# DISCRIMINATION AND INTERPRETATION OF ELECTRONIC NOSE DATA USING ICA

*Oliver Tomic**

MATFORSK
Norwegian Food Research Institute
N-1432 Ås, Norway

*Martin Kermit*†

Agricultural University of Norway
Physics Section
N-1432 Ås, Norway

**ABSTRACT**

This work reports on independent component analysis (ICA) as a tool used to discriminate between odour signals. Measurements of six different alcohol solutions were carried out with a commercial gas sensor array system, a so called electronic nose. The solutions were made of either pure propanol or butanol at concentration levels of 0.5%, 1% and 2%. Principal component analysis (PCA), a standard multivariate analysis tool for gas sensor data, needed three principal components (PC) for effective discrimination of the solutions. With ICA, only two independent components (IC) were needed to achieve a similar result. PC-1 and IC-1 gave both meaningful representations of alcohol concentrations in the solutions. However, only a combination of PC2 and PC3 could represent different types of alcohols as effectively as IC2 did.

## 1. INTRODUCTION

With the term Electronic Nose (EN) is understood an array of chemical gas sensors with a broad and partly overlapping selectivity for measurements of volatile compounds combined with multivariate statistical data processing tools. EN belong to the category of rapid analysis instruments, allowing non-destructive analysis of vapours at a high rate with sufficient reproducibility and accuracy.

EN-instruments rely strongly on the recognition and analysis scheme applied to the data produced by its sensors. The need for adaptive recognition algorithms and dimensionality reduction in the data analysis part is key [1].

Thus, the combination of gas sensor array technology and multivariate data processing provides a powerful tool for applications within a broad range of different environments. Various EN applications were developed in the fields of environmental control [2], cosmetic industry [3], automobile industry [4], medical control [5], microbiology [6] and food industry [7, 8].

*E-mail: oliver.tomic@matforsk.no. Corresponding author
†E-mail: martin.kermit@hiof.no

The traditional and widely used method for analysis of EN data is principal component analysis (PCA) [9], with pattern recognition and data reduction being the primary objectives. However, PCA does not always provide the possibility for the user to take full advantage of the instruments potential. When using PCA for pattern recognition in EN data, discrimination sometimes may not be sufficient and thus, interpretation of data may be aggravated. In this article, independent component analysis (ICA) is suggested as a new method for processing EN data with great potential for improved performance of EN. It is suggested that ICA provides both increased dimensionality reduction and more meaningful interpretation of components when compared to PCA.

The article is organized as follows: The next section describes the experimental setup for the electronic nose and the data collected for this analysis. Section 3 reports on the results obtained when ICA and PCA are compared. The last section concludes this work.

## 2. MATERIALS AND METHODS

First described are the alcohol solutions used to generate the signals for the gas sensor array, which then is described next. Thereafter the recording process and data analysis are reported.

### 2.1. Solutions

Six different alcohol solutions were measured with the gas sensor array system. The solutions were: 0.5% propanol (P05), 1% propanol (P1), 2% propanol (P2), 0.5% butanol (B05), 1% butanol (B1) and 2% butanol (B2). Each solution was kept in 30 ml glass vials, with 10 ml of the particular solution. The glass vials were sealed with teflon coated silicon septa and open screw caps.

### 2.2. Gas sensor array

A commercial hybrid gas sensor array instrument manufactured by AppliedSensor Technologies, Linköping was used

