Nonlinear Inverse Dynamic Models of Gas Sensing Systems Based on Chemical Sensor Arrays for Quantitative Measurements

Antonio Pardo, Santiago Marco, Member, IEEE, and Josep Samitier, Associate Member, IEEE

Abstract—Gas sensing systems based on low-cost chemical sensor arrays are gaining interest for the analysis of multicomponent gas mixtures. These sensors show different problems, e.g., nonlinearities and slow time-response, which can be partially solved by digital signal processing. Our approach is based on building a nonlinear inverse dynamic system. Results for different identification techniques, including artificial neural networks and Wiener series, are compared in terms of measurement accuracy.

Index Terms—Gas sensor arrays, neural networks, nonlinear system identification, quartz microbalances, Wiener series.

I. INTRODUCTION

In the past decade, there has been a growing interest in low-cost chemical sensors for domestic [1], industrial [2], and environmental [3] areas. However, low-cost devices such as semiconductor oxide sensors or polymer coated quartz microbalances present a number of problems which preclude their use in critical measurements: strong nonlinearities, slow time response, and nonselectivity, etc. [4]. A partial solution to these problems can be obtained by digital signal processing if an array of different sensors is used [5], [6].

Our system basically consists of a chamber where the chemical sensor array is installed. The mixture concentration in this chamber can be changed using variable flow ratios (see Section II). It can be considered as a multi-input multioutput system (MIMO) where the inputs are the gas flows ratios of the r-gases injected in the chamber $u = (u_1(t), u_2(t), \ldots, u_r(t))$ and the outputs are the sensors responses $y = (y_1(t), y_2(t), \ldots, y_m(t))$ where $m$ is the number of sensors in the array. Note that

$$u_i(t) = \frac{f_i(t)}{\sum_j f_j(t)}$$ (1)

where $f_i(t)$ is the flow of gas $i$.

The measurement equipment based on a chemical sensor array, plus digital signal processing, to be described in this paper, is going to recover the normalized input flows from the sensors outputs (Fig. 1).

If the sensors were linear and in static conditions, the transformation between the gas inputs and the sensor outputs would be given by a matrix

$$y = Au,$$ (2)

If perfect selective sensors were used, the transformation matrix would be reduced to diagonal form. However, real sensors are nonselective so the transformation matrix has to be inverted. Moreover, because more sensors than gases are normally used, the inversion is carried out in a least-squares sense

$$u = (A^T A)^{-1} A^T y,$$ (3)

Because the response of different sensors can be highly correlated, this inversion process can be ill-posed. The resulting estimate of the inputs may be highly sensitive to noise and disturbances affecting the measured sensors outputs. Nevertheless, in practical applications, the problems become more severe because sensors are, in addition to nonselective, nonlinear, and the measurement can take place in nonstatic conditions.

One of the main issues of previous works in this field has been the search for efficient nonlinear static multivariate models of the sensors [7], [8]. Usually, as we have said, more sensors than gases in the mixture are used and, consequently, the equation system is overdetermined and the solution is found by a minimization method (for instance, Levenberg–Marquardt algorithms [9]). In some cases, feedforward multilayer perceptrons have been used to find the inverse solution [10].

We pretend to use chemical sensor arrays as measurement units even if the system response is slow compared to input signal changes. In gas dynamic measurements, in contrast with static conditions, not only the sensor response is important, but also the time to fill the room where the sensor array is installed has to be considered. From a practical point of view, this chamber has necessarily a finite volume, and its dynamics cannot be neglected if the gas input signals change rapidly. Our approach is to build a joint model of the gas sensor array together with the analysis chamber. Because both the sensors and the chamber behave as low pass filters, the recovery of the gas inputs from the sensor array output (deconvolution) is ill-posed [11]. Moreover, this challenge is hindered by the bad behavior of the sensors: static and dynamic nonlinearities, and, in some cases, long term drifts.

Manuscript received August 27, 1997; revised November 30, 1998.

The authors are with the Departament d’Electrònica, Universitat de Barcelona, Barcelona 08028, Spain.

