Continuous Chemical Classification in Uncontrolled Environments with Sliding Windows

Javier G. Monroy[a,*], Esteban J. Palomo[b,c], Ezequiel López-Rubio[b], Javier Gonzalez-Jimenez[a]

[a] *MAPIR-UMA Group, Dept. System Engineering and Automation. University of Málaga, 29071 Málaga, Spain*
[b] *Department of Computer Languages and Computer Science. University of Málaga. Bulevar Louis Pasteur, 35. 29071 Málaga, Spain*
[c] *School of Mathematical Science and Information Technology. University of Yachay Tech. Hacienda San José s/n. San Miguel de Urcuquí. Ecuador.*

Abstract

Electronic noses are sensing devices able to classify chemical volatiles according to the readings of an array of non-selective gas sensors and some pattern recognition algorithm. Given their high versatility to host multiple sensors while still being compact and lightweight, e-noses have demonstrated to be a promising technology to real-world chemical recognition, which is our main concern in this work. Under these scenarios, classification is usually carried out on sub-sequences of the main e-nose data stream after a segmentation phase which objective is to exploit the temporal correlation of the e-nose's data. In this work we analyze to which extent considering segments of delayed samples by means of fixed-length sliding windows improves the classification accuracy. Extensive experimentation over a variety of experimental scenarios and gas sensor types, together with the analysis of the classification accuracy of three state-of-the-art classifiers, support our conclusions and findings. In particular, it has been found that fixed-length sliding windows attain better results than instantaneous sensor values for several classifier models, with a high statistical significance.

Keywords: chemical classification, artificial olfaction, e-nose, online, sliding window, chemical sensor array

*Corresponding author
Email addresses: jgmonroy@uma.es (Javier G. Monroy), ejpalomo@lcc.uma.es (Esteban J. Palomo), ezeqlr@lcc.uma.es (Ezequiel López-Rubio), javiergonzalez@uma.es (Javier Gonzalez-Jimenez)

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

1. Introduction

An electronic nose (e-nose) is a device intended to detect and recognize different chemical substances by means of an array of gas sensors and some kind of pattern recognition algorithm. It has been successfully employed for breath analysis in clinical environments [Gelperin and Johnson, 2008; Guo et al., 2010], animal nutrition [Campagnoli and Dell'Orto, 2013], autonomous robotics [G. Monroy et al., 2015; Loutfi et al., 2005; Schleif et al., 2015], or volatile quantification [Dentoni et al., 2012; G. Monroy et al., 2013], among other applications. Yet, due to practical difficulties, mostly related to the lack of robustness of calibration models, most e-nose systems have only been employed under laboratory environments, being still one step back to be deployed in real-world applications.

The classification of volatile substances is, possibly, the most studied application of e-noses among the scientific community. Traditionally, this has been performed by analyzing the response of the array of gas sensors (providing multiple signals) when exposed to pulse-like gas excitations under well-controlled measurement conditions (i.e. temperature, humidity, exposure time, etc.). Unsurprisingly, dozens of papers report less than 10% classification error rate on this problem (e.g. see [Schaller et al., 1998; Szczurek et al., 2013; Vergara et al., 2012]). However, when the classification is to be performed on a real, uncontrolled scenario, assumptions such as a perfect alignment or equally length of patterns, do not hold [Hu et al., 2013]. This, which is due to the dynamic and chaotic nature of gas dispersal, together with the strong dynamics shown by most gas sensor technologies, notably increases the complexity of the classification problem. To better illustrate this problem, Figure 1 shows the readings of an e-nose composed of an array of 4 metal oxide-semiconductor (MOX) gas sensors when exposed to an ethanol gas source under controlled and uncontrolled environmental conditions. As can be seen, when controlling the environment, reproducible patterns are obtained and key-points can be easily identified in the response: (a) start of the volatile exposition, (b) end of the transient response and start of the steady phase, and (c) end of the volatile exposition and start of the recovery period. In contrast, when no control is performed over the environment (including the gas-emitting source), signals look much more random and chaotic, being difficult to identify distinctive behaviors or patterns.

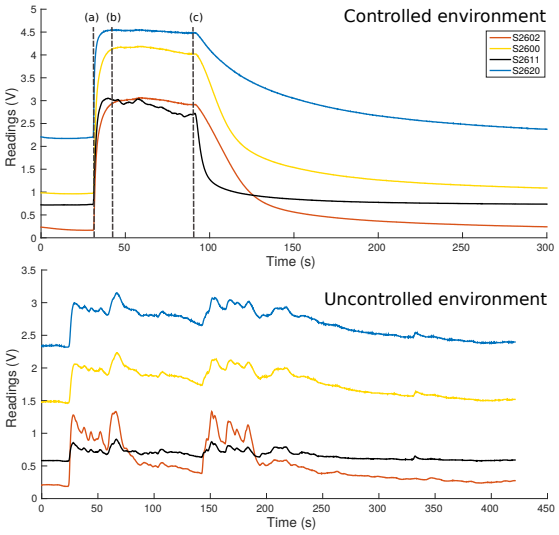

Figure 1: Readings of an e-nose exposed to an ethanol gas source under: (top) well controlled environment and measuring conditions employing a pulse-like excitation, and (bottom) when no control is performed during the measurements. As can be noticed, the regular and well defined patterns on the controlled environment are not present in the uncontrolled case.

Moreover, applications such as the detection of toxic chemicals in human environments, or the localization of gas sources by robots, demand a continuous classification of volatile substances which can not be addressed with the traditional pulse-like excitation procedure. For these applications, chemical recognition can be seen as a particular case of time series classification, characterized by working on sub-sequences of the main data stream. Many are the algorithms proposed in literature for time series segmentation (see [Fu, 2011] for a complete review), ranging from perceptually important points (PIP) to special events detection. Nevertheless, most of these approaches are proposed for uni-variate time series, while e-nose data is fundamentally multi-variate, since its based on an array of gas sensors with dif-

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

ferent dynamic responses. This, together with the aforementioned challenges of real data, make most segmentation approaches difficult to apply to e-nose data, which, in turn, affect negatively the classification rate.

In this article we advocate the use of the well known sliding window approach to avoid feature-based segmentation and study to which extent considering delayed samples contributes to exploit the temporal correlation of e-nose's data. Sliding window can be considered as both an incremental and decremental technique, where in the incremental stage the effect of the newest data is added while in the decremental stage the effect of the oldest data is omitted [He et al., 2015]. This technique is attractive because it is simple and intuitive. Moreover, it is amenable to online applications. We will study the impact of its main parameter, the window length, on the accuracy of three state of the art classifiers, namely: naive Bayes classifiers, decision trees and extreme learning machines. Comparison against the classifier's static version that just account for the instantaneous response of the sensors (window length of one sample) will be provided for a variety of experimental scenarios, e-nose configurations and gas classes (employing three different olfaction data sets). In particular, we carry out two different analyses: first, we study how the window length affects the classification accuracy, by employing a wide range of length values and comparing the resulting performance. Then, we conduct an hypothesis test to see if the classifier's static version and the sliding window version perform equally well, or there is an improvement when employing the latter. Furthermore, a two stages cross-validation method is proposed to provide statistically representative results, concluding that, for online chemical classification in uncontrolled environments, feeding the classifiers with additional delayed samples leads to a small, yet important, improvement (up to 6% units) on the classification accuracy.

The rest of this document is organized as follows. Current approaches to chemical classification with e-noses are reviewed in Section 2. After that, Section 3 introduces the feature extraction based on sliding windows, and Section 4 a variant of the cross validation methodology to cope with the data characteristics of real time odor classification. Then, Section 5 presents extensive experimentation results from which useful conclusions are drawn, and we conclude in Section 6 with a discussion of the results and propose future work.

