)!k!} = 2^N - 1$.

3.2 Chemical compound concentration evaluation

The sensor's response, $F_i(n)$, is given by

$$F_i[n] = h_i[0]c^*[n] + h_i[1]c^*[n-1] + ... + h_i[N]c^*[n-N].$$

This equation can be written as $F_i = \mathbf{H_i}c^*$ where

$$\mathbf{H_i} = \begin{pmatrix} h_i[0] & 0 & \cdots & \cdots & 0 \\ h_i[1] & h_i[0] & 0 & & \vdots \\ h_i[2] & h_i[1] & h_i[0] & 0 & \vdots \\ \vdots & & & \ddots & \ddots \\ h_i[N] & & & & h_i[0] \end{pmatrix}.$$

Stacking this equation for each sensor of the nose yields to:

$$\begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_M \end{pmatrix} = \begin{pmatrix} \mathbf{H_1} \\ \mathbf{H_2} \\ \vdots \\ \mathbf{H_M} \end{pmatrix} c^* \text{ or } F = \mathbf{H}c^*$$

and $c^* = (\mathbf{H^T H})^{-1}\mathbf{H^T}F$. However, the measured data F are noisy. In this case the previous equation generally produces a noisy estimate of the concentration profile. To improve the deconvolution, one can minimize the energy of the second-order derivative of the concentration profile. Moreover, in some situation, it may be interesting to add prior knowledge about the concentration profile, for instance $c^*[0] = 0$ i.e. $sc^* = 0$ where $s = [1\ 0..0]$. So c^* can be estimated by solving the problem:

$$\begin{cases} \text{argmin} & ||F - \mathbf{H}c^*||_2^2 + k||\mathbf{D}c^*||_2^2 \\ \text{subject to :} & \\ & sc^* = 0 \end{cases}$$

where $\mathbf{D}$ is the second-order derivative matrix:

$$\mathbf{D} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

This optimization problem can be analytically solved using the Lagrange multipliers method [14]. The associated Lagrangian is:

$$L(c^*, \lambda) = ||F - \mathbf{H}c^*||_2^2 + k||\mathbf{D}c^*||_2^2 + \lambda s c^*$$

$$= F^{\mathbf{T}}F - 2c^{*\mathbf{T}}\mathbf{H}^{\mathbf{T}}F + c^{*\mathbf{T}}\big(\mathbf{H}^{\mathbf{T}}\mathbf{H} + k\mathbf{D}^{\mathbf{T}}\mathbf{D}\big)c^* + \lambda s c^*$$

where λ is the Lagrange's multiplier; computing its gradient [15] gives:

$$\begin{cases} \frac{\partial L(c^*,\lambda)}{\partial c^*} = -2\mathbf{H}^{\mathbf{T}}F + 2\big(\mathbf{H}^{\mathbf{T}}\mathbf{H} + k\mathbf{D}^{\mathbf{T}}\mathbf{D}\big)c^* + \lambda s^{\mathbf{T}} \\ \frac{\partial L(c^*,\lambda)}{\partial \lambda} = sc^* \end{cases};$$

and nullifying it yields to:

$$\begin{pmatrix} c^* \\ \lambda \end{pmatrix} = \begin{pmatrix} 2\mathbf{H}^{\mathbf{T}}\mathbf{H} + 2k\mathbf{D}^{\mathbf{T}}\mathbf{D} & s^T \\ s & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2\mathbf{H}^{\mathbf{T}}F \\ 0 \end{pmatrix}.$$

To obtain the real concentration profile, we must determine the coefficient c_{ss} (see Eq. 7). This is a traditional regression problem. The classical regression problem consists in estimating an unknown function f based solely on a training set of evaluations $\{(x_1, y_1), ..., (x_n, y_n)\}$ where $y_i = f(x_i)$. The task consists in finding an estimator $\hat{f}$ of f that minimizes a loss function over the training points. The standard loss function $\mathcal{L}$ is the squared error

$$\mathcal{L} = \sum\nolimits_i \big(y_i - \widehat{f}(x_i)\big)^2.$$

However, for our application, the squared relative error

$$\mathcal{L} = \sum\nolimits_i \left(\frac{y_i - \widehat{f}(x_i)}{y_i}\right)^2$$

is more appropriate: approximating a concentration of 2 ppm with an absolute error of 1 ppm is much worst than approximating a concentration of 10 ppm with an absolute error of 1ppm.