Publisher Item Identifier S 0018–9456(98)09805-2.
The recovery of the gas inputs from the sensors output (measurement problem) corresponds to the search for an inverse dynamical model (inverse filter). In fact, this is a general measurement problem: the measurement instrument, or sensor, always distorts the waveform to be measured. In addition, this measured signal is contaminated by noise [12]. In this case, the objective of digital signal processing is to recover the original signals (gas-inputs) from the noisy and distorted observations.

Several procedures can be applied to recover the gas inputs from the sensor signals. A first method is to obtain, first, a forward model of the system and then invert this model. In the estimation of the forward model, the difference between the sensors output and the model predicted output is used as an error criterion. Nevertheless, the inversion of the model is a difficult task if we consider system’s nonlinearities. The problems to build a forward model of the sensor array as a nonlinear dynamic system have been considered in a previous work by the authors [13], [14]. In the present work, we rather try to identify the inverse system and then apply as error criterion the difference between the real inputs and the estimated ones (Fig. 2). This is possible because our test system permits to control the inputs to the system and, in consequence, they are known.

Section II introduces the sensors object of this study and the measurement equipment. Section III describes the different kinds of test signals used and Section IV describes the different model structures used to build the inverse system and it presents a comparative analysis. Finally in Section V we draw some conclusions.

II. SENSORS AND TEST EQUIPMENT

Four polymer coated quartz microbalances with a fundamental frequency of 10 MHz were used to detect a binary mixture of volatile organic compounds: octane and toluene. The reagents were purchased as pro-analysis standards from Merck and Fluka. Each crystal was inserted in an oscillator circuit. Deposition on, or removal from, the crystal surface of analyte molecules induce a frequency shift. Stationary-phase gas materials for gas chromatography were deposited on the crystal surface in order to exploit the sensing effect. The following polymers were utilized [15] (acronymous and producer indicated in parenthesis):

1) Polydimethylsiloxane (PDMS, Wacker-Chemie).
2) Poly(cyanopropyl)methylsiloxane (PCPMS, Hüls-Petrarch).
3) Poly(aminopropylcarboxylic acid)methylsiloxane (PAPMS, 10% amino groups, University of Tübingen).
4) Poly[2-carboxy (D-valine-β-butylamide) propyl]-methysiloxane (Chirasil-Val, 10% valine groups, University of Tübingen).

A computer controlled scanner allowed the sequential monitoring of each output frequency using a Hewlett-Packard 5334B frequency counter. The sensor array was mounted inside a 200 ml brass chamber with a large heat capacity. The temperature was fixed at 303.0 ± 0.1 K by allowing a liquid of constant temperature to flow through the outer part of the brass block. The thermostat used was a microprocessor-controlled Julabo FP30MH instrument. Test vapors were generated from cooled bubblers using synthetic air as carrier gas and then diluted to known concentrations by computer-driven mass flow controllers. All vapors were mixed and temperature stabilized before entering the chamber. The total gas flow was constant with a value of 400 ml/min. The flow of octane and toluene into the chamber are expressed in the rest of the paper as the ratio of gas flow to the total flow.

III. EXPERIMENTS

In order to characterize the behavior of the system, we decided to apply as input signals, step sequences of varying amplitude for both gases independently. The excitation (diluted toluene-octane flowing through the chamber) and purging (synthetic dry air flowing through the chamber) phases had the same length: 20 min. The setpoints correspond directly...
Fig. 3. (a) Static characteristic of sensor 2 (PCPMS) against concentration of octane. Straight line is plotted to emphasize nonlinearity. (b) Rise time (o) while filling the chamber and fall time (×) while purging the chamber for sensor 2 (PCPMS) against concentration of octane.

Fig. 4. “Big net” fully-connected multilayer perceptron and final architecture after pruning.

<table>
<thead>
<tr>
<th></th>
<th>RMSE(o/o)</th>
<th>Error (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation</td>
<td>0.9</td>
<td>350</td>
</tr>
<tr>
<td>Validation</td>
<td>8.6</td>
<td>1060</td>
</tr>
</tbody>
</table>

Fig. 4. “Big net” fully-connected multilayer perceptron and final architecture after pruning.