2. Related Work

This section reviews relevant works on the classification of chemical volatiles under real, uncontrolled environments, and concretely focuses on those aimed at a continuous, real time classification.

Probably, the first work dedicated to this problem was presented by Trincavelli and coauthors [Trincavelli et al., 2009]. Here they proposed a segmentation stage based on the detection of three different phases in the sensors response (transient, steady state and recovery). The authors evaluated different algorithms and feature extraction approaches, concluding that, provided that these phases are reliably identified, online classification of chemical volatiles is feasible.

In [Hatami and Chira, 2013] a set of serially located classifiers with a reject option is proposed for online decision making. The idea is to iteratively classify a short data sequence, or reject the sample leaving it to the next classifier which considers a longer sequence. A set of support vector machine classifiers with radial basis function kernels were used to classify 10 different odors in a wind tunnel experiment, reaching a maximum performance of 94.1% with data sequences of 10 seconds. The main differences with the work presented here are the use of an ensemble of classifiers based on voting agreement instead of a single classifier, and the consideration of a dynamic window of increasing length based on ad hoc thresholds.

More recently, in [Schleif et al., 2015], a generative topographic mapping through time (GTM-TT) is proposed as an unsupervised model for time series odor classification. The work analyzes the impact of the data sequence length on the classification performance, as well as introduces a relevance measure of each sensor in the array. Nonetheless, such analysis is only performed over signals collected in a

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

semi-controlled experimental setup (fixed time exposure and distance to the source). Furthermore, their gas classification framework is fundamentally different from the present proposal because they train an unsupervised neural network per class. Then a given test sample is assigned to the class associated to the network which best represents the sample. On the other hand, our proposal trains a single supervised learning system with samples coming from all the existing classes.

In [Fonollosa et al., 2015], a reservoir computing algorithm (RC) is used to identify and quantify chemicals of interest in an online fashion. A large network of recurrent randomly connected nonlinear units (i.e. neurons) is used to transform the e-nose signals to a higher dimension. Then, the outputs of such a network are used as features of a linear classifier. Given the recurrent connections existent within the neurons of the pool, this approach is able to handle the temporal information carried out by the e-nose data stream without resorting on data sequences. The main difference with the proposed approach is the fact that the dynamic memory inherent of RC algorithms is hidden to the user (because these connections are not trained, but set randomly at the creation of the neurons' network), while in this research work we pursue the analysis of the memory length's impact on the classification accuracy.

Later, in [G. Monroy and Gonzalez-Jimenez, 2015], a sequential Bayesian filtering (SBF) approach is proposed to integrate information from previous e-nose observations without relying on the segmentation of the data stream. Only the most recent e-nose observation is used at each time step for estimating the chemical category, being the filter in charge of refining the class posteriors based on previous estimates.

In [De Vito et al., 2015] a multivariate regression for on-field calibration of e-noses aiming at pollution monitoring is proposed by means of dynamic neuronal networks. The authors tested the approach using a one minute sampling frequency data set gathered with a conventional analyzer over a five weeks period. The main improvement with respect previous works on this research area is the adaptation to dynamic data (since common air-pollution data sets only account for hourly averaged values), extending the static multivariate on-field approach previously proposed by some of the authors [Vito et al., 2009].

Finally, approaches like dynamic time warping (DTW), commonplace in standard time series classification [Niennattrakul et al., 2012], are not easily applicable to chemical discrimination due to the data inconsistencies caused by both the transient response of gas sensors, and the different responses appreciated when exposed to several gases in a sequence.

3. Feature Extraction from Sliding Windows

Generally speaking, a sliding window is a data interval that runs over a larger data sequence (e.g. the temporal data sequence delivered by the e-nose). In this work, we employ fixed-length sliding windows with overlapping time intervals, this way, the segmentation problem is overcome without resorting to any shape analysis of the sensory signals [Niennattrakul et al., 2012], which, as pointed out in the introduction, is not an easy or reliable task due to noise, the sensor dynamics and the cathodic nature of gas dispersal (as illustrated in Figure 1).

The choice of the window length, the most determinant parameter of this technique, involves a trade-off between two factors. On the one hand, narrow windows imply small data sequences from which to exploit the temporal correlation of the data. On the other hand, long windows increase the chance that the gas class has changed over the time period covered by the window, so that the oldest samples are no longer representative of the current gas class. We will deal with this issue in the experimental section, analyzing the impact of the window length for a number of classifiers and scenarios.

Independently of the window length, for each data interval we analyze here how to extract simple though robust features that can be used to classify chemical volatiles in an online fashion. The set of features considered in this work (mean and variance) are introduced in Subsections 3.1 and 3.2, respectively.

Notice that more elaborated features can be extracted from the sliding window, for example, employing curve fitting [Carmel et al., 2003], dynamic principal component analysis [Perera et al., 2006], or wavelet coefficients [Llobet et al., 2002] (see [Yan

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

et al., 2015] for a complete review). Yet, our objective in this work is not to optimize the set of features which achieve the higher classification rate, but to experimentally validate the suitability of a sliding window approach, as well as to analyze the influence of its length, for which the simple mean and variance features are sufficient.

3.1. Mean of the Sliding Window

Let us consider a continuous data signal $s_i(t)$ provided by each of the gas sensors of the e-nose array $i \in \{1, ..., G\}$. We assume that each observed signal is the true, noiseless sensor signal $\hat{s}_i(t)$ affected by a zero mean error term $\epsilon_i(t)$ that models all the sensor inaccuracies:

$$s_i(t) = \hat{s}_i(t) + \epsilon_i(t) \tag{1}$$

$$E[\epsilon_i(t)] = 0 \tag{2}$$

Let's also consider a sliding window of size W over $s_i(t)$, and the mean feature m_i at time instant t given by:

$$m_i(t) = \frac{1}{W} \int_{t-W}^{t} s_i(\tau)\, d\tau \tag{3}$$

In practice, the signal s_i is sampled at fixed time intervals T, thus, the window mean is approximated by:

$$\bar{m}_i(kT) = \frac{1}{N+1} \sum_{j=0}^{N} s_i(kT - jT) \tag{4}$$

where k is the discrete temporal index, and $N = \frac{W}{T}$ is the window size measured in samples.

The expected value of $\bar{m}_i(kT)$ matches the mean value of the underlying, noiseless signal:

$$E[\bar{m}_i(kT)] = \frac{1}{N+1} E\left[\sum_{j=0}^{N} s_i(kT - jT)\right] =$$

$$\frac{1}{N+1}\left(E\left[\sum_{j=0}^{N} \hat{s}_i(kT - jT)\right] + E\left[\sum_{j=0}^{N} \epsilon_i(kT - jT)\right]\right) =$$

$$E\left[\frac{1}{N+1} \sum_{j=0}^{N} \hat{s}_i(kT - jT)\right] \tag{5}$$

where we have applied (2) in the last step. This means that $\bar{m}_i(kT)$ conveys information about the average value of the underlying noiseless sensor signal in the short term near the time instant $t = kT$. In order to measure how close $\bar{m}_i(kT)$ is to the true, noiseless signal $\hat{s}_i$ we can compute the variance of $\bar{m}_i(kT)$ as the average of $N+1$ correlated random variables:

$$\mathrm{var}[\bar{m}_i(kT)] = \mathrm{var}\left[\frac{1}{N+1} \sum_{j=0}^{N} s_i(kT - jT)\right] =$$

$$\frac{1}{(N+1)^2}\left[\sum_{j=0}^{N} \mathrm{var}[s_i(kT - jT)]\right] +$$

$$\frac{2}{(N+1)^2}\left[\sum_{j<l} \mathrm{cov}[s_i(kT - jT), s_i(kT - lT)]\right] \tag{6}$$

