Sensors and Actuators B 166–167 (2012) 320–329

Contents lists available at SciVerse ScienceDirect

Sensors and Actuators B: Chemical

journal homepage: www.elsevier.com/locate/snb

Chemical gas sensor drift compensation using classifier ensembles

Alexander Vergara[a,*,1], Shankar Vembu[a,1], Tuba Ayhan[b,2], Margaret A. Ryan[c], Margie L. Homer[c], Ramón Huerta[a]

[a] BioCircuits Institute, University of California, San Diego, La Jolla, CA 92093, USA
[b] Department of Electronics and Communication Engineering, Technical University of Istanbul, Maslak, TR-34469 Istanbul, Turkey
[c] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

ARTICLE INFO

Article history:
Received 27 September 2011
Received in revised form
26 November 2011
Accepted 17 January 2012
Available online 1 March 2012

Keywords:
Sensor drift
Metal-oxide sensors
Time series classification
Ensemble methods
Support vector machines

ABSTRACT

Sensor drift remains to be the most challenging problem in chemical sensing. To address this problem we have collected an extensive dataset for six different volatile organic compounds over a period of three years under tightly controlled operating conditions using an array of 16 metal-oxide gas sensors. The recordings were made using the same sensor array and a robust gas delivery system. To the best of our knowledge, this is one of the most comprehensive datasets available for the design and development of drift compensation methods, which is freely reachable on-line. We introduced a machine learning approach, namely an ensemble of classifiers, to solve a gas discrimination problem over extended periods of time with high accuracy rates. Experiments clearly indicate the presence of drift in the sensors during the period of three years and that it degrades the performance of the classifiers. Our proposed ensemble method based on support vector machines uses a weighted combination of classifiers trained at different points of time. As our experimental results illustrate, the ensemble of classifiers is able to cope well with sensor drift and performs better than the baseline competing methods.

1. Introduction

In the history of sensors' development, the electronic noses, or simply *e-noses*, are a relatively new addition to the world of sensors, and can be defined as a collection of broadly cross-reactive sensors connected to electronics and an effective pattern recognition system used to detect, classify, and, where necessary, quantify a variety of chemical analytes or odors of concern in a certain area [1,2]. Ideally, these systems would greatly benefit from chemical sensors that would always show an identical response when exposed to the same analyte or a chemical mixture and return to their baseline level immediately after the gas being evaluated is no longer present at the sensor surface. However, in real-life applications, where sensors are operated over a long period of time, such an ideal situation is still largely unrealizable today [3]. The gradual and unpredictable variation of the chemo-sensory signal responses when exposed to the same analyte under identical conditions, a.k.a. sensor drift, has long been recognized as one of the most serious impairments faced by chemical sensors [4–7].

Drift has plagued the sensor research community for many years, deteriorating the performance of classifiers used for gas recognition and augmenting the maintenance costs of chemo-sensory systems, or artificial electronic noses, during real-time operations. In general, sensor drift can be attributed to two predominant sources [8,9]. First, the 'real-drift' (a.k.a. first-order drift) due to the chemical and physical interaction processes of the chemical analytes, in gas phase, occurring at the sensing film microstructure, such as aging (e.g. the reorganization of the sensor surface over long periods of time) and poisoning (e.g., irreversible binding due to external contamination). And second, the 'second-order drift' (or measurement system drift, among many other names), produced by the external and uncontrollable alterations of the experimental operating system, including, but not limited to, changes in the environment (e.g., temperature and humidity variations); measurement delivery system noise (e.g., tubes condensation, sample conditioning, etc.); and thermal and memory effects (e.g., hysteresis or remnants of previous gases). In general, a number of approaches under the notion of sensor drift counteraction have been implemented in the literature, but one of the pioneering works, and perhaps the most systematic sensor drift analysis was performed by Romain and co-workers [10,11], who utilized a very comprehensive dataset, collected over long periods of time in real operating conditions, to provide a deep insight into the sensor drift problem, for both the real and second-order drift. Among the many interesting conclusions drawn from that work,

* Corresponding author. Tel.: +1 8585346758; fax: +1 8585347664.
E-mail address: vergara@ucsd.edu (A. Vergara).
[1] Joint first authors.
[2] Tuba Ayhan carried out most of her work on this article while visiting the BioCircuits Institute at UCSD.

0925-4005/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.snb.2012.01.074

the following three aspects were emphasized: (i) from all the sensing technologies available, metal oxide based gas sensors [12] remain the best option for long term applications for continuous monitoring systems; (ii) a calibration gas is recommended to estimate sensor drift compensation, and (iii), sensor replacement is unavoidable over long periods of time.

In practical applications, it is difficult to empirically differentiate between real drift and second-order drift, if possible at all. Accordingly, it is hard to develop methods to correct different sources of drift because the origin of it cannot be ascertained. Utilizing an effective delivery system allows the chemical sensors to bypass the second-order drift effect, making it possible to exclusively concentrate on the chemical sensors for compensating real drift. In our particular gas delivery system, we can control the second-order drift, too, so we can exclusively address the real drift problem. Thus, in the remainder of this document we use the term *drift* to refer to *real drift*. Concerning real drift reduction, many efforts have been devoted to find the sensor materials that can reversibly interact with the gas so that the detected molecules unbind the sensor material as soon as the gas has been purged out from the sensor surface [13–15]. Other solutions based on periodically changing the sensor working temperature [16,17] have also been implemented in an effort to minimize the effects of irreversibility in the sensors' responses due to poisoning. Undoubtedly, heightened reversibility in the sensor response is necessary for the effective drift counteraction. However, this general treatment only constitutes one facet of the problem—the so-called short-term drift—a substantial study of the sensor variability over longer periods of time is also necessary.

The most commonly used solutions to cope with sensor drift within the chemical sensing community are univariate and multivariate methods, where drift compensation is performed either on each sensor individually or on the entire sensor array [9,18]. Among the multivariate drift compensation methods, unsupervised component correction techniques are the most popular [4,19,20]. These techniques rely on finding linear transformations that normalize the sensor responses across time so that a classifier can be directly applied to the resulting stationary data. For instance, the component correction method presented by Årtusson et al. [4] applies the following transformation to the measurement/data matrix $X \leftarrow X - (X \cdot c)c^{\top}$, where c is the principal component vector(s) of the measurements computed using a reference gas that may approximate the drift direction. The main drawback of these techniques is that they assume the drift direction to be linear in the feature space and, therefore, a linear transformation of the data suffices to correct it. While it is entirely plausible that kernelized versions of component analysis, such as kernel principal component analysis [21], can be applied to account for non-linearities in the drift direction, these techniques have not been investigated much in the chemical sensing community. Also, with the exception of Ref. [20], these techniques require a reference gas that is used to approximate the drift direction by assuming that the reference gas provides the drift direction in all the other gases.