One common algorithm for regression is kernel regression. In this model, the target value is computed as a weighted average of the function values observed at the training point:

$$\hat{y_i} = \frac{\sum_{j \neq i} k_{i,j} y_j}{\sum_{j \neq i} k_{i,j}}$$

where $k_{i,j} = k(x_i, x_j) \geq 0$ is the kernel function. While kernel regression may apply to many types of kernel functions, we focused on the Gaussian kernel generally defined as follows:

$$k(x_i, x_j) = e^{-\alpha||x_i - x_j||_2^2}.$$

The choice of the parameter α requires great care. Its optimal value depends on the noisiness and the smoothness of the function f. It is chosen by minimizing the squared relative error:

$$\alpha = argmin \sum_i \left(1 - \frac{1}{y_i} \frac{\sum_{j \neq i} e^{-\alpha||x_i - x_j||_2^2} y_j}{\sum_{j \neq i} e^{-\alpha||x_i - x_j||_2^2}}\right)^2.$$

A variant of the traditional Gaussian kernel was proposed by Weinberger et al. [16]. The authors proposed to use a Mahalanobis metric instead of the Euclidean distance:

$$k(x_i, x_j) = e^{-||\mathbf{A}(x_i - x_j)||_2^2}.$$

The main difference between a Mahalanobis metric and a Euclidean metric is that the isoline curve is a circle in case of the Euclidean distance while it is an ellipse in case of the Mahalanobis distance. Once again, the matrix $\mathbf{A}$ is chosen by minimizing the squared relative error:

$$\mathbf{A} = argmin \sum_i \left(1 - \frac{1}{y_i} \frac{\sum_{j \neq i} e^{-||\mathbf{A}(x_i - x_j)||_2^2} y_j}{\sum_{j \neq i} e^{-||\mathbf{A}(x_i - x_j)||_2^2}}\right)^2.$$

The two previously mentioned optimization problems can be solved using a gradient descent. If local minima are a concern, one can use several runs with different random initializations and choose the one with minimum error.

4 Experimental results

4.1 Outline of the electronic nose

The selected electronic nose system is based on an array of six functionalized nano-diamond coated SAW sensors each with a fundamental frequency of 433.9 MHz [17]. The sensors were exposed to seven different gases (NH_3, SO_2, H_2S, CH_3OH, C_7H_8, HCN and $DMMP$) at a concentration of 10 ppm, 8 ppm, 6 ppm, 4 ppm and 2 ppm (only 4 ppm and 2 ppm for HCN). Nitrogen was used as the reference and carrier gas to transport the volatile chemical compounds through the gas cell containing the sensors. The temperature of the sensors (22^o C) and the flow rate (200 mL/min) above them were kept constant. Data acquisition was carried out at 10 Hz using the SAGAS instrument [18]. Several cycles exposition (15 sec) - purge (30 sec) were done for each gas at each concentration.