The variance of the observed variables $s_i(kT)$ represents the noise level of the sensors. Let us note σ^2 the maximum variance among all observed variables:

$$\sigma^2 = \max\{\mathrm{var}[s_i(kT)] \mid i \in \{1, ..., G\}\} \tag{7}$$

Taking ρ as the average correlation of distinct variables, from (6) and (7) we can obtain an upper bound for the variance of the average features:

$$\mathrm{var}[\bar{m}_i(kT)] \leq \frac{\sigma^2}{N+1} + \frac{N}{N+1}\rho\sigma^2 \tag{8}$$

A sensible goal is the minimization of $\mathrm{var}[\bar{m}_i(kT)]$ with respect to N, so as to make the computed average features $\bar{m}_i(kT)$ as close to $\hat{s}_i$ as possible. As

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

seen in (8), this depends on ρ and σ^2. If σ^2 is small with respect to ρ, i.e. the observed signal $s_i(kT)$ has a small variability, then the term $\frac{N}{N+1}\rho\sigma^2$ dominates and smaller values of N are better. On the other hand, if ρ is small with respect to σ^2, i.e. the observed signal $s_i(kT)$ exhibits a large variability, then the term $\frac{\sigma^2}{N+1}$ is the dominant one and larger values of N are more advisable. For a system with a fixed sampling interval T this means that there is a value of the sliding window size N that drives $\bar{m}_i(kT)$ as close as possible to the short term mean value of $\hat{s}_i$ that depends on the variability of the observed signal s_i. Large values of N work better for signals with a large variability, and vice versa.

3.2. Variance of the Sliding Window

Since $\bar{m}_i(t)$ does not capture the short term variability of the sensor signal, in order to obtain features that provide information about such variability, we may resort to the variance of the sensor data s_i with respect to $m_i(t)$:

$$v_i(t) = \frac{1}{W}\int_{t-W}^{t} \left(s_i(\tau) - m_i(t)\right)^2 d\tau \tag{9}$$

so that, for a constant signal $s_i(t) = K$, we have $v_i(t) = 0$.

As before, for a sampled signal, this is implemented by averaging at the sampling instants:

$$\bar{v}_i(kT) = \frac{1}{N+1}\sum_{j=0}^{N} \left(s_i(kT - jT) - \bar{m}_i(kT)\right)^2 \tag{10}$$

This way, a feature vector $\mathbf{x}(kT) = \{\bar{m}_1(kT), ..., \bar{m}_G(kT), \bar{v}_1(kT), ..., \bar{v}_G(kT)\}$ can be built from the set of sensor signals $s_i(kT)$, and supplied to a classifier.

4. Sliding Window Cross Validation

Cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. For the case of classification, this involves splitting the available set of samples into three disjoint subsets [Simoes et al., 2014][Zhao et al., 2015]:

- The training set, which is supplied to the classifiers.
- The validation set, which is used to measure the performance of the available classifiers with different sets of tunable parameters, if applicable. Then the classifier and parameter set which yields the best performance over the validation set is chosen.
- The test set, which is used to report the performance of the chosen classifier.

In most practical cases, the split is repeated several times and the chosen classifier and parameter set are those which yield the best average validation performance over the repeated trials. A key underlying assumption of this scheme is that the data samples are independent and identically distributed (i.i.d.), as highlighted in [Molinaro et al., 2005; López-Rubio and Ortiz-de Lazcano-Lobato, 2009]. Thus, given N i.i.d. observations $\mathbf{x}_1, ..., \mathbf{x}_N$ of the feature vector $\mathbf{x}$ with unknown distribution P, the task is to choose a suitable predictor $\psi : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ denotes the set of possible feature vectors and $\mathcal{Y}$ is the set of all classes. We note $\psi(\mathbf{x} \mid P_N, \boldsymbol{\theta})$ to highlight that the prediction depends on the empirical distribution P_N and the specific classifier and its tunable parameters, collectively represented by $\boldsymbol{\theta}$. Also, let y note the actual class. Under these conditions, a common performance measure of classifier $\boldsymbol{\theta}$ is the classification accuracy:

$$\mathcal{A}_{\boldsymbol{\theta}} = \int \mathbb{I}\left(y = \psi(\mathbf{x} \mid P, \boldsymbol{\theta})\right) dP(\mathbf{x}, y) \tag{11}$$

where $\mathbb{I}$ is the indicator function which takes a Boolean value as argument and returns a real number as follows:

$$\mathbb{I}(true) = 1 \text{ and } \mathbb{I}(false) = 0 \tag{12}$$

The quantity $\mathcal{A}_{\boldsymbol{\theta}}$ depends on the unknown data distribution P, so in practice a cross validation procedure is used to split the observed data set into

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

training, validation and test sets with empirical distributions P_N^{Train}, P_N^{Val} and P_N^{Test}, respectively. The average classification accuracy on the validation set can be used to choose a classifier with its parameter set:

$$\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \int \mathbb{I}\left(y = \psi\left(\mathbf{x} \mid P_N^{Train}, \boldsymbol{\theta}\right)\right) dP_N^{Val}(\mathbf{x}, y) \tag{13}$$

Then the reported average accuracy is computed on the test set:

$$\hat{\mathcal{A}}_{\tilde{\boldsymbol{\theta}}} = \int \mathbb{I}\left(y = \psi\left(\mathbf{x} \mid P_N^{Train}, \tilde{\boldsymbol{\theta}}\right)\right) dP_N^{Test}(\mathbf{x}, y) \tag{14}$$

For electronic olfaction this scheme could be used under controlled laboratory conditions where characteristic features that form the samples to supply to the classifiers are extracted from independent experiments, so that each experiment produces only one sample. However, when facing the classification of real time olfaction data, standard cross-validation can not be applied because the i.i.d. assumption would be violated. This is because under continuous operation the features extracted from different time intervals belonging to the same experiment are correlated due to the specific circumstances of each experiment.

Consequently, we need to impose the constraint that all the samples extracted from the same experiment must be put into the same set (training, validation or test), that is, perform cross validation at experiment level.

Let M be the number of available experiments e_i, each with a possibly different number of samples $N(e_i)$, $i \in \{1, ..., M\}$. Then, the M experiments are split into training, validation and test sets with empirical distributions Q^{Train}, Q^{Val} and Q^{Test}, respectively. The new classifier selection procedure is given by:

$$\bar{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \int A(e, \boldsymbol{\theta})\, dQ^{Val}(e) \tag{15}$$

$$A(e, \boldsymbol{\theta}) = \int \mathbb{I}\left(y = \psi\left(\mathbf{x} \mid Q^{Train}, \boldsymbol{\theta}\right)\right) dP_{N(e)}(\mathbf{x}, y) \tag{16}$$

where $P_{N(e)}(\mathbf{x}, y)$ is the empirical distribution of the $N(e)$ samples which belong to experiment e. Finally the reported accuracy is obtained as follows:

$$\hat{\mathcal{A}}_{\bar{\boldsymbol{\theta}}} = \int A(e)\, dQ^{Test}(e) \tag{17}$$

5. Experiments

This section is devoted to evaluate the impact of the sliding window length on the classification of chemical volatiles when carried out in real environments. The objective is to analyze if by increasing the window length (i.e increasing the number of delayed e-nose samples to process) we can better exploit the temporal correlation of the data, and consequently improve the classification accuracy. To this aim, we carry out two different analyses: first, we study the classification accuracy for a wide range of window lengths in order to get an overview and detect possible trends. Then, we perform an hypothesis test to see if the probability distribution of the results derived from employing sliding window with an optimal window length ($N > 1$) differs from that obtained using only one sample ($N = 1$). First of all, the Lilliefors test is used to find out whether the performance results coming from each of the two approaches can be assumed to be Gaussian. If both probability distributions are judged to be Gaussian, then we employ the paired t-test, which returns a test decision for the null hypothesis that the observed difference in the performance comes from a Gaussian distribution with zero mean and unknown variance. If the data for any of the two approaches is not judged to be Gaussian distributed, then we employ the Wilcoxon test which checks the null hypothesis that data come from continuous distributions with equal medians.

