Analusis, **1998**, *26*, 135-141
© EDP Sciences, Wiley-VCH

Detecting information in gas sensor responses using analysis of variance

C. Nicolas[1], A.S. Barros[2], D.N. Rutledge[2], J. Hossenlopp[3], G. Trystram[4] and C. Emonet[4]

[1] *Laboratoire d'Évaluation Sensorielle Nestlé France, 7 Bd. Pierre Carle, BP. 900 Noisiel, 77446 Marne-la-Vallée, France*
[2] *INA P-G, Laboratoire de Chimie Analytique, 16, rue Claude Bernard, 75231 Paris Cedex 05, France*
[3] *CEMAGREF, Équipe Qualité Alimentaire, Domaine de Laluas, 63200 Riom, France*
[4] *INRA - ENSIA, Département de Génie Alimentaire Industriel, 1 Av. des Olympiades, 91305 Massy, France*

Abstract. This article describes how Analysis of Variance may be used to select those regions of the curves generated by a gas sensor array which contain the most discriminant information for a particular application. The Analysis of Variance is performed on each point of the signals generated by the sensor array for a particular set of samples. The Group Variances and the Residual Variances are plotted as functions of their position in the signal. Regions of the signals that vary systematically from one group of samples to another will have high Group Variance values and low and randomly distributed Residual Variance values. This method has shown for a particular set of products for cats analysed with a thirty two polymer sensor array, that the most discriminant information is located at the end of the curves. It has also shown an absence of discrimination between standard and tainted pig fat samples with this sensor array. The advantage of this method is that it can be used on almost any sort of raw signal as a pre-analysis step in order to know whether it is worthwhile going on to more fastidious and time consuming signal analysis procedures.

Key words. Discrimination – gas sensors – foodstuffs – volatile compounds – Analysis of Variance.

Introduction

In food industry, classical analysis of food flavour is done with physico-chemical analytical tools such as GC/MS or with Sensory Analysis techniques. Thus, most of the time, these traditional methods are expensive and time consuming. An alternative technique to evaluate the gas quality of food products is now being widely investigated: gas sensor arrays [1,2]. Gas sensors have a characteristic electrical resistance which varies rapidly with the adsorption of volatile molecules. Electrical signals generated by an array of gas sensors can be analysed with appropriate statistical methods in order to give a representation of different products. At the moment, artificial odour sensing systems arouse much interest in a number of industries as they seem to be a very promising rapid technique for aroma control.

However, there is still much research and development work to be done before the sensors are ready for on line quality control. Sensors developers must work on sensing materials to improve manufacturing repeatability, to reduce their sensitivity to temperature, to humidity, to other interfering gases, to poisoning and drift. Sensor users need to work on sample preparation and above all on data treatment with statistical tools [3]. Much work is being done in both areas. For polymer sensors for instance, different teams have been working on the deposition techniques so that they become as reproducible and as insensitive to process parameters as possible [4,5]. In this article we will mainly focus on the users side by proposing a statistical method to analyse the signals generated by a sensor array and detect discriminant information in them.

Indeed, sensors arrays may generate a huge quantity of data that must be reduced in order to extract the information relevant for a particular application. Of the different applications to be found in the bibliography, only a few give detailed justification for the regions of the sensor signals chosen to characterise the analysed products. Many applications are based on a response vector constituted of the maximum values for all the sensors in the array [1,6,7], without taking into account the dynamic phase or the shape of the curves. *A priori*, nothing indicates that the dynamic phase or the shape do not contain relevant information for the discrimination of the products. One particular approach, to take into account the entire curves, consists in decomposing the curves into a certain number of discontinuous values [8]. The drawback of this method is that it requires a large number of variables to define each sample. To analyse the matrix that is generated in this way by a factor analysis or by a neural network, it is necessary to get a large number of samples to have a reasonably balanced samples/variables ratio. A second approach consists in modelling the curves with a particular function, and then extracting the model parameters [9,10], or in directly extracting a few values for each curve with a step by step descendent Discriminant Factor Analysis [9]. Here an alternative method is presented, that can be applied to systematically analyse the entire curves and to extract the relevant information in them by performing an Analysis of Variance at each point of the signals. This method is applied directly to the raw signals without proceeding to a modelisation step which may be time consuming and source of errors due to incorrect choice of the model or incorrect estimation of the parameters.

Received September 23, 1997; revised February 10, 1998; accepted February 17, 1998.