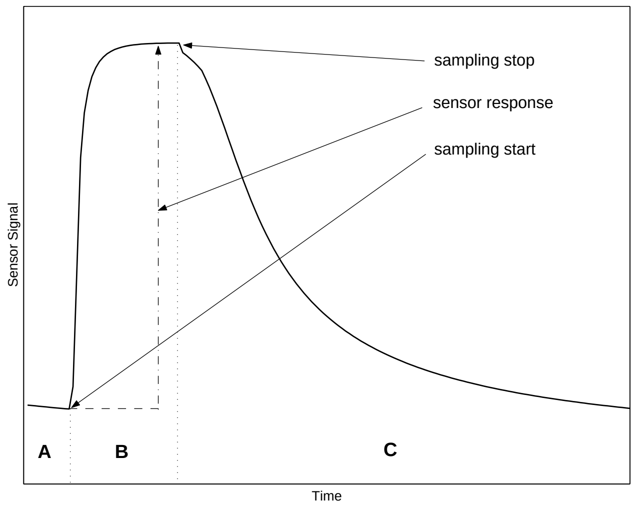


**Fig. 1**. Example of a measurement cycle for one gas sensor. The cycle consists of three phases: *A*: pre-sampling phase, *B*: sampling phase, *C*: recovery phase.

to perform the sample measurements. The gas sensor array, the main part of an EN, consisted of 22 gas sensors, 10 MOSFET[1] sensors and 12 MOS[2]. Each sensor possessed a unique sensitivity profile with varying sensitivity and selectivity towards certain volatile compounds.

Sensor data acquisition was accomplished by dynamic sampling of head-space gas, i.e. gas from the vial, saturated with volatiles from the sample, was pumped over the sensor array. While exposed to the analyte gas, the sensors generated an electronic output signal that was used to determine a sensor response value for the measured sample. Figure 1 displays a typical measurement process for one odour sample. In phase *A* of the measurement cycle, cleaned dry reference air passed over the sensor, providing the baseline signal. During phase *B* sensors were exposed to head-space sample gas from the glass vials and generated a change of electric output signal, triggered by interaction of volatile compounds with the sensor surfaces. By the start of phase *C*, reference air was led again over the gas sensor array for sensor recovery after the sampling process.

### 2.3. Measurements and sensor data

Senstool, the instrument's user software, allowed modification of phase duration in the measurement cycle. The total measurement cycle for the experiment was 200 sec; 20 sec phase *A*, 30 sec phase *B* and 150 sec phase *C*. Length of phase time was chosen according to optimal measurement condition for dynamic sampling. Enough time had to be provided for effective sampling, giving stable sensor signals, and sufficient sensor recovery thereafter. The measurement sequence consisted of 90 measurements, i.e. each of the solutions was measured 15 times. After completion of measurement sequence, a *M x N* data matrix was given, with $N$ representing 22 gas sensors in the array and $M$ representing 90 measurements.

### 2.4. Feature extraction

The sensor responses, i.e. the change in output signal corresponding to baseline (Figure 1) were used for data analysis. Other possible signal parameters not used in this study are on-derivate and integral corresponding to baseline in phase B as well as off-derivate in phase C. Sensor responses from only five gas sensors were used. Those from the remaining sensors were left out, since they did not provide enough variation or useful information in PCA.

### 2.5. Data analysis

The analysis of the data was implemented in Matlab 5.3 using built-in routines for the PCA and the FastICA Matlab package provided by Hyvärinen [10]. A dimensionality reduction was performed according to the criterion of using 99% of the explained variance from the PCA scores. This resulted in using the three largest eigenvalues, thus estimating three components for both PCA and ICA. FastICA was run using the deflation approach and $g(u) = u^3$.

## 3. RESULTS AND DISCUSSION

Figure 2 to 4 display the results of PCA and ICA after the methods were applied on the measurement data. Figure 2 illustrates that ICA clearly outperforms PCA in terms of discrimination of the measured solutions. In the two plots, discrimination between solutions with different concentration levels can be observed along both PC1 and IC1. Both components effectively represent alcohol concentration in the solutions. PC2 discriminates poorly between solutions of propanol and butanol, while IC2 provides a clear separation of the alcohols. IC3 does not contain valuable information as seen in Figure 3 and Figure 4. Obviously, only two IC's are needed for effective discrimination between the measured solutions. In the PCA plot in Figure 3 it can be observed that PC3 separates better between different alcohols than PC2 does in Figure 2. When using PC2 and PC3 as displayed in PCA plot in Figure 4, discrimination between the two alcohols can be obtained. That signifies that PCA needs three PC's for effective discrimination between the solutions.