<table>
<thead>
<tr>
<th></th>
<th>RMSE(o/o)</th>
<th>Error (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation</td>
<td>0.9</td>
<td>350</td>
</tr>
<tr>
<td>Validation</td>
<td>9.2</td>
<td>1090</td>
</tr>
</tbody>
</table>
TABLE I

<table>
<thead>
<tr>
<th></th>
<th>Toluene</th>
<th>Octane</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exp. 1</strong></td>
<td>Delay: 3 min.</td>
<td>Delay: 1 min.</td>
</tr>
<tr>
<td><strong>Order</strong></td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
<td>1.0%</td>
<td>0.3%</td>
</tr>
<tr>
<td><strong>Exp. 2</strong></td>
<td>Delay: 1 min.</td>
<td>Delay: 2 min.</td>
</tr>
<tr>
<td><strong>Order</strong></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
<td>2.3%</td>
<td>5.7%</td>
</tr>
</tbody>
</table>

Experiment 1: white Gaussian sequences (400 samples) for both gases with mean values equal to 1000 ppm and standard deviations of 360 ppm.

Experiment 2: white Gaussian sequences (400 samples) for both gases with mean values equal to 6500 ppm, and standard deviations of 2800 ppm.

These experiments permit the identification of the inverse system in two different conditions: around two different operation points: 1000 and 6500 ppm and with small (exp. 1) and large deviations (exp. 2), respectively.

IV. INVERSE MODELLING RESULTS AND DISCUSSION

The system dynamics is a combination of the time to fill and purge the test chamber and the sensor dynamics. Unfortunately, both components can not be separated clearly. Taking into account the chamber volume and the total flow, the chamber response can be modeled as a first order linear system with a time constant of about 30 s. However, from Fig. 3 we can observe that the rising and falling times are not constant but depend on the concentration. For high concentrations, the stationary response has not been reached even after 20 min. It can also be observed that when the chamber is filled the sensor responses are slower than when it is purged. Moreover, the step response of different sensors is different. All these considerations indicate that the observed dynamics can not be attributed exclusively to the chamber. On the contrary it can be deduced that the sensor dynamics have an important contribution to the dynamic behavior of the complete system. On the other hand, from the step responses we observe that the static response is also nonlinear, specially for octane. For the different sensors the terminal based nonlinearities in the static characteristic are in the range of 7–15% of the full scale output.

Despite this fact, and as a starting point in our modeling procedure, we decided to find a suitable linear parametric inverse model. From the 400 samples long available sequences, 300 points were used for estimation, and 100 points for validation purposes. Among the different possibilities we chose a MIMO finite impulse response (FIR) model, mainly because of the low expected noise in the output signal (laboratory conditions) and because this model is linear in the parameters permitting the estimation in a one step least-square procedure. No significative improvement was obtained by the addition of poles (auto-regressive with exogenous (ARX) inputs models). The inverse FIR model describes the system by the following:

\[ u_i[n-k] = \sum_{j=1}^{n} B_{kj}(q) y_j[n] + e_i[n] \]  

where \( B(q) \) is a polynomial in the delay operator \( q \) and \( i = 1, \ldots, r \). The delay operator is defined by: \( q[n] = u[n-1], u[n], y[n] \) and \( e_i[n] \) are the output (setpoints), the input (sensor responses) and the residual term respectively. \( k \) is the measurement delay, that indicates the time we have to wait to have the most reliable measurement of the gas inputs. It is necessary to wait 1–3 min for future outputs of the sensors to deduce the current injected gas flows. The optimum FIR models are shown in Table I. As an error criterion we have used the power of the prediction error residuals in the validation set. For convenience, and to easily compare experiments 1 and 2, we have normalized the power of the residuals to the power of the input signals (gas concentrations), relative mean square error (RMSE)

\[ \text{RMSE}_i = \sqrt{\frac{1}{n} \sum_{n} (u_i[n] - \bar{u}_m[n])^2} \]

where \( u_i \) is the actual concentration, \( \bar{u}_m \) is the measured concentration with the inverse model and \( \langle u_i \rangle \) is the mean concentration value. Subindex \( i \) denotes the gas.