Article available at http://analusis.edpsciences.org or http://dx.doi.org/10.1051/analusis:1998125

Analysis of Variance (ANOVA) is undoubtedly a very commonly used technique for determining if variables are significantly sensitive to changes in factor levels. In Sensory Analysis for example the Analysis of Variance is used to determine whether a human panel perceived different products as being similar or not for various attributes [11]. However, Analysis of Variance is mostly used to analyse individual variables and not signals. In signals generated by gas sensors the information is spread all along the curves and it is necessary to detect the regions in the signals that vary systematically from one group of products to another. The application of the Analysis of Variance to each point of the curve seems to be appropriate to find out these interesting regions. This approach has been previously used to analyse NMR relaxation curves in a similar way [12-14].

Material and methods

Data acquisition

Twelve different moist products for cats were analysed with 32 polymer gas sensors of the Aroma Scan instrument manufactured by the University of Manchester [15,16]. Four replicates were done for each of the twelve products, on 200 g from four different cans (same batch, same manufacturing date). These 200 g were put in a plastic bag, the plastic bag was inflated with pure 99.999% nitrogen (with a controlled humidity of 50%). The sample bags were left for five minutes at room temperature (20 °C, constant) for headspace generation. The measurement took place using the following cycle: during one minute the pure nitrogen at 50% humidity was passed through the sensor chamber. The sensors' base resistances were calculated with this reference nitrogen. The headspace from the sample bag was then pumped for 100 seconds (150 mL/min). The sensors were then cleaned with the headspace of a solution of water and 2% butanol for one minute. Finally the reference nitrogen was passed through the sensors again in order to reset them to their base resistances.

Data treatment

For each measurement, each of the 32 sensors generates a curve (101 points for each sensor at one point per second) with an ascending phase and a second phase during which the slope decreases greatly (Fig. 1). In recent studies published on gas sensor, several approaches have been used to determine the regions of the curves that contain the relevant information for the discrimination of the products. In particular, Vernat-Rossi et al. [9] compared results obtained for two different analyses on the discrimination of dried sausages of different origins and different aromas, with metal oxide sensors. The first study was a Discriminant Factor Analysis on the most informative points in the dynamic phase as determined by a preliminary step by step Discriminant Factor Analysis. The second study was a Discriminant Factor Analysis on the parameters resulting from a modelisation of the dynamic phase using a Gompertz sigmoïd. The results of this comparison showed that, for this particular study, the relevant information was contained in the initial dynamic phase of the curves (first ten seconds of the analysis). For other applications [1,6,7,17] the relevant information is contained in the second phase where the slope is weaker.

In the present study a univariate technique, Analysis of Variance, is applied to the raw signal to determine which regions of the curves are most informative.

With the Aroma Scan instrument, at a fixed time $t = x$, the signal $I_{t=x}$ generated by a particular sensor can be defined as:

$$I_{t=x} = 100 \times (R_{t=x} - R_{t=0})/R_{t=0}$$
$$R_t = \text{resistance of the sensor at time } t. \qquad (1)$$

A matrix was created in which each sample was defined by the 32 sets of 101 points generated by each sensor, giving 3232 variables for each sample. Products were divided into twelve groups with four samples in each (four replicates of each product).

The Analysis of Variance used by Vackier et al. [12] and by Rutledge et al. [13,14] to analyse NMR relaxation curves was applied to determine the overall influence of the product on the sensor responses and to determine which regions of the curves are the most sensitive to the product effect. Calculations were performed using a program developed in