4.2 Resolution of the optimization problem

In this section, we compare the performances of simulated annealing (SA) [19], particle swarm optimization (PSO) [20] and $(\lambda+\mu)$ evolution strategy $\big((\lambda+\mu)$-ES$\big)$ [21]. The hyper-parameters of the algorithms were set empirically. The candidate solution of SA was perturbed by adding a Gaussian random variable to it, the acceptance probabilities were generated using the Boltzmann function and the temperature is linearly decreased at each step. The inertia weight of the 100 particles of PSO was set to 0.729, their social and cognitive acceleration coefficients were set to 1.494, a ring neighbourhood topology was used. The population of $(\lambda+\mu)$-ES was set to $\lambda = 100$ individuals, $\mu = 20$ children were generated at each iteration using a Gaussian mutation and the 1/5 success rule, their parents were randomly selected using the roulette wheel method. During the optimization process, if a candidate solution did not satisfy the constraints, it was projected into the feasible space: the variables $T_{i,m}$, $T_{i,v}$ and $A_{i,m}$ are clamped within their respective range. Concerning the equality constraint, the variables $A_{i,m}$ and $A_{i,v}$ are projected into the feasible space by multiplying them by a constant α

$$\alpha = \frac{SS_i}{\frac{A_{m,i}}{1-T_{m,i}} + \frac{A_{v,i}}{1-T_{v,i}}}$$

such that

$$\frac{\alpha A_{m,i}}{1-T_{m,i}} + \frac{\alpha A_{v,i}}{1-T_{v,i}} = \alpha\big(\frac{A_{m,i}}{1-T_{m,i}} + \frac{A_{v,i}}{1-T_{v,i}}\big) = SS_i.$$

The algorithms were run until they converged. The parameters K_m, K_v, τ_m, τ_v were then retrieved using Eq. 5 and Eq. 6. Since the real values of the K_m, K_v, τ_m and τ_v are not known, we assessed the accuracy of the previously mentioned algorithms by comparing the value of the objective function after a certain number of evaluations. We also assessed the reproducibility of the algorithms by comparing the average standard deviation of the estimated parameters over 50 runs performed with each sample of the database. Tab. 1 shows the experimental results. It was experimentally established that PSO is the most accurate

Table 1: Experimental results

	SA		$(\lambda+\mu)$-ES		PSO	
Accuracy	1%		16%		83%	
Reproducibility	K_i	T_i	K_i	T_i	K_i	T_i
(std. dev.)	63.83	0.14	64.62	0.17	42.70	0.14

and reproducible heuristic: the objective function was the lowest in 83% of the database and it had the lowest standard deviation. Hence, we concluded that it is the most appropriate heuristic to estimate the parameters of the sensors. Figure 3 shows the evolution of the fitness of the best candidate solution over 10 runs performed with the same sample.

4.3 Odour recognition

In this section, we experimentally compare the performances obtained by using the amplitudes of the contributions estimated with PSO (K_m and K_v), the signal amplitude during the steady state and the features proposed by the authors of [22]. They propose to use the parameters of three different models of sensor's response: the exponential model, the Lorentzian model; and the double sigmoid model. The following table (Tab. 2) shows the classification success rate over a 5-fold cross-validation process using LMNN. These results show that

Table 2: Classification scores.

K_m and K_v	95.7%
Amplitude	94.2%
Exponential model	91.5%
Lorentzian model	93.5%
Double sigmoide model	89.6%

the amplitudes of the two contributions are suitable features to perform odour recognition whereas the features proposed in [22] yielded to the lower classification success rate. As the amplitudes of the two contributions and of the signals during steady state yielded to the best performance and since the misclassified samples were not the same in both cases, we investigated the performance of concatenating these two feature sets and selecting the most appropriate subset. The application of the Hasse diagram heuristic to the chemical compounds identification problem yielded to a classification rate greater than 97% showing the interest of using both descriptors and the efficiency of this heuristic.

4.4 Concentration estimation

In this section, we experimentally compare the accuracy of the concentration estimation process. The following table (Tab. 3) shows the accuracy of the KR and MLKR regression algorithm over a 5-fold cross-validation process. Figure

Table 3: Regression algorithms relative error.