In this paper, we take a completely different approach to solve the mentioned problem, in which we do not make any of the above-mentioned assumptions. Instead, we use a supervised machine learning method, namely, an ensemble of classifiers to cope with sensor drift. To the best of our knowledge, such a machine learning approach, that automatically detects and copes with sensor drift, has not been applied in the chemical sensing community before, although it has been shown to yield promising results on problems with drifting concepts in machine learning and data mining [22–25]. Utilizing a comprehensive dataset of a multi-component gas classification problem recorded from metal-oxide gas sensors over a course of 36 months, we investigate the feasibility of our ensemble of classifiers methodology to mitigate the drift effect in chemical gas sensors. It is important to note that the ensemble method used in this paper *complements*, rather than *competes* against, the existent component correction methods mentioned above; since component correction methods are essentially a pre-processing technique, we can indeed use the ensemble method on the pre-processed data, too. In the remainder of this paper, we first describe the experimental setup, the dataset, and the feature extraction methods considered in this work (Section 2). We then describe the drift compensation algorithm (Section 3), followed by a detailed description of our experimental findings (Section 4). And finally, we present the concluding comments drawn from the results presented in this paper (Section 5).

2. Data collection

We apply our drift compensation method to an extensive dataset[3] recorded by a metal-oxide gas sensor array. In this section, we describe the experimental setup, the recording protocol, and the signal processing algorithms used for feature extraction.

2.1. The experimental setup

We used a sixteen screen-printed commercially available metal-oxide semiconductor gas sensors array, manufactured and commercialized by Figaro Inc. [12], for our experiments. The custom design used in the sensing technology includes an independently controlled RuO_2 (Ruthenium Oxide) electrical heating line and a metal oxide semiconductor film as a sensor material printed onto the measuring electrodes (noble metal). The obtained sensor element is mounted onto an alumina substrate and then connected by lead wires to the pins of the sensor package. The resulting array, populated by sensor devices (4 of each) tagged by the manufacturer as TGS2600, TGS2602, TGS2610, TGS2620 is placed into a 60 ml-volume test chamber, where the odorants of interest, in gaseous form, are to be injected for trials. To generate the required dataset, we connect the said test chamber to a computer-controlled continuous flow system, which provides versatility for conveying the chemical compounds of interest at the desired concentrations to the sensing chamber with high accuracy and in a highly reproducible way while keeping the total flow constant. In particular, our system utilizes three digital, computed-supervised mass flow controllers (MFCs) (provided by Bronkhorst High-Tech B.V. [26]), each of them with different maximum flow levels (200, 100, and 20 ml/min, ±1% of accuracy). Such devices connect to different pressurized gas cylinders, which contain, diluted in dry air, either the carrier gas or the chemical analytes to be measured. To maintain the moisture level constant at 10% R.H. (measured at $25 \pm 1\,^{\circ}C$) during the entire measurement process, we utilize synthetic dry air as background for all measurements, provided by Airgas Inc. [27]. Then, the analytes under analysis (i.e., ammonia, acetaldehyde, acetone, ethylene, ethanol, and toluene) are added to this background in random order. The total flow rate across the sensing chamber is set to 200 ml/min and kept constant for the whole measurement process. The response of the gas sensor array was measured when the operating temperature of sensors was fixed at 400 °C, which, according to the deterministic one-to-one look-up table provided by the manufacturer [12], is attained via a built-in heater that is driven by an external DC voltage source set at 5 V. Finally, to ensure that reproducible response patterns are acquired during each measurement, the sensors were pre-heated for several days prior the experiment process gets started.

[3] The dataset will be made available on the UCI repository upon acceptance of the paper.