[1]The MOSFET (metal-oxide-semiconductor field-effect transistor) gas sensors consist of three layers; a doped silicon semi-conductor, a thick oxide layer (SiO2) as insulator and on top a thin catalytic metal layer.

[2]The MOS are metal oxide semiconductor gas sensors. They consist of a metal-oxide semi-conducting film coated onto a ceramic core with an integrated heater.

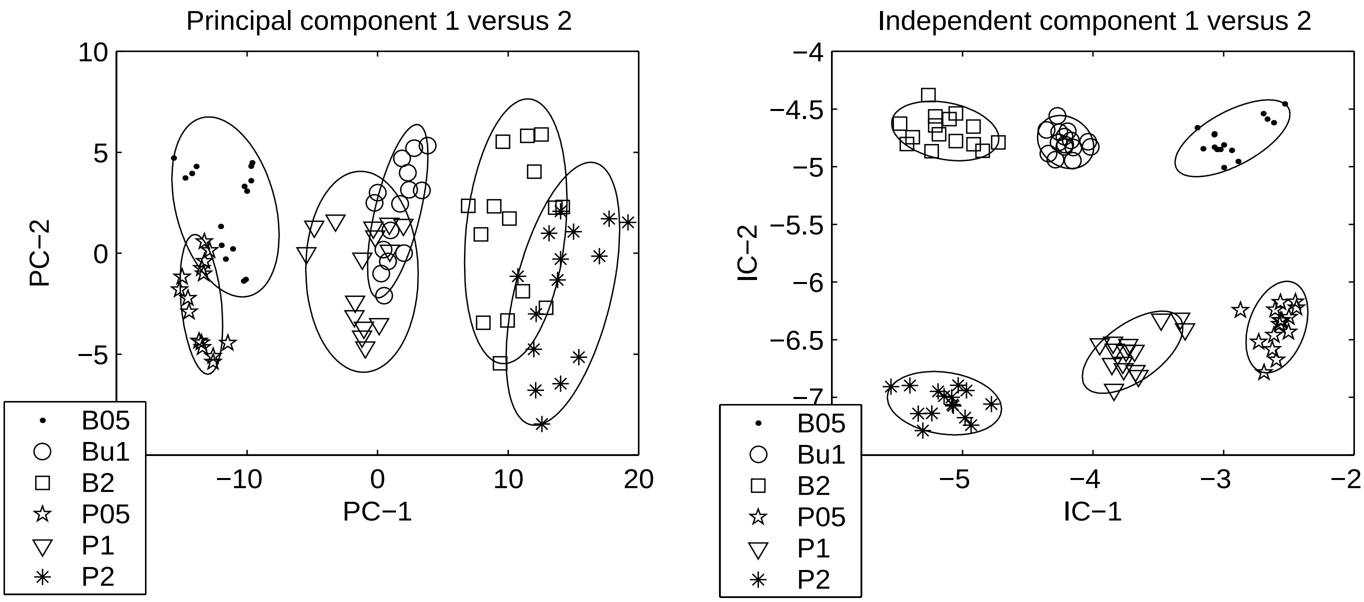


**Fig. 2**. *Left*: PCA scatter plot. Solutions of the same alcohol with increasing concentration are spread from left to right along PC1. PC2 provides little separation between the two alcohols. *Right*: ICA scatter plot. Solutions with increasing concentration are spread from right to left along IC1. IC2 clearly separates solutions based on different alcohols.

## 4. CONCLUSION

In order to make better use of EN's potential, the technology needs to be improved further. New data processing methods like ICA can be instrumental in achieving this goal, with advanced pattern recognition, data reduction and easier interpretations of measurement data compared to the standard EN data processing tool PCA. In this experiment, ICA outperformed PCA with increased discrimination between the six measured alcohol solutions, in terms of alcohol concentration and different type of alcohol while using less components. Within EN technology, ICA seems to be a promising new data processing tool that helps users make better use of the potential that the instruments provide in a wide range of environments.