Kernel regression	9%
Metric learning kernel regression	7%

4 summarizes the concentration profile estimation workflow. The first figure represents the responses of the sensors exposed to 8 ppm of ammonia from 0 sec to 10 sec and to 4 ppm of ammonia from 10 sec to 15 sec. The second figure represents the estimated impulse responses using PSO. And the third figure represents the estimated concentration profile using the deconvolution process and the kernel regression method ($\lambda = 10^6$). One can notably remark that the concentration estimate is near the setpoint during the steady state.

5 Conclusion

The optimization problem developed in this study enables the determination of the parameters of the mass loading effect and of the viscoelastic contribution to SAW sensor's frequency shift. This work was motivated by the fact that these features are independent of the concentration profile c and belong to a higher dimensional space than the traditional features used for compound identification and hence may carry more information about them. This assumption was verified experimentally by comparing the classification rate obtained with these features and the one obtained with the steady state amplitude. Moreover, we showed that the classification success rate can reach 97% by using these two feature's space. Moreover we showed that the features we introduced, make possible the determination, with a relative error $< 10\%$, of the temporal profile of the concentration by performing a deconvolution.

Acknowledgement

This research was supported by the French trans-governmental CBRN-E R&D program.

References

[1] A. Loutfi and S. Coradeschi and G. Kumar Mani and P. Shankar and J. Bosco Balaguru Rayappan, *Electronic noses for food quality: A review*, Journal of Food Engineering 144 (103-111), 2015.

[2] E. A. Baldwin and J. Bai and A. Plotto and S. Dea, *Electronic Noses and Tongues: Applications for the Food and Pharmaceutical Industries*, Sensors 11 (4744-4766), 2011.

[3] V. Bhasker Raj and A. T. Nimal and Y. Parmar and M.U. Sharma and V. Gupta, *Investigations on the origin of mass and elastic loading in the time varying distinct response of ZnO SAW ammonia sensor*, Sensors and Actuators B: Chemical 166, 2012.

[4] D.S. Ballantine and R.M. White and S.J. Martin and A.J. Ricco and E.T. Zellers and G.C. Frye and H. Wohltjen, *Acoustic Wave Sensors: theory, design and physico-chemical applications*, Applications of Modern Acoustics, 1997.

[5] B. Tard, *Études des interactions gaz-surfaces diamant par gravimétrie sur résonateur à onde acoustique*, Université Pierre et Marie Curie, 2013.

[6] A. V. Oppenheim and R. W Schafer, *Discrete-Time Signal Processing third edition*, Pearson, 2010

[7] A. R. Meyer and R. Rubinfeld , *Generating Functions*, Technical report 2005.

[8] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

[9] M. Gaudioso and W. Khalaf and C. Pace, *On the use of the SVM approach in analyzing an electronic nose,*, Proceedings of the 7th International Conference on Hybrid Intelligent Systems (42-46), 2007.

[10] T. A. Roppel and D. Wilson, *Improved chemical identification from sensor arrays using intelligent algorithms*, Advanced Environmental and Chemical Sensing Technology 4205 (206-266), 2001.

[11] S. Gney and A. Atasoy, *Multiclass classification of n-butanol concentrations with k-nearest neighbor algorithm and support vector machine in an electronic nose*, Sensors and Actuators B: Chemical 166 (721-725), 2012.

[12] K. Q. Weinberger and J. Blitzer and L. K. Saul, *Distance Metric Learning for Large Margin Nearest Neighbor Classification*, Advances in Neural Information Processing Systems 18, 2006.

[13] S. Derksen and H. J. Keselman, *Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables*, Bristish Journal of Mathematical and Statistical Psychology 45 (265-282), 1992.

[14] R. T. Rockafellar *Lagrange Multipliers and Optimality*, SIAM Rev 35 (183-238), 2006.

[15] K. B. Petersen and M. S. Pedersen *The Matrix Cookbook*, Technical report, 2012.

[16] K. Q. Weinberger and G. Tesauro, *Metric Learning for Kernel Regression*, Proceedings of the International Conference on Artificial Intelligence and Statistics, 2007.