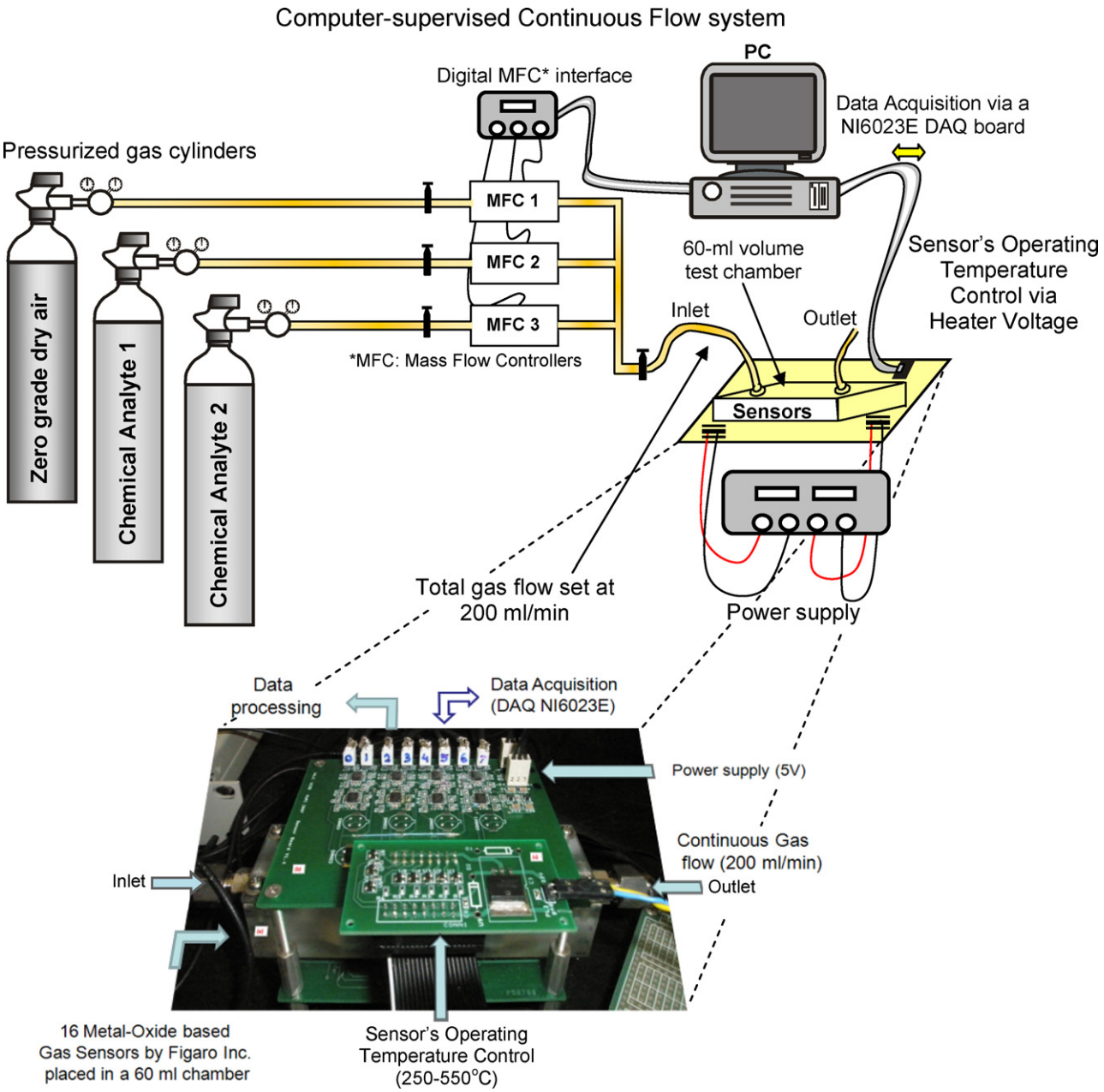

Fig. 1. Experimental setup used for data acquisition. The sensor responses are recorded in the presence of the analyte in gaseous form diluted at different concentrations in dry air. The measurement system operates in a fully computerized environment with minimal human intervention, which provides versatility in conveying the odors of interest (at the desired concentrations) to the sensing chamber with high accuracy, and simultaneously in keeping constant the total flow. Therefore, no changes in the flow or flow dynamics are reflected in the sensor response (i.e., only the presence of an odorant will be reflected in the sensor response). Moreover, since the system continuously supplies gas to the sensing chamber (either clean dry air or a chemical component), the amount of gas molecules in the sensing chamber is homogeneously distributed.

The sensor response is read-out in the form of the resistance across the active layer of each sensor; hence each measurement produces a 16-channel time series sequence. The data acquisition board collects the data from the gas sensors and controls the analog voltage signal to every sensor heater. This voltage is used to control and vary the sensor heater's operating temperature utilizing a made-in-home, LabVIEW [28] environment program running on a PC platform. The experimental setup is illustrated in the diagram shown in Fig. 1.

2.2. The dataset

Our dataset consists of a six-gas/analyte classification problem dosed at different concentrations, in which the goal is to discriminate the six different analytes regardless of their concentration. The odor identity and concentration values in parts-per-million by volume (ppmv) are listed in Table 1.

Each of the possible gas type-concentration pairs was sampled in no particular order. The resulting dataset consists of 13,910 recordings (time series sequences) collected over a period of 36 months. The exact distribution of the number of measurements per month is shown in Table 2. Notice that some of the months shown in the table do not contain any of the measurements for one or more gases; other gas analytes and/or complex mixtures that are not considered in this study were collected during this specific period of time utilizing the same array of sensors and experimental apparatus, which in turn increased the complexity of the problem due to interference from these external analytes that may

Table 1
Analytes and concentrations in the dataset.

Analytes	Concentrations in ppmv
Ammonia	50, 60, 70, 75, 80, 90, 100, 110, 120, 125, 130, 140, 150, 160, 170, 175, 180, 190, 200, 210, 220, 225, 230, 240, 250, 260, 270, 275, 280, 290, 300, 350, 400, 450, 500, 600, 700, 750, 800, 900, 950, 1000
Acetaldehyde	5, 10, 13, 20, 25, 30, 35, 40, 45, 50, 60, 70, 75, 80, 90, 100, 120, 125, 130, 140, 150, 160, 170, 175, 180, 190, 200, 210, 220, 225, 230, 240, 250, 275, 300, 500
Acetone	12, 25, 38, 50, 60, 62, 70, 75, 80, 88, 90, 100, 110, 120, 125, 130, 140, 150, 170, 175, 180, 190, 200, 210, 220, 225, 230, 240, 250, 260, 270, 275, 280, 290, 300, 350, 400, 450, 500, 1000
Ethylene	10, 20, 25, 30, 35, 40, 50, 60, 70, 75, 90, 100, 110, 120, 125, 130, 140, 150, 160, 170, 175, 180, 190, 200, 210, 220, 225, 230, 240, 250, 275, 300
Ethanol	10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 100, 110, 120, 125, 130, 140, 150, 160, 170, 175, 180, 190, 200, 210, 220, 225, 230, 240, 250, 275, 500, 600
Toluene	10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100

Table 2
Dataset details. Each row corresponds to samples collected during a period of one month for six gases.

Month ID	Number of examples						
	Ammonia	Acetaldehyde	Acetone	Ethylene	Ethanol	Toluene	Total
month1	76	0	0	88	84	0	248
month2	7	30	70	10	6	74	197
month3	0	0	7	140	70	0	217
month4	0	4	0	170	82	5	261
month8	0	0	0	20	0	0	20
month9	0	0	0	4	11	0	15
month10	100	105	525	0	1	0	731
month11	0	0	0	146	360	0	506
month12	0	192	0	334	0	0	526
month13	216	48	275	10	5	0	554
month14	0	18	0	43	52	0	113
month15	12	12	12	0	12	0	48
month16	20	46	63	40	28	0	197
month17	0	0	0	20	0	0	20
month18	0	0	0	3	0	0	3
month19	110	29	140	100	264	9	652
month20	0	0	466	451	250	458	1625
month21	360	744	630	662	649	568	3613
month22	25	15	123	0	0	0	163
month23	15	18	20	30	30	18	131
month24	0	25	28	0	0	1	54
month30	100	50	50	55	61	100	416
month36	600	600	600	600	600	600	3600

potentially affect the sensor life. Additionally, as observed in Table 2, the last batch of recordings, which contains 3600 measurements from the same analytes, was purposely collected five months after the sensors had been powered off. This 5-month gap is extremely significant for this study not only because it allows us to validate our suggested method on the annotated set of measurements collected five months later, but also because it is during this period of time that the sensors were prompted to severe contamination since external interferents could easily and irreversibly get attached to the sensing layer due to the lack of the operating temperature.