Three olfaction data sets composed of real e-nose experiments are considered in this section (see Section 5.1 for a detailed description of each data set).

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

These data sets provide a wide variability with respect the number and type of gas sensors, the environmental and measuring conditions, or the nature of gases to be recognized, something fundamental to obtain pertinent conclusions. Furthermore, a set of three well-known classifiers: naive Bayes Friedman et al. [1997], decision trees Rokach and Maimon [2008] and extreme learning machines (ELM) Huang et al. [2006], are used in this analysis to provide comparison. More precisely, for decision trees the Gini's diversity index is used as the criterion to choose whether to split at a given node [Breiman et al., 1984; Coppersmith et al., 1999].

Feature extraction is performed as stated in Section 3, therefore, input samples are represented by computing their mean $\bar{m}_i$ and variance $\bar{v}_i$ features for each sensor i in the e-nose array, resulting in a feature vector of dimension twice the number of sensors used in the data set. That is, supposing an e-nose composed of G gas sensors, the feature vector extracted from each sliding window would be $\mathbf{x} = [\bar{m}_1, ..., \bar{m}_G, \bar{v}_1, ..., \bar{v}_G]$.

Moreover, fourteen different window's lengths are analyzed, ranging from the static case scenario where only the most recent e-nose observation is used to predict the gas class (one-sample window's length), up to a window of 14 seconds. This maximum length has been chosen empirically, having in mind that long data sequences increase the chance that the gas class has changed over the time period covered by the window. Thus, although 14 seconds is, in practice, a quite wide window, we provide here the results to study the tendency of the classification accuracy.

To guarantee the statistical relevance of the results, in this work we carry out a double cross validation process. First, we perform cross validation at experiment level as proposed in Section 4, for 20 times. The exact data partition varies with the performed analysis: for the overview study, we employ 80% for training and 20% for testing (since there is no parameter to optimize, we do not have any validation set), while for the hypothesis test we employ 60% for training, 20% for validation and 20% for testing. Then, for each partition, we perform again cross validation, randomly selecting 80% of the data. This second phase is repeated ten times. To better understand this process, we provide the pseudo-codes detailing the steps involved in each of the two analyses in Algorithms 1 and 2, respectively.

Algorithm 1 Pseudo-code of the Overview Analysis

```
for each window length do
    for i=1..20 do                ▷ Cross validation at experiment level
        Data partition (80% train, 20% test)
        Feature extraction
        for j=1..10 do            ▷ Cross validation at sample level
            Data partition (80% of original partitions)
            Train classifiers
            Get accuracy for test subset
        end for
        Get average accuracy (j=1..10)
    end for
    Get overall average accuracy (i=1..20)
end for
```

A description of the data sets, detailing the number of sensors that comprise the e-nose array, the number of gas classes to discriminate, as well as the measuring conditions, are described in Subsection 5.1. Then, experimental results for each data set are provided in Subsections 5.2, 5.3 and 5.4.

5.1. Data Sets Description

Three publicly available data sets of olfactory time series have been employed in this work. A short description detailing the environmental conditions, and the number of classes and sensors is provided next. Furthermore, for the sake of comparison, a summary of their main characteristics is provided in Table 1.

Table 1: Summary of the Three Olfaction Data Sets Used in This Work.

Data Set	Classes	Sensors	Environment
UMA	4	4	Controlled - gas pulses
UCI-1	2	8	Uncontrolled - wind tunnel
UCI-2	10	8	Uncontrolled - wind tunnel

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

Algorithm 2 Pseudo-code of the Hypothesis Test Analysis

```
for i=1..20 do ▷ Cross validation at experiment level
    Data partition (60% train, 20% validation, 20% test)
    for each window length do
        Feature extraction
        for j=1..10 do ▷ Cross validation at sample level
            Data partition (80% of original partitions)
            Train classifiers
            Get accuracy for validation subset
        end for
        Get average accuracy (j=1..10)
    end for
    Get optimal window length (max validation accuracy)
    Get accuracy for test subset using one-sample
    Get accuracy for test subset using the optimal window-length
end for
Get overall average accuracies (i=1..20) for the one-sample and optimal window-length cases
Lilliefors test to check for Gaussianity
t-test or Wilcoxon test to check for significant performance differences between one-sample and optimal window-length
```

- **Data set UMA**: This data set aims at providing a reference to the chemical classification under well controlled measuring conditions. To this end, we consider here a subset of the MAPIR-UMA olfaction data set[1]. Concretely, we account a total of 69 time series gathered when exposing an e-nose to gas pulses of four different analytes: acetone, ethanol, butane and alcoholic beverages[2]. From the available gas sensors in the e-nose array, we select four, namely TGS2602, TGS2600, TGS2611 and TGS2620, which are sampled at a frequency of 7Hz. To illustrate the type of signals in this data set, Figure 2-top plots two out of the 69 time series in it, corresponding to two different gas classes. As can be seen, because the semi-controlled measuring conditions, the sensors' responses follow step-like shapes, which could be considered as a pattern. However, for the purpose of this work, we will apply the sliding window approach similarly to other data sets, ignoring the patterns present on the data. Concretely, we will consider samples in the time interval [30,180] seconds, leaving out those before and after because they carry little information about the gas identity.

- **Data set UCI-1**: This olfaction data set from the UCI database (see Fonollosa et al. [2014]), is based on 180 time series recorded in a wind tunnel facility at a sampling rate of 10Hz. Two different gas mixtures are presented to an e-nose composed of eight MOX gas sensors. Since in this work we do not account for mixtures of gases, we consider this data set as a two-classes problem, that is, we will identify class 1 as the mixture of ethylene and methane, and class 2 as the mixture of ethylene and carbon monoxide. Taking into account the initial baseline period of the time series of this data set (60 seconds approximately), we will only process samples in the time interval [80s,290s], also removing samples long after the gas exposure. Figure 2-middle shows an illustrative example of the time series of this data set.

- **Data set UCI-2**: The last data set considered in this work is formed from a subset of the extensive data set presented in [Vergara et al., 2013]. Concretely, we consider here two subsets formed by setting the sensor's heater voltage to $5v$, the fan speed to $5500rpm$, down-sampled at $10Hz$, and selecting the fifth e-nose (from

[1] Data set from the MAchine Perception and Intelligent Robotics (MAPIR) group, University of Malaga, Spain. For a full description and download, please visit: http://mapir.isa.uma.es/mapirwebsite/index.php/robotic-olfaction/139-odor-classification

[2] Notice that due to the similarity between all the alcoholic beverages in the original data set, we consider all of them as one single class.