[17] E. Chevallier and E. Scorsone and P. Bergonzo, *New sensitive coating based on modified diamond nanoparticles for chemical SAW sensors*, Sensors and Actuators B: Chemical 154, 2011.

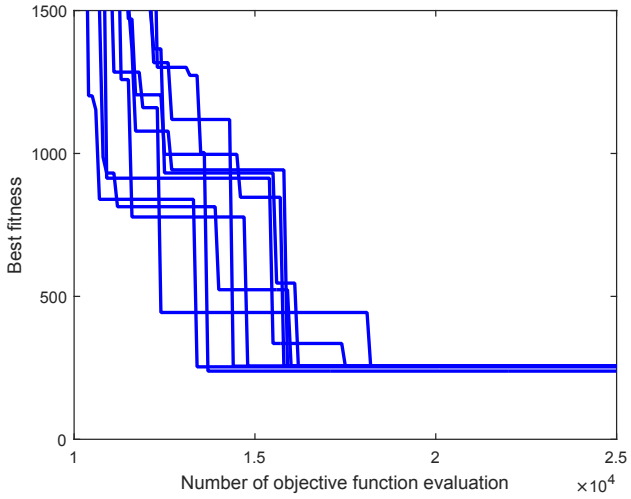
[18] M. Rapp and J. Reibel and A. Voigt and M. Balzer and O. Bülow, *New miniaturized SAW-sensor array for organic gas detection driven by multiplexed oscillators*, Sensors and Actuators B: Chemical 65, 2000.

[19] S. Kirkpatrick, *Optimization by simulated annealing: Quantitative studies*, Journal of statistical physics 34 (975-986) 1984.

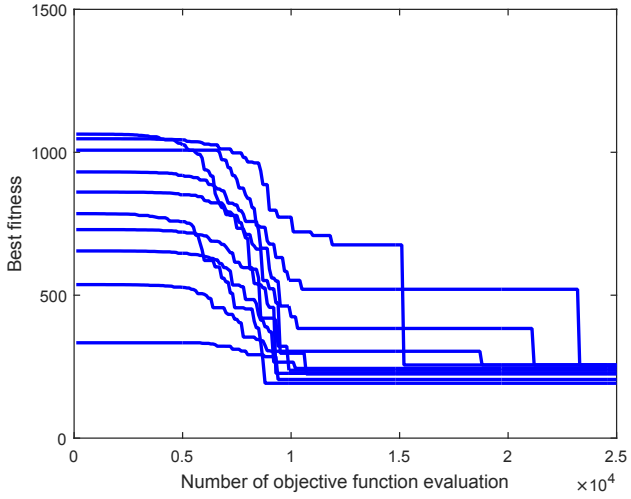
[20] F. Van Den Bergh, *An analysis of particle swarm optimizers*, Phd Thesis University of Pretoria, 2006.

[21] N. Hansen and D. V. Arnold and A. Auger, *Evolution Strategies*, Technical report, 2013.

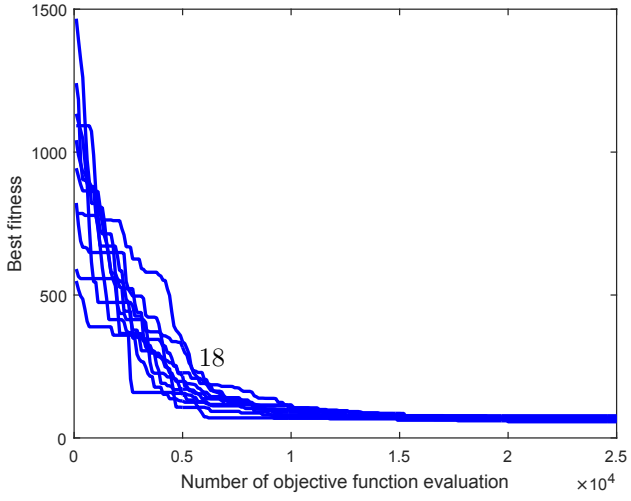
[22] L. Carmel and S. Levy and D. Lancet and D. Harel, *A feature extraction method for chemical sensors in electronic noses*, Sensors and Actuators B: Chemical 95 (67 - 79), 2003.