To generate the dataset, we followed a measurement procedure consisting of the following steps. First, a constant flow of zero-grade dry air was circulated through the sensing chamber while the gas sensor array was kept at a stable operating temperature (400 °C).[4] This step was done to measure the baseline steady-state sensor response (i.e., the sensor response in the presence of no chemical analytes). Afterwards, the desired concentration of the odorant was injected by the continuous flow system into the sensing chamber. Finally, in the third step (cleaning phase) the vapor was vacuumed away from the sensor array and the test chamber was cleaned with dry air before the concentration phase of a new measurement. The acquisition time of these measurements took at least 300 s to complete, divided into 100 s for the gas injection phase and at least 200 s for the recovery (cleaning) phase. For processing purposes, we considered the whole sensor response after subtracting the baseline from each record. The sampling rate was set to 100 Hz. Finally, the measurement process herein described was replicated for subsequent measurements.

2.3. Data processing and feature extraction

The Figaro metal-oxide gas sensors are known to have a slow response to a chemical analyte. This response, under tightly controlled operating conditions (i.e., constant air flow and fixed operating temperature), typically involves a monotonically saturating smooth change in the conductance/resistance across its sensing layer due to the adsorption/desorption reactions of the chemical analyte occurring at the micro-porous surface of the sensor. The amount and speed of these reactions depend on (i) the analyte identity, (ii) the analyte concentration, (iii) the active layer (i.e., sensor type), and (iv) the surface temperature (i.e., the sensors' operating temperature). Since the last two factors are fixed throughout the entire measurement procedure of this analysis, the sensor-analyte identity/concentration interaction process becomes the only factor that, as a pair, shapes the response profile, and, thus, that defines the identity of the chemical analyte of interest [29]. Accordingly, features reflecting the whole sensing dynamics at the sensor surface are of special interest in our drift compensation analysis.

Feature extraction plays an important role in every chemosensory application [30]; it is defined as a transformation mapping

[4] We do not have access to the actual sensing surface temperature due to packaging, but a look-up table relating it to the heater voltage can be found upon request in Ref. [12].

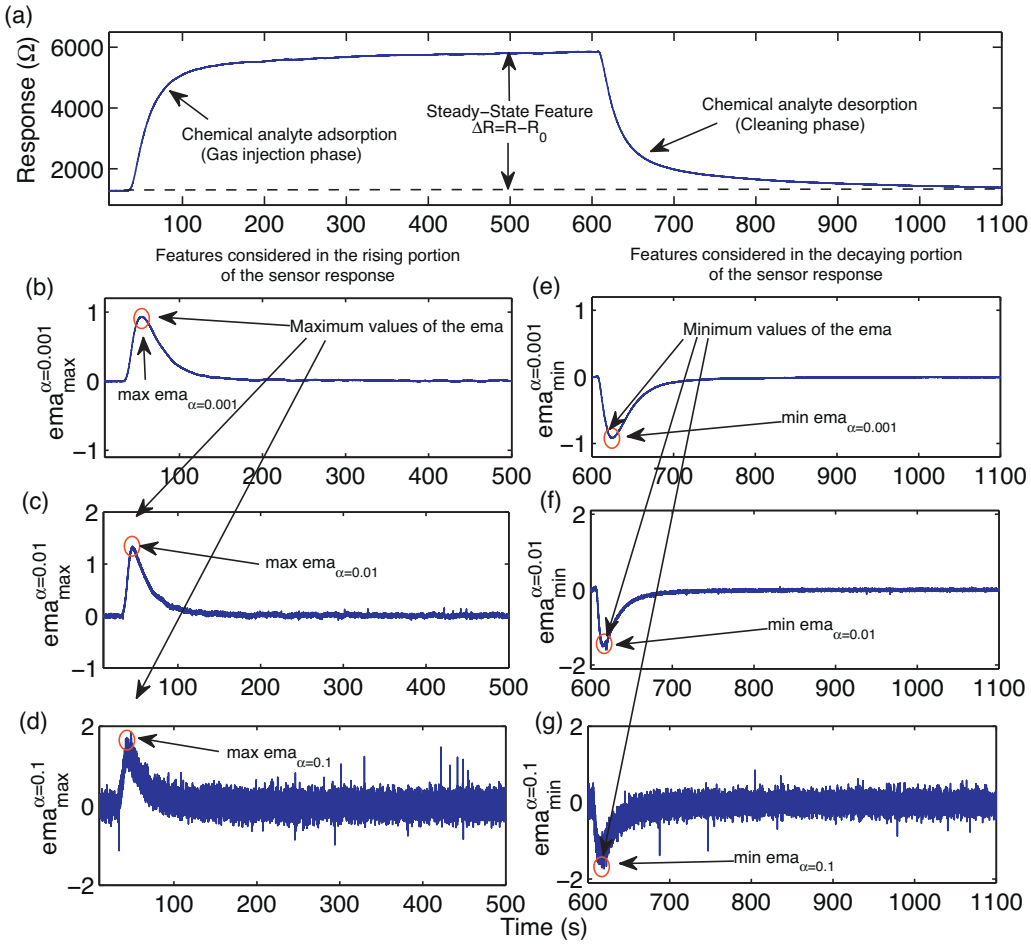

Fig. 2. Panel (a), typical response of a metal-oxide based chemical sensor to 30 ppmv of Acetaldehyde. The curve shows the three phases of a measurement: baseline measurement (made with pure air), test gas measurement (when the chemical analyte is injected, in gas form, to the test chamber), and recovery phase (during which the sensor again is exposed to pure air; the recovery time is usually much longer that the gas injection phase). Panels (b)–(d), exponential moving average of the rising portion of the sensor response (gas injection) for $\alpha = 0.1$, $\alpha = 0.01$, and $\alpha = 0.001$. The maximum values of the graphs (i.e., $\max_k ema_\alpha r[k]$), represent the features extracted from the chemical sensor in response to the analyte. Panels (e)–(g), exponential moving average of the decaying portion of the sensor response (cleaning phase) for $\alpha = 0.1$, $\alpha = 0.01$, and $\alpha = 0.001$. The minimum values of the graphs (i.e., $\min_k ema_\alpha r[k]$), are the features we extracted to represent the sensor signal during its cleaning phase. Features are marked with red circles in the plot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the sensor response to a space of lower dimension preserving the most meaningful portion of information contained in the original sensor signal. In this work, we consider two distinct types of features that exploit the whole dynamic processes occurring at the sensor surface, including the ones that reflect its adsorption, desorption, and steady-state (or final) responses of the sensor element. On the one hand, we utilize the steady-state feature, which is the "gold-standard" for chemo-sensory feature extraction [31]. It is defined as the difference of the maximal resistance change and the baseline,