## 5. REFERENCES

[1] T. Roppel, D. M. Wilson, Biologically-inspired pattern recognition for odor detection, Pattern Recognition Letters 21 (2000) 213–219.

[2] R. Stuetz, G. Engin, R. Fenner, Sewage odour measurements using a sensory panel and an electronic nose, Water Sci. Technol. 38 (1998) 331–335.

[3] Q. Lucas, N. Mignard, P. Dauneau, O. Dutheil, Innovative sensor array analysis of cosmetic raw materials, in: Proceedings of ISOEN 99, Tuebingen, Germany .

[4] S. Garrigues, T. Talou, D. Nesa, Quartz crystal microbalance sensors based electronic nose for qc in the automotive industry, in: Proceedings of ISOEN 2000, Brighton, UK .

[5] J. Stetter, W. Penrose, C. McEntegart, R. Roberts, Prospects for infectious disease diagnosis with sensor arrays, in: Proceedings of ISOEN 2000, Brighton, UK .

[6] T. Bachinger, U. Riese, R. Eriksson, C. Mandenius, Monitoring cellular state transitions in a production scale cho-cell process using an electronic nose, Journal of Biotechnology 76 (2000) 61–71.

[7] J. Haugen, Headspace analysis of foods and flavours: Theory and practice, Kluwer Academic/Plenum Publishers, New York, 2001.

[8] J. Haugen, O. Tomic, F. Lundby, K. Kvaal, E. Strand, L. Svela, K. Jorgensen, Analysis of off-flavours in raw cow's milk with a commercial gas-sensor system, Electronic Noses and Olfaction 2000 .

[9] J. Hertz, A. Krogh, R. G. Palmer, Introduction to the theory of Neural Computation, Addison-Wesley Publishing Company, 1991, the Advanced Book Program.

[10] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Transactions on Neural Networks 10 (3) (1999) 626–634.