RMSE for experiments 1 and 2 in the validation set is shown in Table I. In this table, by order we denote the number of samples for every sensor which are inputs to the model, in other words, the order of the polynomial \( B(q) \) plus one. Clearly, FIR models performed much better when the input concentrations had only a standard deviation of 360 ppm. It indicates that for small excursions of the inputs, the nonlinear behavior could be accurately represented by a local linear approximation.

Because our previous work [14] on forward nonlinear modeling permitted us to find models with less than 1% prediction errors in an extended concentration range (exp. 2), we expected that moving to nonlinear models will improve the measurement accuracy. Among nonlinear models, two main model structures were tested: Wiener series expansions [16] and artificial neural networks (ANN) [17].

Among neural networks, feedforward multilayer perceptrons, generalized radial basis functions (RBF) [17] and Elman recurrent networks [18] have been evaluated as models for the
TABLE II
COMPARISON BETWEEN THE PERFORMANCE OF DIFFERENT INVERSE NONLINEAR MODELS

<table>
<thead>
<tr>
<th>Gases</th>
<th>Wiener Kernel</th>
<th>Feedforward net</th>
<th>RBF net</th>
<th>Elman net</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toluene</td>
<td>delay: 1 min.</td>
<td>delay: 1 min.</td>
<td>Delay: 1 min.</td>
<td>delay: 1 min.</td>
</tr>
<tr>
<td></td>
<td>Order: 3</td>
<td>Order: 3</td>
<td>Order: 3</td>
<td>order: 1</td>
</tr>
<tr>
<td></td>
<td>1st order kernel</td>
<td>5 neurons in the hidden layer</td>
<td>80 neurons in the hidden layer</td>
<td>20 neurons in the hidden layer</td>
</tr>
<tr>
<td></td>
<td>diagonal terms of 2nd order self and cross kernels</td>
<td>RMSE: 1.2% (310 ppm)</td>
<td>RMSE: 0.5% (200 ppm)</td>
<td>RMSE: 3.7% (540 ppm)</td>
</tr>
<tr>
<td>Octane</td>
<td>delay: 2 min.</td>
<td>delay: 2 min.</td>
<td>Delay: 2 min.</td>
<td>delay: 2 min.</td>
</tr>
<tr>
<td></td>
<td>Order: 4</td>
<td>Order: 4</td>
<td>Order: 4</td>
<td>order: 1</td>
</tr>
<tr>
<td></td>
<td>1st order kernel</td>
<td>3 neurons in the hidden layer</td>
<td>75 neurons in the hidden layer</td>
<td>4 neurons in the hidden layer</td>
</tr>
<tr>
<td></td>
<td>diagonal terms of 2nd order self and cross kernels</td>
<td>RMSE: 2.6% (450 ppm)</td>
<td>RMSE: 0.9% (270 ppm)</td>
<td>RMSE: 4.6% (600 ppm)</td>
</tr>
<tr>
<td></td>
<td>RMSE: 0.6% (220 ppm)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. Measurement error by RBF nets against the number of neurons in the hidden layer. "*" Estimation zone, toluene; “+” Validation zone, toluene; “x” Estimation zone, octane; “o” Validation zone, octane.

inverse system. For the calibration of nonlinear sensors with ANN we should mention the constructive methods for building neural networks, which have been discussed in [19] and [20]. These methods permit the interpolation of an arbitrary function from a set of input–output pairs. Nonlinearities are concentrated in the input layer, and weight adjustment is carried out in a single pass over the calibration data. Moreover, the procedure also provides the optimum calibration points. However, this procedure produces large nets when applied to multi-input systems. Its application to the identification of multi-input dynamic inverse systems its not presented here. It deserves further work.

Two independent ANN have been built, trained and tested: one for every gas. Minimum RMSE in the validation zone has again been chosen as criterion to select the number and connectivity of neurons. Moreover, we have checked that the estimated models fulfill the validation criteria proposed by Billings for nonlinear systems [21].