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

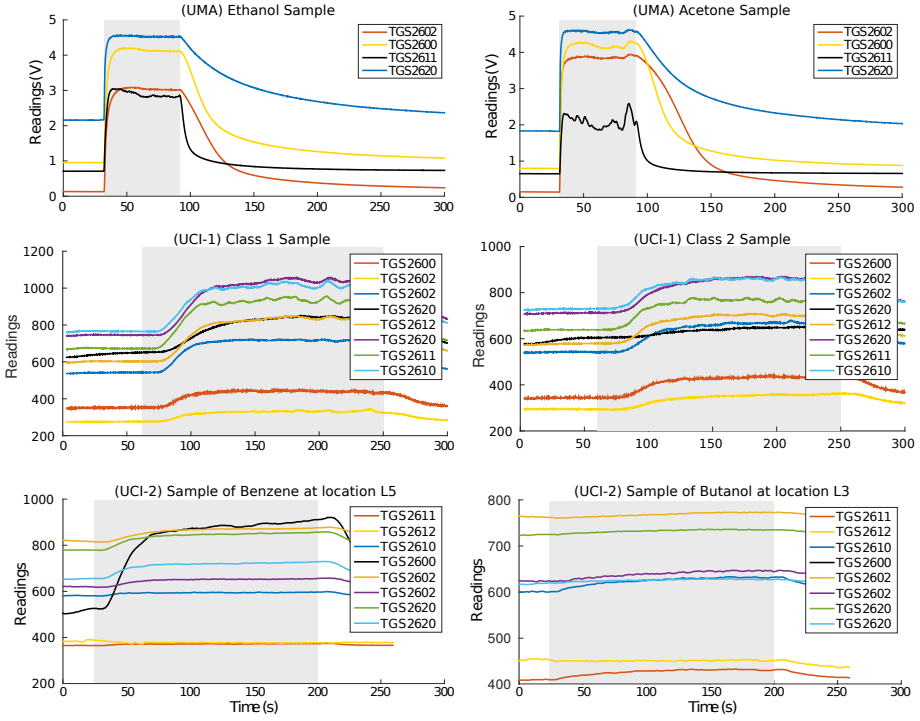

Figure 2: E-nose readings from two of the time series of each data set: (top) UMA, (middle) UCI-1 and (bottom) UCI-2. The volatile exposure times are marked as a shaded-grey region. As can be seen, different number and type of gas sensors and chemical volatiles are considered.

the nine available) located at positions $L3$ and $L5$. These settings lead to two data set (hereinafter UCI-2_L3 and UCI-2_L5) composed of 200 time series recorded in a wind tunnel with an e-nose composed of eight Figaro gas sensors. The main difference with respect the previous data set is the fact that here we account for ten different gas classes, with the consequently complexity increase. As in previous cases, to discard the initial baseline period as well as samples long after the gas exposure, we restrict the feature extraction to samples in the time interval [25s,255s]. Figure 2-bottom plots some samples of this data set where it is noticeable the absence of steady state values.

5.2. Results from Data Set UMA

Given the controlled environmental conditions employed to gather the samples of this data set, and in order to gain some understanding about its influence over the classification rate, we initially perform a PCA analysis by considering sliding windows of length one second. The resulting data is shown in Figure 3, where the first three principal components are plotted. The variances explained by each principal component are: 87.36%, 9.13% and 2.66%, respectively. It is interesting to notice the high class separability as the window slides through the step-like signal (see Figure 2-top), which indicates that classification is feasible.

Placing emphasis on the repeatability of the data samples, we can study which parts of the signal contain more discriminatory information attending to the considered features, and to the window length. This analysis is shown in Figure 4, where the average classification accuracy of a naive Bayes classifier is plotted for different lengths and positions of the

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

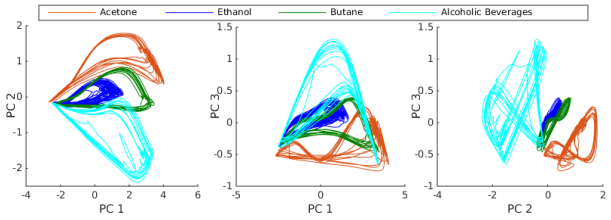

Figure 3: Classes of the data set UMA after performing a PCA analysis over sliding windows of length one second.

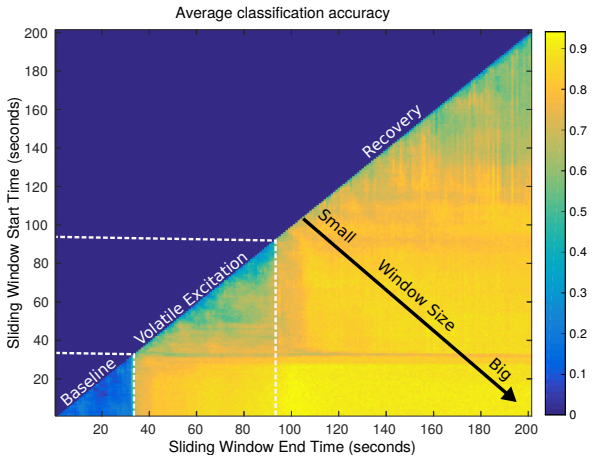

Figure 4: Average classification accuracy of a naive Bayes classifier for different lengths and positions of the sliding window within the UMA data set.

sliding window. Different conclusions can be drawn from this analysis:

- In general, small windows (close to the diagonal) yield lower accuracy than windows with longer lengths. This is due to the fact that no class transitions are considered in this data set, thus, the longer the window, the more information about the nature of the gas.

- The only two exceptions to the previous point are windows that, despite their short length, are placed in the transient phases of the signals (i.e. after the start and end of the volatile exposition (seconds 30 and 90 approximately)). This suggests that transient samples are more discriminant than steady state ones.

- Before and long after the e-nose has been exposed to the volatiles, the classification rate drops because the sensor signals are almost constant and carry no gas information.

- The maximum average classification is achieved when the window starts in between 0 and 30 seconds, and spans until 100 to 200 seconds, that is, when both transient phases (rise and recovery) are included. Nevertheless, in real applications, long windows have the drawback of increasing the chance of mixing data from different gas classes within the window span.

Focusing now on the overview analysis, Figure 5 presents the classification accuracies for the set of window lengths and classifiers considered in this work. These results are obtained from the double cross-validation procedure which initially partitions the data set into 80% train (corresponding to 55 experiments) and 20% test (14 experiments). Yet, it is important to notice, that, from one single experiment several samples are extracted by applying the sliding window approach.

From these results, we can observe that most classifiers are able to correctly classify the gas samples, even when only the most recent reading of the e-nose is used as feature (one-sample window length). Both, ELM and decision trees, show barely no accuracy variation when increasing the window length, while naive Bayes shows an accuracy improvement of 2% units, approximately. This fact suggests that, for this data set, using the one-sample classifier version is probably the best option, since the amplitude values of the e-nose's sensors are discriminant enough. As already mentioned, given the low variability of the data samples it is not surprising to obtain such high classification rates, which is also supported by the high class separability appreciated in the PCA analysis of the data (see Figure 3).

Hypothesis Test

From the overview analysis we have concluded that there is barely no improvement in using the sliding window approach for this UMA dataset. Yet, we perform here the hypothesis test to ascertain if our

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

Table 2: Comparison Between One-sample and Sliding Window with Optimal Window Length Approaches for the UMA data set - Average Classification Accuracies and p-value of the hypothesis tests for a significance level of 5%.

Classifier	One Sample		Sliding Window			t-test	Wilcoxon
	Accuracy mean(std)	Lilliefors h(p-value)	Opt.length mean(std)	Accuracy mean(std)	Lilliefors h(p-value)	p-value	p-value
NaiveBayes	0.68(0.03)	0(0.5000)	8.80(4.61)	0.72(0.03)	1(0.0010)	-	0.0003
Tree	0.97(0.01)	0(0.0686)	5.55(4.09)	0.98(0.01)	0(0.5000)	0.0005	-
ELM	0.86(0.01)	0(0.1973)	7.80(5.08)	0.87(0.01)	0(0.5000)	0.0001	-

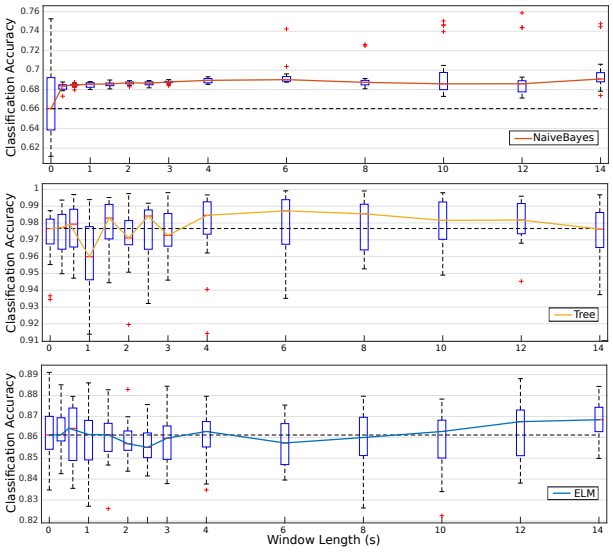

Figure 5: Boxplot of the classification accuracy for data set UMA. For comparison, an horizontal dashed line is plotted at the median accuracy value corresponding to the one-sample window length.

original conclusions are supported by an statistical analysis. Table 2 compares the average classification accuracy for the cases of using a one-sample classifier versus the case of employing a sliding window approach with the optimal window length. We also include the results of the Lilliefors test (to check for the Gaussianity of the data from both approaches), and the p-value for the t-test or Wilcoxon test, respectively.