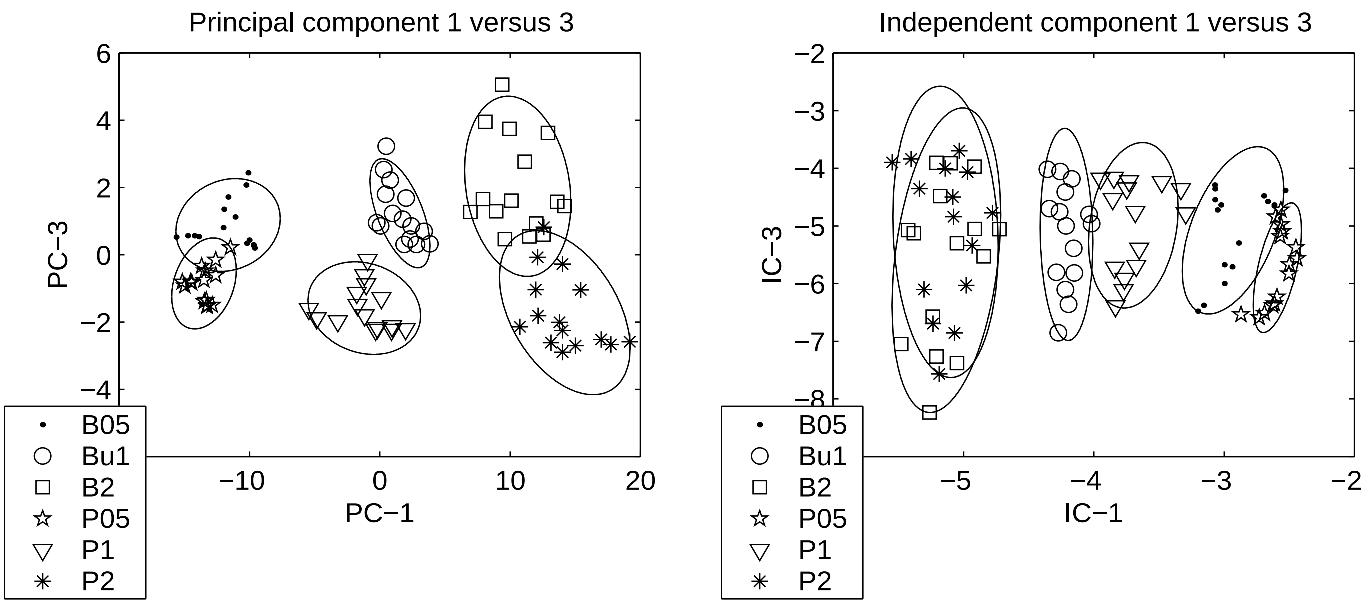


**Fig. 3**. *Left*: PCA scatter plot using the first and the third principal components. PC3 provides little separation between solutions based on different alcohols *Right*: ICA scatter plot using the first and the third independent components. IC3 contains no useful information.

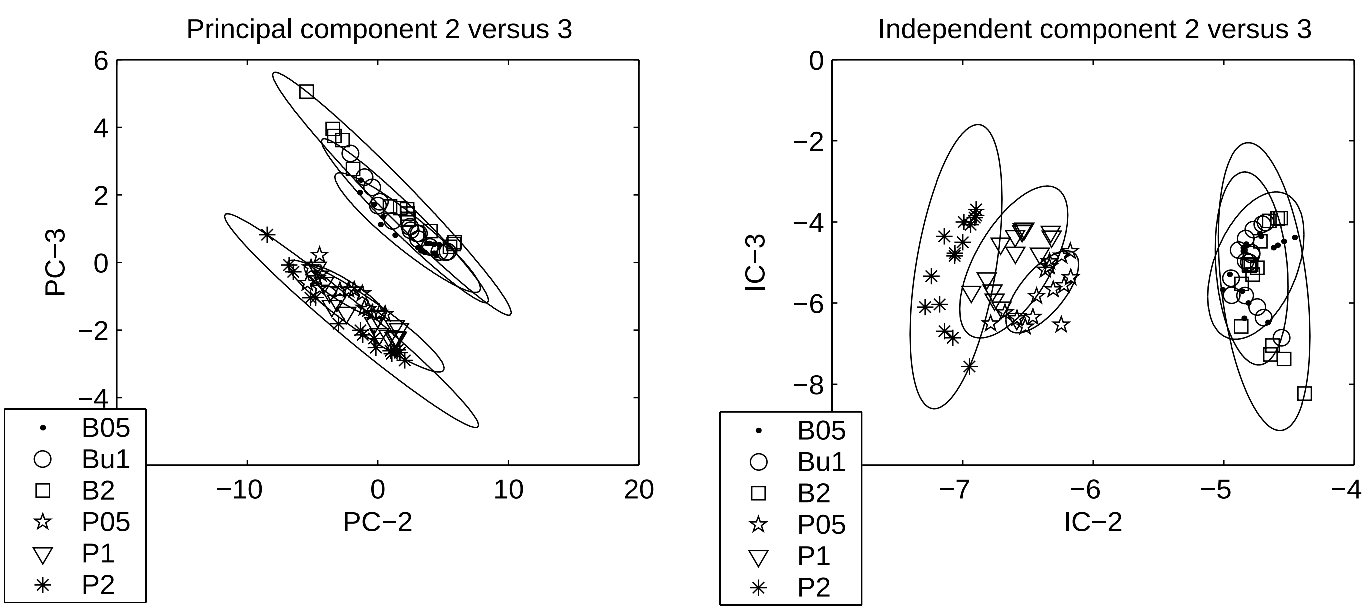


**Fig. 4**. *Left*: PCA scatter plot using the second and the third principal components. PC2 and PC3 are needed for effective discrimination between the two alcohols *Right*: ICA scatter plot using the second and the third independent components. IC2 alone efficiently discriminates between the alcohols.