$$\Delta R = \max_k r[k] - \min_k r[k], \tag{1}$$

and its normalized version expressed by the ratio of the maximal resistance and the baseline values,

$$||\Delta R|| = \frac{\max_k r[k] - \min_k r[k]}{\min_k r[k]}, \tag{2}$$

where $r[k]$ is the time profile of sensor resistance, k is the discrete time indexing the recording interval $[0, T]$ when the chemical vapor is present in the test chamber.

On the other hand, we also extract an aggregate of features reflecting the sensor dynamics of the increasing/decaying transient portion of the sensor response during the entire measurement procedure under controlled conditions. In particular, we utilize the exponential moving average (ema_α), a transform borrowed from the field of econometrics originally introduced to the chemical sensor community in [32] that converts the increasing/decaying and saturating discrete time series $r[\cdot]$ collected from the chemical sensor into a real scalar $f_\alpha\{r[\cdot]\}$, by estimating the maximum value (or minimum for the decaying portion of the sensor response) of its exponential moving average transform (ema_α), calculated by,

$$y[k] = (1-\alpha)y[k-1] + \alpha(r[k] - r[k-1]), \tag{3}$$

where $k = 1, 2, \ldots, T$, $y[0]$, its initial condition, set to zero ($y[0] = 0$), and the scalar α ($\alpha \in \{0, 1\}$) being a smoothing parameter of the operator that defines both the quality of the feature $f_\alpha\{r[\cdot]\}$ and the time of its occurrence along the time series.[5] In this work, we set three different values for α ($\alpha = 0.1$, $\alpha = 0.01$, and $\alpha = 0.001$)

[5] The readers are referred to Ref. [32] for a more detailed analysis and discussion on this feature.

Table 3
Features extracted from the time series data.

Steady-state features	Transient features	
	Rising portion	Decaying portion
ΔR	$\max_k ema_{\alpha=0.001}(r[k])$	$\min_k ema_{\alpha=0.001}(r[k])$
$\|\|\Delta R\|\|$	$\max_k ema_{\alpha=0.01}(r[k])$	$\min_k ema_{\alpha=0.01}(r[k])$
	$\max_k ema_{\alpha=0.1}(r[k])$	$\min_k ema_{\alpha=0.1}(r[k])$

as already used in our previous work [32] to obtain three different feature values starting from the pre-recorded rising portion of the sensor response and three additional features with the same α values for the decaying portion of the sensor response, covering thus the entire sensor response. Fig. 2 shows the typical signal response of a chemical sensor in the presence of 30 ppmv of Acetaldehyde and its ema_α representation for the three different α values. Finally, as this figure illustrates, by applying the above-mentioned transform to each of the 16 channels (sensors) in the pre-recorded time-series, we map the sensor array response into a 128-dimensional feature vector, which resulted from a combination of the 8 features described above $\times 16$ sensors (see Table 3). For a more detailed discussion on these features, the readers are referred to Ref. [32].

3. Drift compensation method

We use an ensemble of classifiers [22,24,25] to detect and cope with sensor drift. Consider a binary classification problem with a set of features x as inputs and a class label (a gas/analyte in our problem) y as output. At every time step t, we receive a batch of examples $S_t = \{(x_1, y_1), \ldots, (x_{m_t}, y_{m_t})\}$ of size m_t. We train a classifier $f_t(x)$, for example a support vector machine (SVM) [33], using the current batch of examples. The final classifier $h_{t+1}(x)$ at time step $(t+1)$ is a weighted combination of classifiers, i.e., $h_{t+1}(x) = \sum_{i=1}^{t} \beta_i f_i(x)$, where $\{\beta_1, \ldots, \beta_t\}$ is the set of classifier weights. Under the assumption that the distribution of examples in the current batch S_t has not changed much from those in the previous batch S_{t-1}, we use the examples in batch S_t to estimate the weights $\{\beta_1, \ldots, \beta_t\}$. There are several ways to estimate these weights. A simple and intuitive way is to assign weights to classifiers according to their prediction performance on batch S_t. Alternatively, we could solve an optimization problem such as

$$\underset{\beta_1,\ldots,\beta_t}{\operatorname{argmin}} \sum_{i=1}^{m_t} \left(\sum_{j=1}^{t} \beta_j f_j(x_i) - y_i \right)^2. \quad (4)$$

It is also possible to minimize a different loss function than the squared loss used above. For example, if the base classifiers are SVMs, we can minimize the hinge loss $L(f(x), y) = \max(0, 1 - yf(x))$. Minimizing the squared loss as shown above has the advantage of computing the solution $\{\beta_1, \ldots, \beta_t\}$ in closed form. The algorithm in its most general form is given below.

Algorithm 1. Algorithm to cope with concept drift

1: **for** $t = 1, \ldots, T$ **do**
2: Receive $S_t = \{(x_1, y_1), \ldots, (x_{m_t}, y_{m_t})\}$
3: Train a classifier (SVM) on S_t
4: Estimate the weights $\{\beta_1, \ldots, \beta_t\}$ using one of the techniques described in the text
5: **end for**
6: Output final classifier: $\{\beta_1, \ldots, \beta_T\}$ and $\{f_1, \ldots, f_T\}$

For multiple classes, we can use one-vs-one or one-vs-all strategy [34] to train a classifier for every pair of classes or for every class respectively. Let L be the number of classes. We can estimate a set of weights $\{\beta_1^c, \ldots, \beta_t^c\}$ for every class $c \in \{1, \ldots, \binom{L}{2}\}$ or $c \in \{1, \ldots, L-1\}$ by solving (4). Alternatively, if we want to assign weights according to the prediction performance of classifiers on the most recent batch, we can simply estimate a single set of weights $\{\beta_1, \ldots, \beta_t\}$ using the multi-class classifier prediction accuracies on every batch. In this case, predictions are made according to a weighted majority voting, i.e.,

$$h_{t+1}(x) = \underset{y \in \{1,\ldots,L\}}{\operatorname{argmax}} \sum_{t: f_t(x)=y} \beta_t. \quad (5)$$

We used this strategy in all our experiments due to its simplicity and ease of implementation. Most importantly, we can directly use the output of existing software—for instance, the widely used LibSVM [35]—to estimate the weights without the need to solve an additional optimization problem.