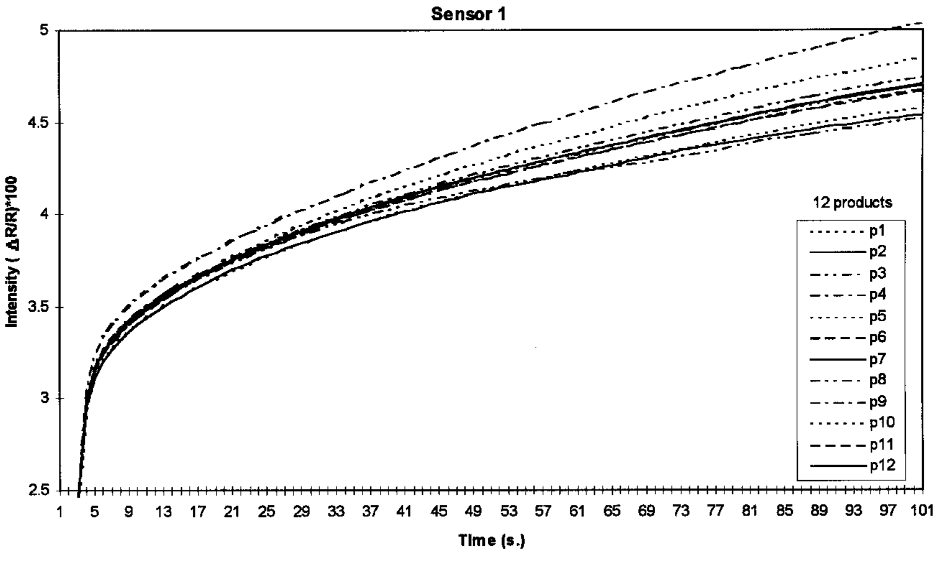

Fig. 1. Average responses from sensor 1 to the volatile compounds of the twelve products (in the first seconds of the analysis, all the curves merge into one single curve, at the end of the analysis, curves from different products become quite different).

the Laboratory of Analytical Chemistry of INA P-G, and validated using MATLAB routines. This Analysis of Variance was done for each column of the matrix, i.e. for each of the 3232 variables. ANOVA calculates the variability at each point in the curve which is due to the predefined grouping of the samples, and the variability which is not explained by the groups.

Each of the twelve groups of products P_i (i = 1 to 12) is characterised by an average value $y_{i.}$ and an estimated variance σ_i^2 (calculated using the 4 replicates, j = 1 to 4), which measures the dispersion of the values of the variable within each of the twelve groups. An average dispersion within groups can be calculated, which is also called within-group or Residual Variance (2):

$$V_R = \sum_{i,\ j} (y_{ij} - y_{i.})^2 \,/\, n-k$$

(n = 4 × 12 = 48 and k = 12). (2)

The dispersion of the group averages $y_{i.}$ compared to the grand average $y_{..}$ is called the between-group or Group Variance (3):

$$V_G = \sum_{i=1}^{k} n_i (y_{i.} - y_{..})^2 \,/ k-1 \ . \qquad (3)$$

If the groups have similar values for a given variable, the between-group Variance of the variable is very low. If the groups are very different, the between-group Variance tends to be higher. In order to analyse the significance level of the differences for a given variable, one usually studies the Group Variance/Residual Variance ratio (Fisher F value). The calculation of the F ratio and of the associated probability is undoubtedly a good statistical criterion to estimate the significance level of the influence of an individual variable on the factor studied. But here, the variables studied are points in a signal. We found that it is often more interesting to plot the between-group Variance values as function of their position in the signal and look at the evolution of this plot. The regions of the curves that have a high Group Variance compared to the Residual Variance may be considered as containing relevant information to discriminate the products. Similarly, structure in the Residual Variance can also indicate the presence of information which has not been explained by the grouping of the samples used as the ANOVA factors. Sometimes high F values may be found even when Group Variance values are not very great, simply because of very small values for Residual Variance. Furthermore, when there are no important differences between samples, other than that due to the groups, the Residual Variance will not only be low but also distributed in a random fashion. Thus, F plot can be quite "noisy" and difficult to interpret. For that reason, Group Variance values were studied rather than F values.

Results and discussion

The Analysis of Variance shows that the Group Variance increases almost constantly, for all sensors, from the beginning of the curve to the end of the curve (Fig. 2). Furthermore, the variance between the twelve groups is significantly higher than the Residual Variance (Fig. 3) and the closer to the end of the curves, the more the difference becomes significant for all 32 sensors. In this particular case the Residual Variance values are structured, indicating that some information in the signals has not been taken into account by the grouping of the samples. Figure 4 shows the F ratio. As the F ratio value for a significant difference between groups at the 95% probability level (11 and 36 degrees of liberty) is 2.059, nearly the whole curve is significant.

These results show that it is best, in this particular study, to use the values at the end of the curves to discriminate the products. Although in the case of this particular application, it may have been interesting to study even longer

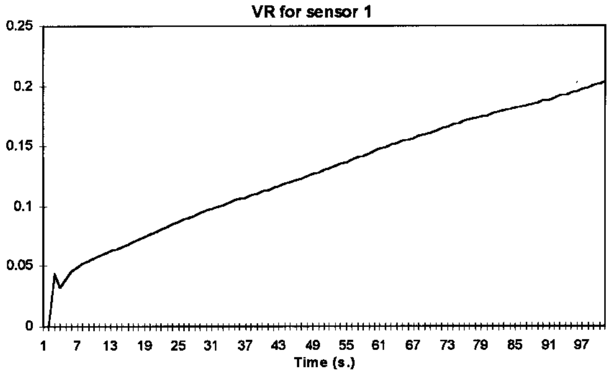

Fig. 3. Evolution of the Residual Variance within the twelve groups of moist cat food products.