As can be seen, the accuracy results confirm the conclusions obtained in the overall analysis, that is, there is not a significant improvement when using sliding windows (both decision trees and ELM shown an average improvement of 1% unit), except for the the case of naive Bayes, which shows an improvement of 4% units. Despite this, the null hypothesis is always rejected (with high statistical significance), which means that even if the improvement is small, sliding windows have a positive effect on the classification performance.

5.3. Results from Data Set UCI-1

This second data set presents a more challenging scenario since there is no control over the environment conditions while recording the e-nose samples, (see Section 5.1). This can be seen in the PCA plot shown in Figure 6, obtained by employing a window length of one second. As can be noticed, there is a big difference when compared to the PCA of data set UMA (see Figure 3). In this case, neither repeatability, nor high separability in the classes is appreciated, which corroborates the increased difficulty present in the classification of odor volatiles in uncontrolled environments. In this case scenario, the variances explained by each principal component are: 79.67%, 10.41% and 8.63%, respectively.

Under this challenging scenario, the classification accuracies of the three classifiers are shown in Figure 7. The data partitioning corresponding to the

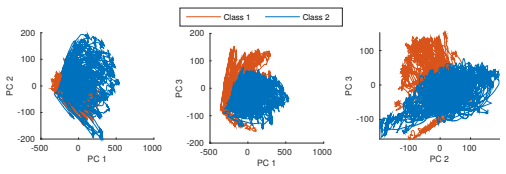

Figure 6: Classes of the data set UCI-1 after performing a PCA analysis over sliding windows of length one second.

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

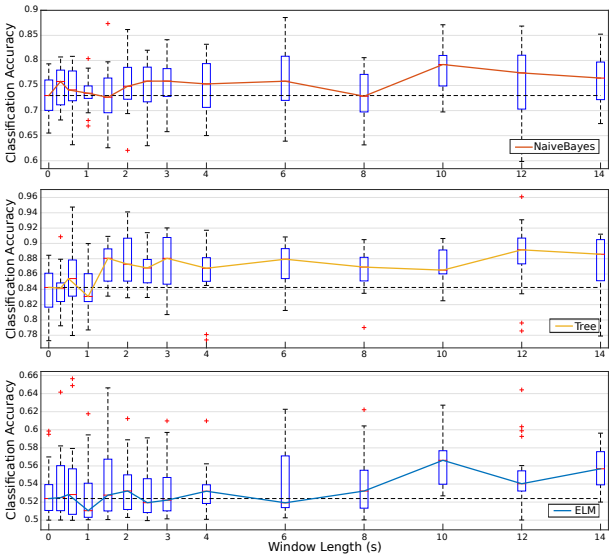

Figure 7: Boxplot of the classification accuracy for data set UCI-1. For comparison, an horizontal dashed line is plotted at the median accuracy value corresponding to the one-sample window length.

first level cross-validation leads to 144 experiments for training and 36 experiments for testing. Some conclusions can be drawn from these results:

- As expected, there is a noticeable decrease in the overall accuracy performance when compared to the case of controlled gas pulses (data set UMA). The reason behind this drop is the increased dynamics on the e-nose readings due to the presence of turbulent airflows during the recording process. The only exception is naive Bayes classifier, which presents a slightly increase on its performance.

- An improvement on the classification accuracy up to 6% units can be accomplished when considering delayed samples by means of sliding windows. Yet, the optimal window length depends on the classifier, which suggest that an optimization phase should be carried out to determine the best value (see Hypothesis Test subsection below). This finding is, obviously, only applicable to classifiers which do not internally consider the time dependencies of the data. That is, works like [Fonollosa et al., 2015; G. Monroy and Gonzalez-Jimenez, 2015] where high classification rates are achieved using only the most recent e-nose observation, already consider the time correlation of the data within the classification algorithm, and thus information from past samples are also used during the estimation of the data class.

- Small windows (0.3 to 2 seconds) present accuracies quite similar to the one-sample case. This is due to the few samples considered by those windows and the consequently difficulty to exploit the temporal correlation. Opposite to this, long windows present the higher accuracy improvements, yet, for real scenarios, they imply an increase in the chance that the data-generating process has changed over the time period covered by the window (e.g. class transition). The latter cannot be proven with real experiments because of the difficulty to obtain a precise ground truth when gas transitions are present. To the best knowledge of the authors, no publicly available data set considers gas transitions with a reliable ground truth.

Hypothesis Test

Here we carry out the second analysis of the data, employing in this case train, validation and test subsets as described in Algorithm 2. The objective is to discern if the use of sliding windows improves the classification performance (as suggested by the overview analysis) when compared with the case of using the most recent e-nose observation to classify. Table 3 summarizes the results of such analysis for the UCI-1 dataset. Two main conclusions can be extracted from these results. First, with respect to the classification accuracy, it can be seen how the optimal sliding window yields improvements in the range (1 - 4)% units, depending on the classifier employed. Second, related to the null hypothesis we see how it is rejected in all cases except for the Tree classifier. For this case, results from both configurations are generated from a probability distribution with similar mean value (therefore no significant improvement), and the differences in the results may be due to ran-

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

Table 3: Comparison Between One-sample and Sliding Window with Optimal Window Length Approaches for the UCI-1 data set - Average Classification Accuracies and p-value of the hypothesis tests for a significance level of 5%.

Classifier	One Sample		Sliding Window			t-test	Wilcoxon
	Accuracy mean(std)	Lilliefors h(p-value)	Opt.length mean(std)	Accuracy mean(std)	Lilliefors h(p-value)	p-value	p-value
NaiveBayes	0.73(0.04)	0(0.1013)	11.30(5.54)	0.74(0.05)	0(0.5000)	0.0460	-
Tree	0.80(0.04)	0(0.5000)	5.81(4.94)	0.82(0.07)	0(0.1155)	0.2839	-
ELM	0.53(0.02)	0(0.2025)	5.37(4.78)	0.57(0.06)	1(0.0021)	-	0.0006

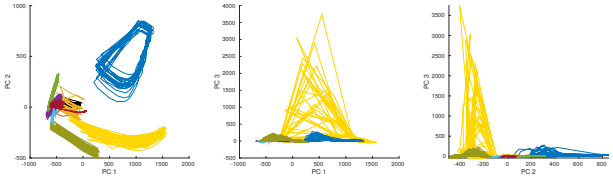

Figure 8: Classes of the data set UCI-2_L3 after performing a PCA analysis over sliding windows of length one second.

dom fluctuations.

5.4. Results from Data Set UCI-2

As with the previous data sets, we start by plotting the PCA of the data when considering sliding windows of length one second to get an insight about the data structure. Figure 8 plots the first three principal components for the UCI-2_L3 data set (we omit the corresponding PCA for UCI-2_L5 because of its similarity), where, as in the case of data set UCI-1, not clear separability among classes is noticeable (except for Class 1). The total variances explained by each principal component are 84.82%, 12.01%, and 2.25%, respectively. As mentioned in the data set description, to the turbulence transport phenomena we now add the complexity due to the high number of chemical classes to recognize.