4. Experimental results

In all our experiments, we trained multi-class SVMs (one-vs-one strategy) with RBF kernel using the publicly available LibSVM software [35]. The features in the training and test datasets were scaled appropriately to lie between -1 and $+1$. The kernel bandwidth parameter γ and the SVM C parameter were chosen using 10-fold cross validation by performing a grid search in the range $[2^{-10}, 2^{-9}, \ldots, 2^4, 2^5]$ and $[2^{-5}, 2^{-4}, \ldots, 2^9, 2^{10}]$, respectively.

We first established the fact that the sensors are drifting and that the drift is degrading the performance of classifiers. We trained a multi-class classifier on data collected during the first two months and tested it on data from the remaining months. Details on the number of measurements collected during each month for a period of three years is given in Table 2.

The classifier's performance measured in terms of prediction accuracy is shown in Fig. 3. We see that the performance gradually degrades with time, which in turn serves as a clear indicator of sensor drift and the way it has affected the accuracy of the classifier. Notice as well that on month 18 (ID: month18, in Table 2), we have only 3 measurements from class 4 (ethylene) and, therefore, the performance suddenly rises to 100%. We consider this issue to be an exception.

We then considered four settings as described below:

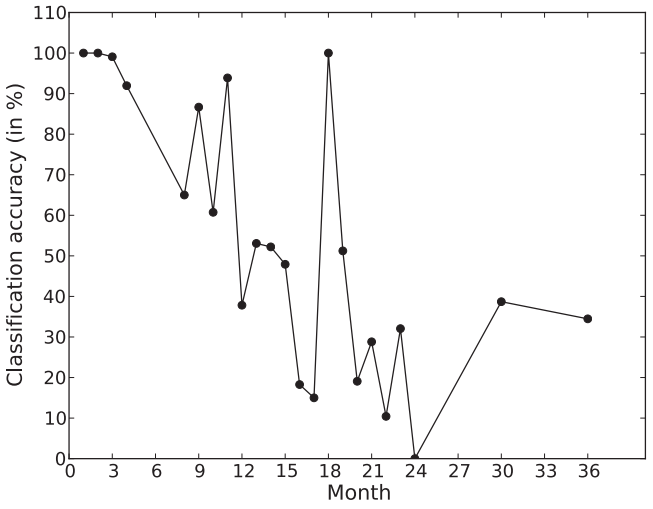

Fig. 3. Performance of the classifiers when trained on months 1 and 2 and tested on months 3–36.

Table 4
Data set details. Each row corresponds to months that were combined to form a batch.

Batch ID	Month IDs
batch1	month1, month2
batch2	month3, month4, month8, month9, month10
batch3	month11, month12, month13
batch4	month14, month15
batch5	month16
batch6	month17, month18, month19, month20
batch7	month21
batch8	month22, month23
batch9	month24, month30
batch10	month36

Setting 1 For every month, we trained a multi-class classifier with data from *only* the previous month and tested it on the current month.

Setting 2 For every month, we trained an ensemble of multi-class classifiers using the method in Algorithm 1.

Setting 3 Same as Setting 2 but with uniform weights on the individual (base) classifiers.

Setting 4 Same as Setting 1 but with the test sets modified using the component correction method (PCA) [4].

The SVM trained in Setting 1 is a strong baseline because it sees the most recent batch of examples which is not corrupted by drifted data from the past. If we were to train an SVM using the entire data until a time point t, then the performance would degrade even further. Note that the ensemble of classifiers trained in Setting 3 with uniform weights essentially uses *all* the data from the past. We would like to emphasize that our experimental setup is robust in the sense that the baseline methods are strong. Similar experimental setup was used in Ref. [36] which was one of the first methods to address concept drift in the machine learning community. Setting 4 is similar to the experimental setup described in Refs. [19,20].

In order to be able to train classifiers under these settings, we need sufficient number of examples in each class and month.[6] We therefore combined measurements from 36 months to form 10 batches in such a way that the number of measurements was as uniformly distributed as possible. Details on the number of measurements in each batch is given in Table 4. Fig. 4 shows the performance of an SVM (black line) trained on batch 1 and tested on batches 2–10. Note that this curve is estimated with the same SVM model used in Fig. 3 but tested on data from batches instead of months.

Once again, we see that the performance of the classifier is degrading with time due to drift. We found similar behaviors when we trained several SVMs on batches 2–5 and tested them on successive batches. These results are again shown in Fig. 4. The complete set of results, i.e., the accuracy of classifiers trained on batches 1–9 and tested on successive batches, is given in Table 5. The first five rows in this table correspond to the results plotted in Fig. 4. These classifiers were eventually used in our ensemble method with appropriate weights. Fig. 5 shows the performance of classifiers under the settings 1, 2, 3, and 4 described above. As expected, classifiers trained under setting 1, i.e., using the most recent batch of examples, perform better than the classifier trained with data from only batch 1. We believe the drop in the accuracy when testing on batches 5 and 6 utilizing the classifiers trained on batches 4 and 5, respectively, is an artefact on the dataset. We suspect that the concentration of the gases used for training has a high impact on the performance of the classifiers over time. Accordingly, as part of an ongoing investigation, we intend to address what concentration levels have to be used to calibrate the gas sensor array. Interestingly, the classifiers trained under Setting 1 were able to cope with drift to some extent and, as mentioned above, we believe they are a natural and strong baseline for any drift-correcting machine learning algorithm to compare against.