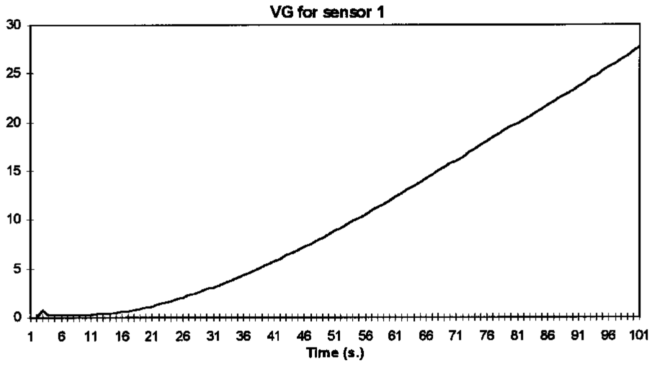

Fig. 2. Evolution of the Group Variance between the twelve groups of moist cat food products.

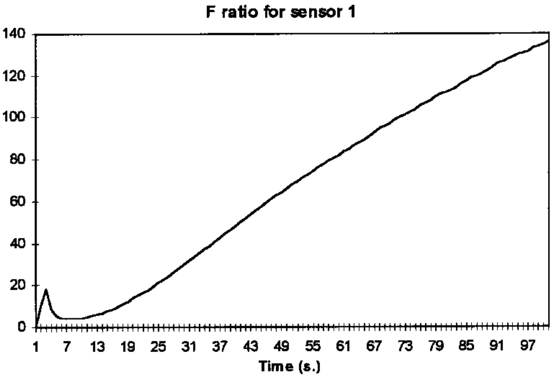

Fig. 4. Evolution of the F ratio for the twelve groups of moist cat food products.

measurement times, the data of the end of these signals gives a satisfactory discrimination between products.

In other cases this Analysis of Variance could be very interesting to demonstrate an absence of discrimination. Standard and boar tainted pork fat samples were analysed (5 grams heated at 80 °C during 30 minutes, five replications for each group). The measurement cycle was the same as that used for the moist cat food, measurements being done using an autosampler - the headspace analysis itself lasted 80 s. instead of 101 s. (Fig. 5). Results of the ANOVA, carried out on these data, gave Group Variances and Residual Variances which show that there is no discrimination between these two groups. Figures 6, 7 and 8 show respectively the Group Variances, the Residual Variances and the *F* ratio for sensor 1. Here the minimum theoretical *F* ratio value should have been 5.12 to have a 95% significance level (1 and 9 degrees of liberty). In this application, the method shows that variability between groups is not significantly higher than within groups (standard or tainted). These results may be due to problems of variability in meat products, lack of sensitivity of the sensors to the molecules responsible for the taint, or high water contents at 80 °C that can mask the sensors' responses to the volatile molecules.

In order to be more complete, we also modelled the signals obtained with cat food samples and with pig fat samples in order to extract representative parameters for the curves shape. These models were calculated using MATLAB routines. Cat food signals were modelled with a four parameters function:

$$i(t) = p(1) + p(2) \times t + p(3) \times \exp\,(t^{p(4)}) \; . \tag{4}$$

This function was used to decompose the thirty two signals generated by the four replicates on the twelve analysed products. Figure 9 shows the estimated curve for the first measurement of product 1 on sensor 1, the sum of square errors being 0.0878.

Pig fat samples were modelled with a three parameters function (these parameters were calculated separately for the time intervals [4 s; 35 s] and [35 s; 80 s]):

$$i(t) = p(1) + p(2) \times \exp(-\,t \times p(3)) \; . \tag{5}$$

This function was used to decompose the thirty two signals generated by the five replicates on the two types of pig fat products where decomposed. Figure 10 shows the estimated curves for the measurement on the first standard pig fat sample by sensor 1. Here the sums of square errors were respectively 0.0462 and 0.0003 for the first (Fig. 10a) and the second (Fig. 10b) parts of the curve. It could be stressed that it was very difficult to decompose all the curves from the second phases (35 s to 80 s) for the pig fat samples.

