Three-way electronic nose data

Thomas Skov & Rasmus Bro
The Food Technology group, a section of the Department of Food Science, KVL, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

Problem
These data was obtained in order to investigate if bad licorices could be differentiated from good licorices.

The problem of bad licorices was revealed through several complaints from consumers that reported a burned taste of the licorices. The licorice company therefore initiated an investigation to find a method that could distinguish between bad and good licorices. A third group of fabricated bad licorices (licorices dried for a longer time) were included in the data set, to mimic the observed burned taste of the bad licorices.

An electronic nose combined with multivariate tools (chemometrics) showed applicable for this purpose. Due to the time-consuming data analysis with several pre-processing steps and model investigations an innovative GUI (Graphical User Interface), 'SENSABLE - Analysis of sensor based data', was made.

**Get the data from** http://www.models.kvl.dk

The data is available in MATLAB 6.5 format. Download the data and write load data in MATLAB to inspect the data. The data is given in a structure array (NoseData) which contains further information about the data.

If you use the data we would appreciate that you report the results to us as a courtesy of the work involved in producing and preparing the data. Also you may want to refer to the data by referring to:

T. Skov and R. Bro. *A new approach for modelling sensor based data*, Submitted for publication (2004)

Data
The $\underline{\mathbf{X}}$ matrix is arranged as: Samples $\times$ Time $\times$ Sensor. See Figure 1.
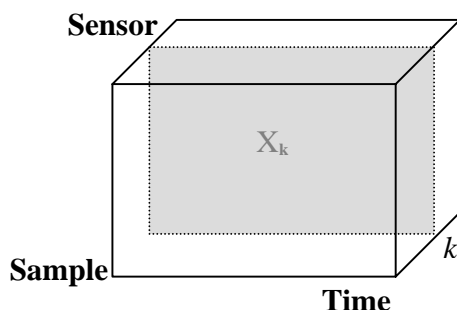
Figure 1. Structure of data from the electronic nose, with the *k* slab (*k* sensor) shown.

The sample mode consists of 6 good licorice samples, 6 bad licorice samples and 6 fabricated bad licorice samples. The time mode is a continuous time scale where the sensor signal has been measured over 120 sec every ½ sec. The first time is the baseline signal (i.e. signal of carrier gas). Twelve sensors, all based on Metal Oxide Semiconductor (MOS) technologies, were used to register the volatile compounds from the samples.

This means that the size of $\underline{\mathbf{X}}$ will be $18 \times 241 \times 12$.

Y-matrix: Because we know that the samples belong to three specific groups a discriminant variable is included as a kind of external information. Ideally this is done to maximize the separation between the groups and to minimize the variation within each group.
A typical response of the twelve sensors from a general sample is shown in Figure 2.
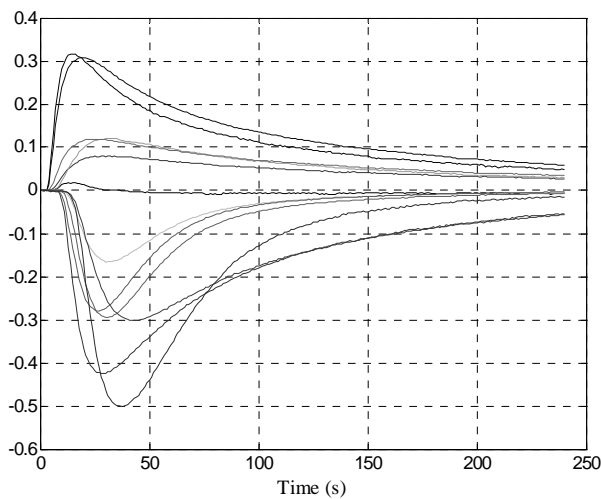
Figure 2. Baseline corrected sensor signal $((S-S_0)/S_0)$ for one licorice sample for the twelve sensors.