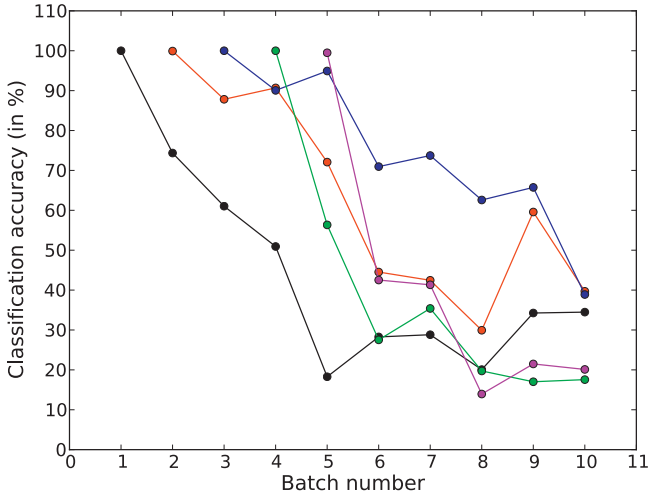

Fig. 4. Performance of the classifiers trained on batches 1–5 and tested on successive batches. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

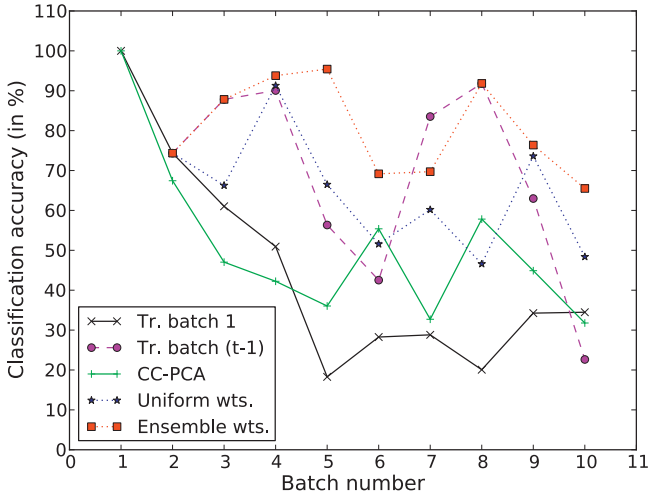

Fig. 5. Performance of the classifiers trained under Setting 1 (•), **Setting 2** (■), Setting 3 (⋆) and Setting 4 (+) described in the text. The black, continuous line corresponds to the setting where a classifier was trained with batch 1 and tested on successive batches. Setting 2 corresponds to our proposed method (Algorithm 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

As shown in Fig. 5, the classifier ensembles were able to perform better than these baseline classifiers when tested on most of the batches with significant improvements in accuracy on several batches. From the indicated figure, we can make the following statements:

- The classifier ensemble performs better than the SVM trained in Setting 1 at batches/time points $t = 4$, 5, 6, 9, and10. The performance is comparable at time points $t = 3$ and $t = 8$. Note that at $t = 2$, we do not have an ensemble and, therefore, every setting is essentially the same. The SVM in Setting 1 performs better than our approach *only* at $t = 7$. As mentioned above, this SVM trained

[6] Note that this requirement is not necessary if we train classifiers in a *purely online* fashion [24,25], but in this paper we focus only on batch training using SVMs.

Table 5
Performance of the classifiers trained on batches 1–9 and tested on successive batches.

Batch ID	Classification accuracy (in %) on batches 2–10								
	2	3	4	5	6	7	8	9	10
batch1	74.36	61.03	50.93	18.27	28.26	28.81	20.07	34.26	34.48
batch2		87.83	90.68	72.08	44.52	42.46	29.93	59.57	39.69
batch3			90.06	94.92	70.96	73.73	62.59	65.74	38.89
batch4				56.35	27.52	35.40	19.73	17.02	17.56
batch5					42.52	41.32	13.95	21.49	20.11
batch6						83.53	88.44	65.74	49.97
batch7							91.84	69.15	54.28
batch8								62.98	37.69
batch9									22.64

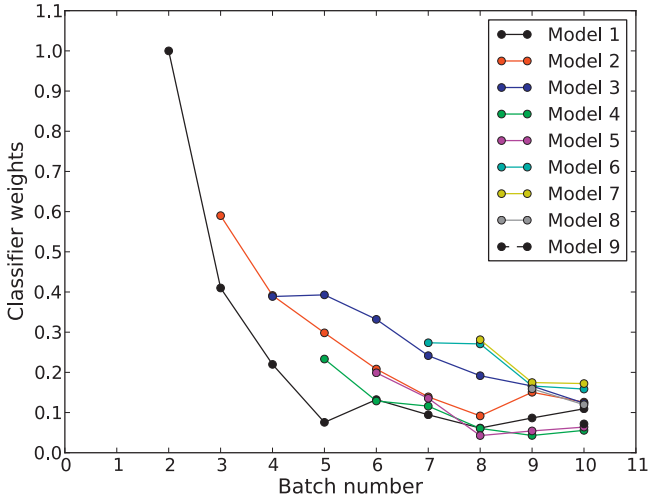

Fig. 6. Classifier weights used in the ensembles (model 1 through model 9). At every point on the *x*-axis (time/batch) the corresponding points in the *y*-axis are the weights of the individual classifiers used in the ensemble. Notice that these weights sum up to 1 at every point of time/batch (*x*-axis) and that the contribution of an individual classifier to the ensemble gradually decreases with time.

in Setting 1 is a very strong baseline and, thus, performing better than or as well as this setting is a positive result.

- The classifier ensemble performs better than the classifier trained in Setting 3 at *all* the time points, i.e., from $t = 3$ to $t = 10$. This result clearly demonstrates the effectiveness of ensemble methods to automatically detect and cope with sensor drift. Fig. 6 shows how the classifier weights used in the ensemble change with time.
- Finally, although the component correction method (Setting 4) helped to slightly improve the performance of SVM trained in Setting 1, it performed worse than all the other methods on most of the batches. Note that we applied the component correction method six times by treating every one of our six gases as a reference gas. In Fig. 5, we plot only the results obtained from the best reference gas, i.e., at every batch/time point, we apply the component correction method six times and report the best performance in terms of classifier accuracy. Fig. 7 shows the results for all the reference gases.