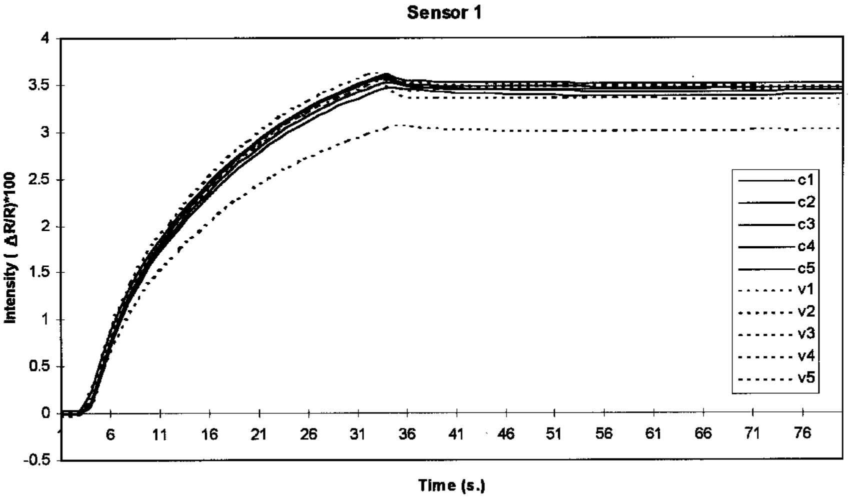

Fig. 5. Responses from sensor 1 to the volatile compounds of the standard pig fat samples (c1 to c5) and of the tainted pig fat samples (v1 to v5).

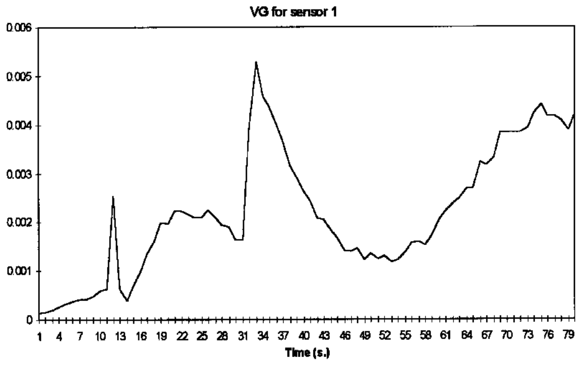

Fig. 6. Evolution of the Group Variance between the two categories (standard or tainted) of pork fat samples.

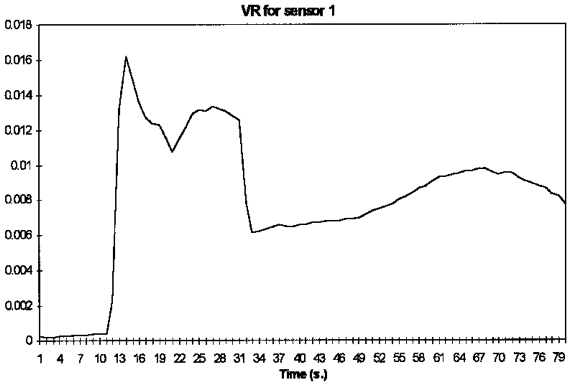

Fig. 7. Evolution of the Residual Variance within the two categories (standard or tainted) of pork fat samples.

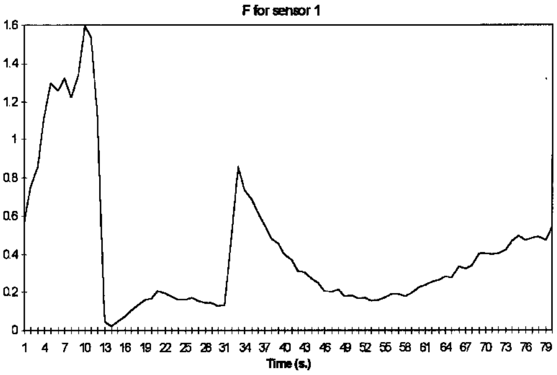

Fig. 8. Evolution of the *F* ratio for the two categories (standard or tainted) of pork fat samples.

Table I gives the results of the Analysis of Variance performed on the four parameters calculated for the cat food sample signals. We can notice that parameter *p*(2) which represents the slope at the end of the signal is the one that discriminates the products best. Table II gives the results of the Analysis of Variance performed on the parameters calculated for the pig fat samples (results are given only for the decompositions on the first phase of the signals, as the decompositions of the second parts were very uncertain). These results confirm that there is no relevant information in the curves to discriminate standard and boar tainted fat samples.