The classification accuracy with respect the window length is shown in Figure 9 and Figure 10 for both subsets of the UCI-2 data set, L3 and L5 respectively. The following conclusions can be drawn from these results:

- A small, yet important, improvement in the classification accuracy (up to 5% units) is appreciated when employing sliding windows (even for small ones). This corroborates that it is possible to exploit the temporal correlation of real e-nose data by considering delayed samples. The only exception is ELM which is not suited for this complex data set, providing estimations below 0.3 accuracy.

- Although this data set introduces more classes in the classification problem than data set UCI-1, the classification accuracies are, on average, higher. One possible reason is the fact that data set UCI-1 considers mixtures of gases, which may mislead the classifiers. Despite this, both data sets show similar behaviours with respect to the window length, which is the focus of this research.

Hypothesis Test

As with the previous data sets, we perform now the hypothesis test to check if the null hypothesis that the two competing approaches perform equally well is supported by the results corresponding to this data set. Tables 4 and 5 depict the results obtained when comparing the one-sample and the sliding window approaches, and provides the p-values for the t-test or the Wilcoxon test, respectively. Again, small yet important improvements can be noticed when using the optimal window length (between 1% and 6% units). Furthermore, these improvements seem to be significant, since the p-values are in most cases much lower than the 0.05 significance level, which means that the null hypothesis is rejected. The only exception is the combination of dataset UCI-2_L3 with decision tree classifier, which not only provides a small improvement of 1% units, but also leads to the rejection of the null hypothesis.

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

Table 4: Comparison between One-sample and Sliding Window with Optimal Window Length Approaches for the UCI-2_L3 data set - Average Classification Accuracies and p-value of the hypothesis tests for a significance level of 5%.

Classifier	One Sample		Sliding Window			t-test	Wilcoxon
	Accuracy mean(std)	Lilliefors h(p-value)	Opt.length mean(std)	Accuracy mean(std)	Lilliefors h(p-value)	p-value	p-value
NaiveBayes	0.89(0.03)	0(0.1273)	9.55(5.82)	0.90(0.03)	0(0.4482)	0.0469	-
Tree	0.93(0.03)	0(0.5000)	9.25(4.70)	0.94(0.04)	1(0.0018)	-	0.2184
ELM	0.30(0.05)	0(0.3924)	4.92(4.72)	0.35(0.03)	0(0.2416)	0.0005	-

Table 5: Comparison between One-sample and Sliding Window with Optimal Window Length Approaches for the UCI-2_L5 data set - Average Classification Accuracies and p-value of the hypothesis tests for a significance level of 5%.

Classifier	One Sample		Sliding Window			t-test	Wilcoxon
	Accuracy mean(std)	Lilliefors h(p-value)	Opt.length mean(std)	Accuracy mean(std)	Lilliefors h(p-value)	p-value	p-value
NaiveBayes	0.90(0.03)	1(0.0496)	1.18(0.41)	0.93(0.03)	1(0.0311)	-	0.0040
Tree	0.90(0.04)	0(0.5000)	8.30(5.65)	0.93(0.03)	1(0.0051)	-	0.0031
ELM	0.29(0.04)	0(0.5000)	2.29(3.45)	0.35(0.06)	1(0.0440)	-	0.0004

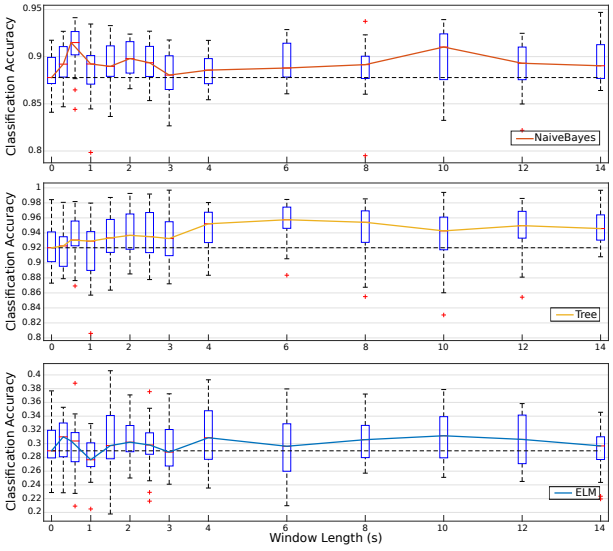

Figure 9: Boxplot of the classification accuracy for data set UCI-2_L3. For comparison, an horizontal dashed line is plotted at the median accuracy value corresponding to the one-sample window length.

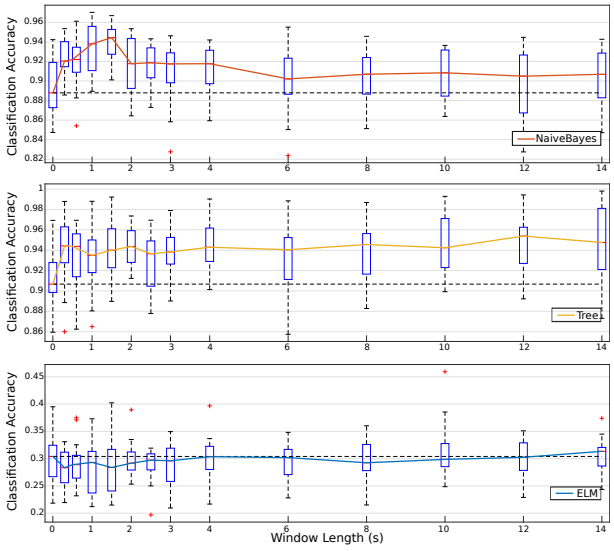

Figure 10: Boxplot of the classification accuracy for data set UCI-2_L5. For comparison, an horizontal dashed line is plotted at the median accuracy value corresponding to the one-sample window length.

6. Conclusions and Future Work

In this work we have reviewed the problem of continuous chemical classification with an electronic nose under uncontrolled environmental conditions, analyzing to which extent feature extraction based on fixed length sliding windows is a feasible option to

Authors' accepted manuscript: Chemometrics and Intelligent Laboratory Systems, 2016.
The final publication is available at: http://dx.doi.org/10.1016/j.chemolab.2016.08.011

exploit the temporal correlation of the data. Results obtained by employing three publicly available olfaction data sets covering different e-nose configurations, gas classes and environmental conditions, as well as three well-known classifiers, demonstrate that a sliding window methodology improves the classification accuracy with respect to the static configuration, which only accounts for the instantaneous values of the e-nose's sensors. More specifically, this work has led to the following findings:

- Accounting for delayed samples of the e-nose by means of sliding windows leads to improvements in the classification accuracy up to 6% units with respect the case of using only the most recent sample, depending on the data set and the classifier at hand. The statistical relevance of these results is supported by the use of a double cross-validation process (experiment and sample levels), and the results of an hypothesis test which, in general terms, corroborates the results with high significance.

- In 10 out of the 12 combinations of (data set - classifier) considered in this work, the hypothesis test rejects with high significance the null hypothesis that results from both approaches (one-sample and sliding windows) are similar. This means that even if the improvement derived of employing sliding windows is small, they have a positive effect on the classification performance.

- Related to the optimal window length, results presented in Tables 2, 3, 4 and 5 show that the optimal window length considerably varies with the data and classifier used. In general, long windows are preferred over short ones (the average among all experiments is 6.68 seconds), but an optimization phase is required to determine the optimal value.

- We can conclude that it is advisable to exploit the temporal correlation inherent on e-nose data, by means of the simple sliding window approach used in this work, or by more elaborate mechanisms, since, although small, improvements can be achieved with little computational cost.

Future work will consider new features to analyze to which extent they affect the optimal window length, as well as dealing with the problem of real time classification of chemical volatiles when performed in motion.