One of the main problems with feedforward nets is the selection of the appropriate number of neurons and their connectivity. The number of inputs corresponds to the number of sensors (four) times the model order, which has been kept equal to the order of the optimum FIR model. That is, 12 input neurons for toluene and 16 input neurons for octane. In addition we have to determine the number of neurons in the hidden layer. To cope with this problem, two different paths have been followed. The first procedure has been to train the nets with an increasing number of neurons in the hidden layer.
A maximum of seven neurons has been considered. Then we choose the feedforward net which provides lower RMSE (see Table II). A second strategy has been to build a “big net,” keeping in mind that we have to prevent overfitting. A “big net” is a multilayer perceptron with a number of free weights larger than the number of patterns in the estimation zone (for example, we tried 16-19-1 for octane with just 300 patterns for estimation). In order to prevent overfitting we stopped the training procedure when the error in the validation zone started to increase. Results in Fig. 4 show that while the error in the estimation zone was small, the net has bad generalization capability resulting in high errors in the validation zone. In the search for better performance by reducing the size of the net we have used a magnitude based pruning algorithm [17]. However, the obtained results were not satisfactory. A partially connected net with a reduced number of neurons presents a slight increase in the prediction error in the validation zone. As it can be observed from the comparison of Fig. 4 and Table II, this second procedure does not improve the results achieved with smaller but fully connected networks (first approach).

Two (one for toluene and another one for octane) generalized RBF neural nets with weighted inputs and common spread

Fig. 6. Comparison between real and measured concentrations using Wiener Kernels. The error is also displayed. Measurement points: dots, model prediction: solid line, error: dashed line.

Fig. 7. Final summary of the better accuracy obtained for the different model architectures tested.
parameter for all the neurons have been used, and trained with an orthogonal least-squares (OLS) learning algorithm [22]. These nets present the optimum results with less neurons than input patterns. Fig. 5 shows the evolution of the error in the validation zone against the number of neurons in the hidden layer showing an optimum point at 80 neurons. Compared to feedforward nets trained with the back-propagation algorithm (or faster alternatives as the Levenberg–Mardquardt method [23]) the used OLS has the advantage that always converges to the same point. In contrast, feedforward nets need to be trained several times, with different starting conditions, in order to be sure that the algorithm has not been trapped in a local minimum. Even one training session with the OLS algorithm is much faster than the Levenberg–Mardquardt method to train the feedforward perceptron.

Disappointing results (see also Table II) were obtained with the Elman networks. This networks do not need a tapped delay input chain because the neuron themselves present a dynamic behavior by virtue of the inclusion of delay elements and recurrent connections. This is the reason way the order in Table II is fixed to 1. This really means that we rely in the internal recurrent connections to generate the dynamical behavior of the system. In our experiments the obtained results do not improve in a significant manner the results obtained by linear methods.

Finally, we consider Wiener series expansions as a method for nonlinear system modeling. The mathematical expression, up to second order, can be written as

\[
u[n-k] = u_o + \sum_{j=1}^{4} \sum_{m=0}^{N-1} k^{(1)}_{j}[m] (y_j[n-m] - y_j o) + \sum_{j=1}^{4} \sum_{m=0}^{N-1} \sum_{p=0}^{N-1} k^{(2)}_{jp}[m,p] (y_j[n-m]y_j[n-p] - P_{jp}) + \sum_{j=1}^{4} \sum_{l=j+1}^{4} \sum_{m=0}^{N-1} \sum_{p=0}^{N-1} k^{(2)}_{jl}[m,p] (y_j[n-m]y_l[n-p] - P_{jl})
\]

where, \( k^{(1)}_{j} \) is the first-order kernel, \( k^{(2)}_{jp}[m,r] \) are the second-order self-kernels and \( k^{(2)}_{jl}[m,r] \) are the second-order cross-kernels. \( P_{jl} \) is the cross-power between signals \( y_j \) and \( y_l \). \( u_o \) and \( y_j o \) are the mean values of \( u[n] \) and \( y_j[n] \). \( k \) is the measurement delay.

Although Wiener series expansions are essentially a non-parametric model structure, models with finite memory can be estimated by least-squares [23]. This is a much more efficient method than the estimation by cross-correlation techniques [24] which need very long records for accurate estimation, especially for multi-input systems [25], [14]. Our procedure consist of two steps: first a rough estimation by cross-correlation and second a selection of the samples of the kernels which are going to be estimated by least-squares. Several manual iterations were necessary to find the best models but fortunately the estimation was fast. Using this procedure the order of the model was selected and in addition we discovered that only the diagonal terms of the second order kernels (self- and cross-kernels) were significant. This resulted in a smaller model. The best results are shown in Table II. Fig. 6 shows a very good agreement in the injected flows of toluene and octane, real and estimated by Wiener Kernels, and the measurement error for every gas.