This complementary analysis on parameters representative from the curve shape, compared to the results of the Analysis of Variance performed on each point of the raw curves, shows that this second method is a very good tool to show up if the curves contain interesting information. This can be done before proceeding to use the fastidious and often uncertain signal decomposition procedures required to have the characteristic parameters of the signal.

Conclusion

Table I. Results of the Analysis of Variance performed on the four parameters characterising the signals of the cat food samples: the significance level of the differences between groups is given by *p* (groups are significantly different with a 95% probability if $p<0.05$).

Capteurs / p	*p(1)*	*p(2)*	*p(3)*	*p(4)*
c1	0.0000	0.0000	0.0000	0.0010
c2	0.0000	0.0000	0.0000	0.0045
c3	0.0000	0.0000	0.0000	0.1016
c4	0.0000	0.0000	0.0000	0.0924
c5	0.4991	0.0000	0.6655	0.0027
c6	0.9427	0.0000	0.8977	0.2234
c7	0.0000	0.0000	0.0000	0.0346
c8	0.0000	0.0000	0.0000	0.0234
c9	0.0000	0.0000	0.0000	0.0156
c10	0.0000	0.0000	0.0000	0.1775
c11	0.0000	0.0000	0.0000	0.0086
c12	0.0000	0.0000	0.0000	0.0127
c13	0.0000	0.0000	0.0000	0.0121
c14	0.0000	0.0000	0.0000	0.3823
c15	0.3626	0.0000	0.1268	0.0005
c16	0.3046	0.0000	0.1104	0.0005
c17	0.2749	0.0000	0.2749	0.0103
c18	0.0062	0.0000	0.0032	0.0000
c19	0.0000	0.0000	0.0000	0.4967
c20	0.0000	0.0000	0.0000	0.1081
c21	0.0522	0.0000	0.0141	0.0000
c22	0.9835	0.0003	0.9451	0.1331
c23	0.8989	0.0000	0.8003	0.0407
c24	0.8345	0.0000	0.7041	0.0148
c25	0.0022	0.0000	0.0017	0.0000
c26	0.0000	0.0000	0.0000	0.0081
c27	0.0000	0.0000	0.0000	0.0000
c28	0.0587	0.0000	0.0004	0.3829
c29	0.0000	0.0000	0.0000	0.0114
c30	0.0000	0.0000	0.0000	0.0065
c31	0.0000	0.0000	0.0000	0.1594
c32	0.4853	0.0000	0.5911	0.4055

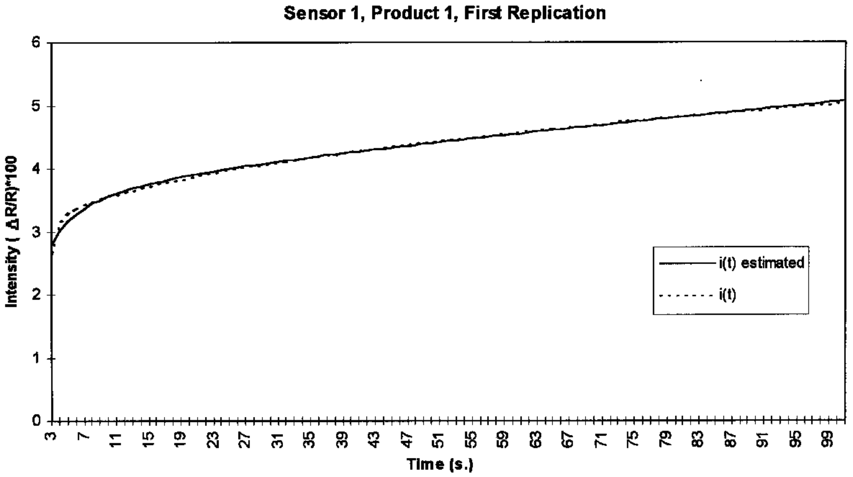

Fig. 9. Estimated and real curve for the first measurement of product 1 on sensor 1 (points for seconds 1 and 2 of the acquisition where the signal has not started to rise yet were not taken into account).