(a) Simulated Annealing.

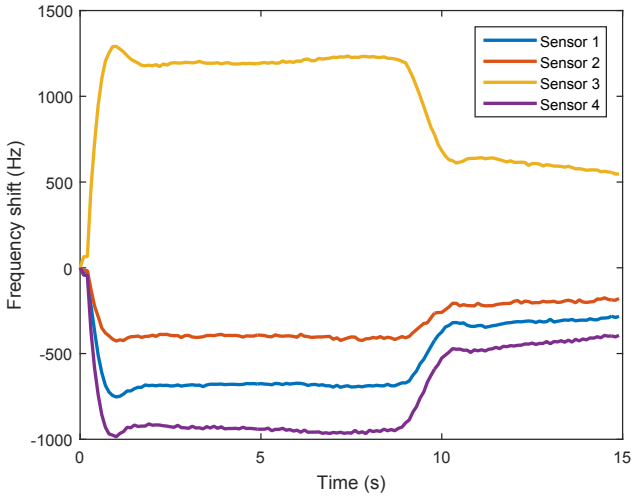


(b) $(\lambda + \mu)$ Evolution Strategy

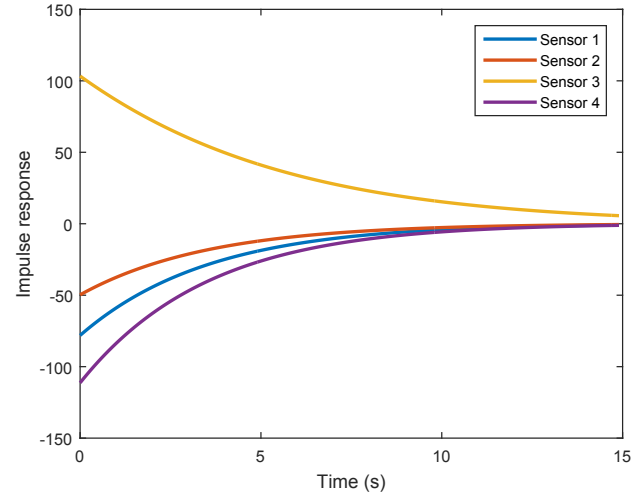


(c) Particle Swarm Optimization

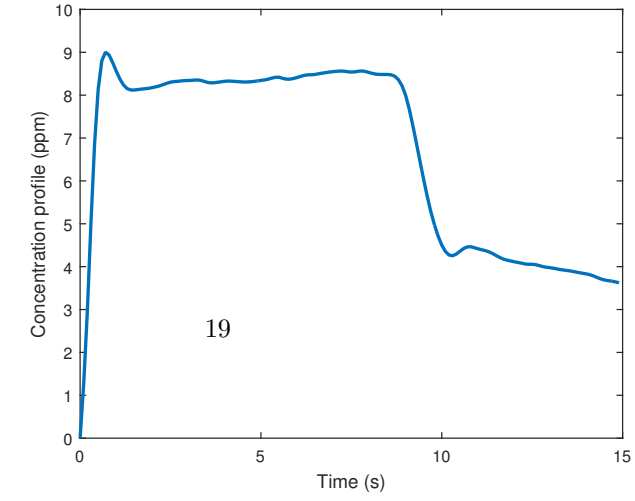
Figure 3: Evolution of the best candidate solution's fitness during the optimiza-



(a) Signals acquisitions.



(b) Impulse responses identification with PSO.



(c) Deconvolution and scaling with kernel regression.