As a summary, Fig. 7 plots the measurement errors achieved with every model structure considered in the work. The obtained results with the static calibration are also shown for reference. From the results, we observe that among the different model structures, the best ones are Wiener Kernels and Generalized RBF neural networks that show similar performances. Feedforward has good results but not as good as we expected and Elman nets showed RMSE values close to those offered by linear models (compare Tables I and II). The errors for the two best models are under 1% RMSE. We consider that this error is inherent to the experiment and it is due to the mass flow controllers that feature an accuracy of 1% full scale (180 ppm).

From the complete results it is clear that multilayer perceptrons, either fully or partially connected, do not improve the performance of the Wiener Kernel. In consequence, it seems worthless to use perceptrons when a faster method is available. From our experience, this could be the case when systems exhibit only mild nonlinearities and a limited number of input patterns are available. Probably, when working with more nonlinear systems, the truncation of the Wiener series to second order terms will not be enough and, in these conditions superior performance of the perceptrons could be expected.

V. Conclusion

Polymere coated quartz microbalance sensor arrays permit the analysis of multicomponent gas mixtures in dynamic conditions were static models result in measurement errors. Sensor problems including nonlinearities and dynamic effects can be solved with proper digital signal processing of the sensor signals. Our proposal, the identification of nonlinear inverse models has proved to be a valuable tool. The measuring system consisting of a chemical sensor array plus digital signal processing (inverse model) has been able to extract which were the gas flows into the chamber with a very small error near the experimental limit of the system. The ill-posedness of the inverse problem has been solved by straight identificatican of the inverse system from available experimentation (estimation zone). Among the different model structures tested, the authors prefer Wiener kernels because of its small errors and fast estimation. Among ANN, generalized RBF neural networks can also be used providing better results than multilayer perceptrons in similar conditions. Rather poor results were obtained with Elman networks.

Acknowledgment

The authors would like to thank A. Hierlemann, J. Mitrovics, M. Schweizer, U. Weimar, and W. Göpel, Tübingen University, for providing the measurements needed to complete this work, and F. A. M. Davide, Telecom Italia, for fruitful discussions.
REFERENCES


Antonio Pardo received the diploma in physics from the University of Barcelona, Barcelona, Spain in 1991 and is pursuing the Ph.D. degree in the Department of Applied Physics and Electronics, University of Barcelona.

His research interests focus on signal processing for gas sensors and pattern recognition.

Santiago Marco (M’98) received the B.S. degree in physics and the Ph.D. degree (with honors) in the development of a novel silicon sensor for in vivo measurements of the blood pressure, both from the University of Barcelona, Spain, in 1988 and 1993, respectively.

He has been Associate Professor at the University of Barcelona since 1995. From 1989 to 1990, he worked in the electrooptical characterization of deep levels in GaAs. From 1990 to 1993, he was a regular visitor of the Centro Nacional de Microelecronicà, Barcelona. In 1994, he was a Visiting Professor at the University of Roma, Italy, working in electronic noses. He has published about 25 papers in scientific journals and books. His current research interests are the modeling, simulation, and test of Microsystems and the development of intelligent gas sensor arrays.

Josep Samitier (A’95) studied physics at the University of Barcelona, Barcelona, Spain, from 1977 to 1982.

He has been Full Professor at the University of Barcelona since February 1995. He was a Research Fellow from 1983 to 1984 in the Applied Physics Department, University of Barcelona, in the field of GaAs-MESFET devices and electrooptical characterization of III-V semiconductors. From February 1984 to June 1985, he was a Visiting Research Fellow at the Laboratoire D’Electronique Philips, LEP, Paris. In 1988, he was appointed Associate Professor at the same university working in the development of Microsystems and electronic instrumentation. Current research and developed projects concern the design, test, and signal conditioning of Microsystems and the design of interface circuits for sensors and Microsystems. At present, he coordinates the projects in electronic instrumentation of the electronic engineering R&D activities.