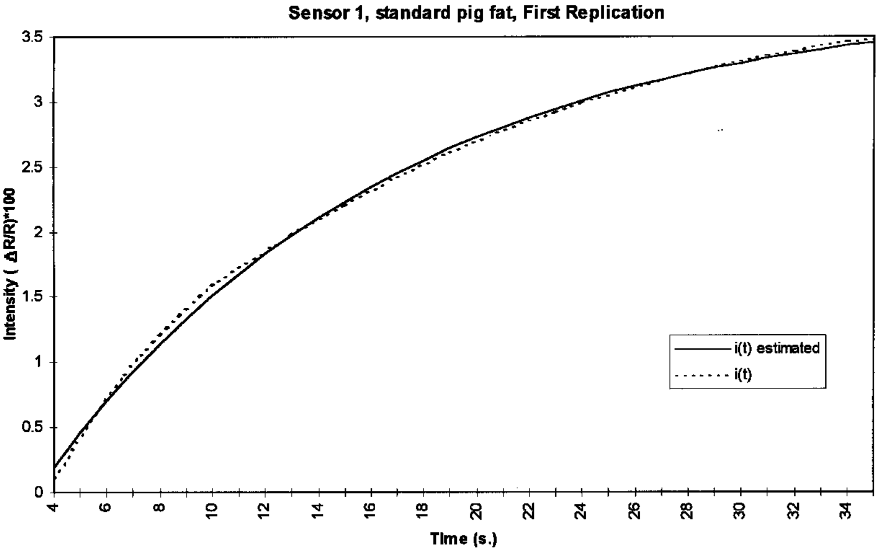

Fig. 10a. Estimated and real curve for the measurement on the first standard pig fat sample by sensor 1. Acquisition for the interval [4 s; 35 s] (points for seconds 1, 2 and 3 of the acquisition where the signal has not started to rise yet were not taken into account).

Table II. Results of the Analysis of Variance performed on the three parameters characterising the signals of the pig fat samples (signals from 4 to 35 seconds are considered here): groups would be significantly different with a 95% probability if $p<0.05$.

Capteurs / p	*p(1)*	*p(2)*	*p(3)*
c1	0.4181	0.3112	0.8535
c2	0.4967	0.3380	0.8390
c3	0.4329	0.3373	0.9949
c4	0.4449	0.3339	0.9806
c5	0.3996	0.3204	0.9781
c6	0.3940	0.3340	0.9668
c7	0.4726	0.3568	0.8986
c8	0.4831	0.3719	0.7905
c9	0.3622	0.3748	0.7945
c10	0.4241	0.3802	0.7496
c11	0.2115	0.3585	0.8415
c12	0.5078	0.3576	0.8093
c13	0.4424	0.3696	0.8026
c14	0.4845	0.3782	0.6411
c15	0.5661	0.5038	0.3467
c16	0.3587	0.3536	0.7585
c17	0.5496	0.3931	0.5138
c18	0.4292	0.3950	0.4159
c19	0.4160	0.3807	0.6083
c20	0.3976	0.3991	0.4732
c21	0.3799	0.3501	0.6264
c22	0.4685	0.5407	0.3650
c23	0.3920	0.3848	0.5124
c24	0.4104	0.4084	0.5504
c25	0.2310	0.2434	0.3465
c26	0.3655	0.3828	0.6629
c27	0.4176	0.3771	0.6493
c28	0.4308	0.3918	0.7430
c29	0.4073	0.3857	0.7047
c30	0.4476	0.3916	0.6911
c31	0.3924	0.3782	0.6371
c32	0.3637	0.3669	0.6173

The method proposed is an interesting and rapid data analysis tool to detect the presence of relevant information in the sensor responses. In the cat food example, it was shown that the relevant information to discriminate the products is contained at the end of the curves. This could be explained by the fact that adsorption and diffusion phenomena on sensor surfaces have probably stabilised [18], so that the differences between products are more obvious. However the gas sensitivity mechanism of the polymers is poorly understood at present. Different hypotheses show that the interaction between organic vapours and organic semi-conductors depends on the type of polymer, on the type of counter-ion, on the analysed vapour [1]. Each application must therefore be studied separately and the ANOVA is certainly well adapted to do this. It has also been shown that it can detect when there is no discriminant information in the curves. We believe that this method can contribute to the improvement and the diffusion of the gas sensor technology as it gives a good basis to know if a particular sensor array gives a pertinent response for a given problem.

Acknowledgements

The authors wish to thank Pascale Carel, of the Friskies Research and Development Center of Amiens, for her help and support and Dr. Bertrand Heyd, from the ENSIA, for fruitful discussions and help with curve parameter calculation.