Acknowledgments

This work is partially supported by the Ministry of Economy and Competitiveness of Spain under grants DPI2014-55826-R and TIN2014-53465-R. It is also partially supported by the Autonomous Government of Andalusia (Spain) and the European Regional Development Fund under projects TEP530, TIC6213 and TIC-657. All of them include funds from the European Regional Development Fund (ERDF). The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga.

References

Breiman, L., Friedman, J., Stone, C. J., Olshen, R., 1984. Classification and Regression Trees. Wadsworth Statistics/Probability. Chapman and Hall/CRC.

Campagnoli, A., Dell'Orto, V., 2013. Potential application of electronic olfaction systems in feedstuffs analysis and animal nutrition. Sensors (Switzerland) 13 (11), 14611–14632.

Carmel, L., Levy, S., Lancet, D., Harel, D., 2003. A feature extraction method for chemical sensors in electronic noses. Sensors and Actuators B: Chemical 93 (1-3), 67–76.

Coppersmith, D., Hong, S. J., Hosking, J. R., 1999. Partitioning nominal attributes in decision trees. Data Mining and Knowledge Discovery.

De Vito, S., Delli Veneri, P., Esposito, E., Salvato, M., Bright, V., Jones, R., Popoola, O., 2015. Dynamic multivariate regression for on-field calibration of high speed air quality chemical multi-sensor systems.

Dentoni, L., Capelli, L., Sironi, S., Rosso, R. D., Zanetti, S., Torre, M. D., 2012. Development of an electronic nose for environmental odour monitoring. Sensors 12 (11), 14363–14381.

Fonollosa, J., Rodriguez-Lujan, I., Trincavelli, M., Vergara, A., Huerta, R., 2014. Chemical discrimination in turbulent gas mixtures with MOX sensors validated by gas chromatography-mass spectrometry. Sensors (Switzerland) 14 (10), 19336–19353.

Fonollosa, J., Sheik, S., Huerta, R., Marco, S., 2015. Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. Sensors and Actuators B: Chemical 215, 618–629.

Friedman, N., Geiger, D., Goldszmidt, M., Nov. 1997. Bayesian network classifiers. Mach. Learn. 29 (2-3), 131–163.

Fu, T. c., 2011. A review on time series data mining. Engineering Applications of Artificial Intelligence 24 (1), 164–181.

G. Monroy, J., Blanco, J.-L., Gonzalez-Jimenez, J., 2015. Time-variant gas distribution mapping with obstacle information. Autonomous Robots, 1–16.

G. Monroy, J., Gonzalez-Jimenez, J., 2015. Real-time odor classification through sequential Bayesian filtering. In: 16th International Symposium on Olfaction and Electronic Nose (ISOEN).

G. Monroy, J., Lilienthal, A. J., Blanco, J.-L., Gonzalez-Jimenez, J., Trincavelli, M., 2013. Probabilistic gas quantification with MOX sensors in open sampling systems-a gaussian process approach. Sensors and Actuators, B: Chemical 188, 298–312.

Gelperin, A., Johnson, A., 2008. Nanotube-based sensor arrays for clinical breath analysis. Journal of Breath Research 2 (3).

Guo, D., Zhang, D., Li, N., Zhang, L., Yang, J., 2010. A novel breath analysis system based on electronic olfaction. IEEE Transactions on Biomedical Engineering 57 (11), 2753–2763.

Hatami, N., Chira, C., April 2013. Classifiers with a reject option for early time-series classification. In: IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL). pp. 9–16.

He, Z., Zhou, H., Wang, J., Chen, Z., Wang, D., Xing, Y., 2015. An improved detection statistic for monitoring the nonstationary and nonlinear processes. Chemometrics and Intelligent Laboratory Systems 145, 114 – 124.

Hu, B., Chen, Y., Keogh, E., 2013. Time series classification under more realistic assumptions. In: Proceedings of the SIAM Data Mining Conference (SDM). pp. 578–586.

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70 (1-3), 489–501.

Llobet, E., Brezmes, J., Ionescu, R., Vilanova, X., Al-Khalifa, S., Gardner, J., Bârsan, N., Correig, X., 2002. Wavelet transform and fuzzy artmap-based pattern recognition for fast gas identification using a micro-hotplate gas sensor. Sensors and Actuators B: Chemical 83 (1-3), 238–244.

López-Rubio, E., Ortiz-de Lazcano-Lobato, J. M., 2009. Automatic model selection by cross-validation for probabilistic PCA. Neural Processing Letters 30, 113–132.

Loutfi, A., Broxvall, M., Coradeschi, S., Karlsson, L., 2005. Object recognition: A new application for smelling robots. Robotics and Autonomous Systems 52 (4), 272–289.

Molinaro, A. M., Simon, R., Pfeiffer, R. M., 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics 21 (15), 3301–3307.

Niennattrakul, V., Srisai, D., Ratanamahatana, C. A., 2012. Shape-based template matching for time series data. Knowledge-Based Systems 26 (0), 1 – 8.

Perera, A., Papamichail, N., Barsan, N., Weimar, U., Marco, S., 2006. On-line novelty detection by recursive dynamic principal component analysis and

gas sensor arrays under drift conditions. Sensors Journal, IEEE 6 (3), 770–783.

Rokach, L., Maimon, O., 2008. Data Mining with Decision Trees: Theroy and Applications. World Scientific Publishing Co., Inc., River Edge, NJ, USA.

Schaller, E., Bosset, J. O., Escher, F., 1998. Electronic noses and their application to food. LWT - Food Science and Technology 31 (4), 305–316.

Schleif, F.-M., Hammer, B., G. Monroy, J., Gonzalez-Jimenez, J., Blanco, J.-L., Biehl, M., Petkov, N., 2015. Odor recognition in robotics applications by discriminative time-series modeling. Pattern Analysis and Applications, 1–14.

Simoes, P., Izumi, N., Casagrande, R., Venson, R., Veronezi, C., Moretti, G., da Rocha, E., Cechinel, C., Ceretta, L., Comunello, E., Martins, P., Casagrande, R., Snoeyer, M., Manenti, S., 2014. Classification of images acquired with colposcopy using artificial neural networks. Cancer Informatics 13, 119–124.

Szczurek, A., Krawczyk, B., Maciejewska, M., 2013. VOCs classification based on the committee of classifiers coupled with single sensor signals. Chemometrics and Intelligent Laboratory Systems 125 (0), 1–10.

Trincavelli, M., Coradeschi, S., Loutfi, A., 2009. Odour classification system for continuous monitoring applications. Sensors and Actuators B: Chemical 139 (2), 265–273.

Vergara, A., Fonollosa, J., Mahiques, J., Trincavelli, M., Rulkov, N., Huerta, R., 2013. On the performance of gas sensor arrays in open sampling systems using inhibitory support vector machines. Sensors and Actuators B: Chemical 185, 462–477.

Vergara, A., Vembu, S., Ayhan, T., Ryan, M. A., Homer, M. L., Huerta, R., 2012. Chemical gas sensor drift compensation using classifier ensembles. Sensors and Actuators B: Chemical 166-167 (0), 320–329.

Vito, S. D., Piga, M., Martinotto, L., Francia, G. D., 2009. Co, NO2 and NOx urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. Sensors and Actuators B: Chemical 143 (1), 182–191.

Yan, J., Guo, X., Duan, S., Jia, P., Wang, L., Peng, C., Zhang, S., 2015. Electronic Nose Feature Extraction Methods: A Review. Sensors 15 (11), 27804–27831.

Zhao, Z., Xu, S., Kang, B., Kabir, M., Liu, Y., Wasinger, R., 2015. Investigation and improvement of multi-layer perception neural networks for credit scoring. Expert Systems with Applications 42 (7), 3508–3516.