References

1. Gardner, J. W.; Bartlett, P. N. in: Sensors and Sensory Systems for Electronic Noses, Gardner, J. W.; Bartlett, P. N. Eds., Kluwer Academic Publishers, 1992; Vol. 212.
2. Gardner, J. W.; Bartlett, P. N. *Sensors Act. B* **1994**; *18-19*, 211-220.
3. Gardner J. W.; Bartlett, P. N. *Trends Anal. Chem.* **1996**, *5*(9), 486-493.
4. Gardner J. W.; Pike, A.; de Rooij, N. F.; Koudelka-Hep, M.; Clerc, P. A.; Hierlemann, A.; Göpel W. *Sensors and actuators B* **1995**, *26-27*, 135-139.

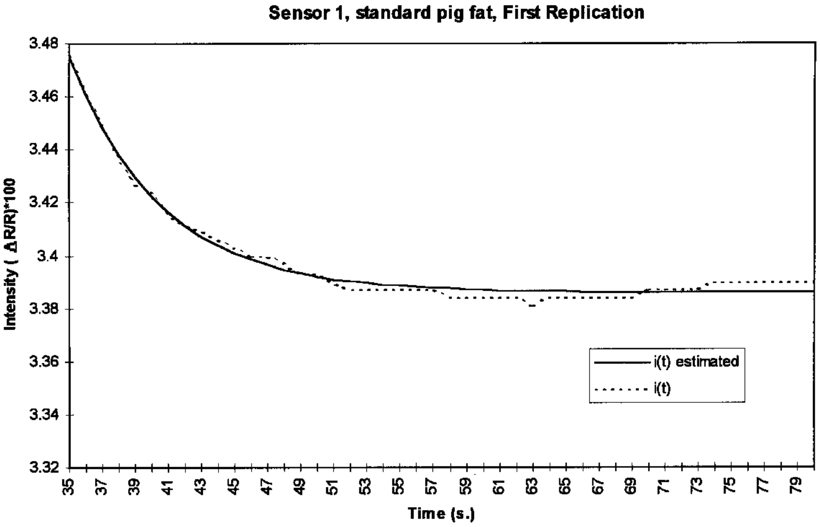

Fig. 10b. Estimated and real curve for the measurement on the first standard pig fat sample by sensor 1. Acquisition for the interval [35 s; 80 s].

5. Neaves P. I.; Hatfielf, J. V. *Sensors Actuat. B* **1995**, *26-27*, 223-23.
6. Gardner, J. W. *Sensors Act. B* **1991**, *4*, 109-115.
7. Gardner, J. W.; Bartlett, P. N. in: Olfaction and taste XI, Kurihara, K.; Suzuki, N.; Ogawa, H. Eds., Springer-Verlag, Tokyo, 1994; pp 690-693.
8. Mielle, P. *Trends Food Sci. Tech.* **1996**, *7*, 432-438.
9. Vernat-Rossi, V.; Vernat, G.; Berdague, J. L. *Analusis* **1996**, *24*, 309-315.
10. Eklov, T.; Martensson, P.; Lundström, I. *Anal. Chim. Acta* **1997**, *18343*, 1-9.
11. de Kermadec, F. Méthodes statistiques permettant d’expliquer l’appréciation hédonique par les caractéristiques sensorielles, Thèse de doctorat en science, Université Montpellier II, 1996.
12. Vackier, M. C.; Barros, A. S.; Rutledge, D. N. *J. Mag. Reson. Anal.* **1996**, 321-327.
13. Rutledge, D. N.; Barros,A. S.; Gaudard, F. *Mag. Reson. Chem.* **1997**, *35*, S13-S21.
14. Rutledge, D. N. *Analusis* **1997**, *25*(1), 9-14.
15. Persaud, K. C.; Khaffaf, S. M.; Pisanelli, A. M. *Measurem. Contr. **1996***, 17-20.
16. Persaud, K. C.; Qutob, A. A.; Travers, P.; Pisanelli, A. M.; Szyszko, S. in: Olfaction and taste XI Kurihara, K.; Suzuki, N.; Ogawa, H. Eds., Springer-Verlag, Tokyo, 1994; pp 708-709.
17. Aishima, T. *Anal. Chim. Acta* **1991**, *243*, 293-300.
18. Gardner, J. W.; Bartlett, P. N.; Pratt, K. F. E *IEE Proc.-Circuits Devices Syst*. **1995**, *142*(5) 321-333.