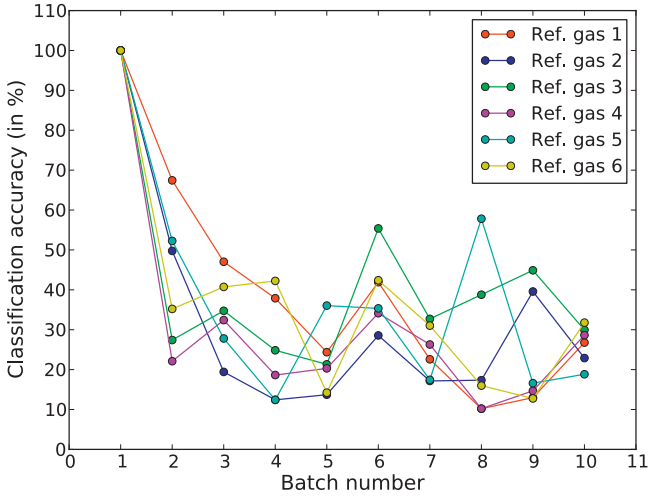

Fig. 7. Performance of the classifier trained under Setting 4 with component correction. The individual plots correspond to the performance of classifier trained with batch 1 and tested on batches at subsequent time points after applying the component correction method for every one of the six reference gases.

As a final remark, we would like to mention that our proposed method for sensor drift compensation is a *supervised* learning algorithm. If we have trained a classifier ensemble until a certain point in time, say T, and would like to use it for future time points, then we have to repeatedly add classifiers to the ensemble and update their weights so that it can track the drifting direction. To do this, we need labels for the data since the classifiers are trained in a supervised manner. In a more realistic scenario like real-time deployment where we may not have labels for data in future time points, then we have to update the classifier ensemble in a *semi-supervised* way [37]. For instance, at time $T+1$, we may use the classifier ensemble trained until time T to provide labels for the new batch of data, and then we add another classifier to the ensemble using these new labels. Under the assumption that the data at time $T+1$ has not drifted much from the data at time T, this is a reasonable strategy to fill in missing labels. We leave the design and investigation of such semi-supervised learning algorithms [37] for future work.

5. Conclusions

Over a period of 3 years we have collected a comprehensive dataset using an array of 16 metal-oxide gas sensors to analyze the problem of sensor drift. The data was collected under tightly controlled conditions for 6 different gases at several concentrations ranging from 10 to 1000 ppmv. To the best of our knowledge this is one of the most comprehensive datasets available for the design and development of drift compensation methods, which will be available online as a supplementary material of this article, hopefully aiding in algorithmic improvements in the future. We have reported significant benefits using machine learning techniques based on ensembles of classifiers as a drift correction method for chemo-sensory applications. As we have demonstrated, the proposed method not only manages to mitigate well the effect of drift in chemical sensors, but also outperforms competing methods. An important observation from our sensor drift analysis is that the compensations made by the proposed method remain valid for long periods of time. The 6-month gap between the recordings from the first 30 months and the last recorded month did not invalidate the compensation of drift on the same task, despite the fact that the

sensors were exposed to poisoning, as they were powered-off during this period of time.

Finally, the level of generality that the proposed drift counteraction method attains is high from the following perspective. By design, the drift counteraction method using classifier ensembles operates on samples collected on-the-fly in an online fashion which is important for real-time operations. The technique does not make any assumptions about the nature of drift, and it is also agnostic to the base classifier used in the gas identification task. As a consequence, this solution has the potential to be of use in many other applications where data is collected over an extended period of time with drifting underlying distributions, and, most importantly, it can be readily translated to more realistic gas sensing applications, including but not limited to the identification and localization of chemical analytes in complex environments (e.g., in a wind tunnel), gas distribution mapping, and gas plume tracking using robotic platforms. We held this issue as an arguable position here that we would address in further studies.

Acknowledgements

This work has been supported by U.S. office of Naval Research (ONR) under the contract number N00014-07-1-0741, by Jet Propulsion Laboratory under the contract number 2010-1396686, and by the US Army Medical and Materiel Command and by the United States Army Research Institute of Environmental Medicine (USARIEM), under contract number W81XWH-10-C-0040 in collaboration with Elintrix. The authors also thank Joey Reeds and Travis Wong, from Elintrix, for the helpful and fruitful discussions during the elaboration of this work as well as Joanna Zytkowicz for proofreading and revising the manuscript.

References

[1] J.W. Gardner, P.N. Bartlett, Electronic Noses: Principles and Applications, Oxford University Press, New York, 1999.
[2] K.C. Persaud, G. Dodd, Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose, Nature 299 (1982) 352–355.
[3] M. Holmberg, F. Winquist, I. Lundstrom, F. Davide, C. DiNatale, A. D'Amico, Drift counteraction for an electronic nose, Sensors and Actuators B: Chemical 36 (1–3) (1996) 528–535.
[4] T. Årtusson, T. Eklöv, I. Lundström, P. Mårtensson, M. Sjöström, M. Holmberg, Drift correction for gas sensors using multivariate methods, Journal of Chemometrics 14 (5–6) (2000) 711–723.
[5] W. Göpel, K.-D. Schierbaum, Definitions and typical example, in: Chemical and Biochemical Sensors, Part I, VCH, Weinheim, 1992, pp. 1–28.
[6] F. Davide, C.D. Natale, M. Holmberg, F. Winquist, Frequency analysis of drift in chemical sensors, in: in: Proceedings of the 1st Italian Conference on Sensors and Microsystems, 1996, pp. 150–154.
[7] M. Zuppa, C. Distante, P. Siciliano, K.C. Persaud, Drift counteraction with multiple self-organising maps for an electronic nose, Sensors and Actuators B: Chemical 98 (2–3) (2004) 305–317.
[8] M. Holmberg, T. Artusson, Drift compensation, standards, and calibration methods, Handbook of Artificial Olfaction Machines, WILEY-VCH Weinheim, Germany, 2003, pp. 325–346.
[9] A. Hierlemann, R. Gutierrez-Osuna, Higher-order chemical sensing, ACS Chemical Reviews 108 (2008) 563–613.
[10] A.-C. Romain, J. Nicolas, Long term stability of metal oxide-based gas sensors for e-nose environmental applications: an overview, Sensors and Actuators B: Chemical 146 (2) (2010) 502–506.
[11] A.-C. Romain, P. André, J. Nicolas, Three years experiment with the same tin oxide sensor arrays for the identification of malodorous sources in the environment, Sensors and Actuators B: Chemical 84 (2–3) (2002) 271–277.
[12] Figaro USA, Inc., http://www.figarosensor.com/.
[13] W. Göpel, New materials and transducers for chemical sensors, Sensors and Actuators B: Chemical 18 (1–3) (1994) 1–21.
[14] N. Yamazoe, New approaches for improving semiconductor gas sensors, Sensors and Actuators B: Chemical 5 (1–4) (1991) 7–19.
[15] K. Dobos, R. Strotman, G. Zimmer, Performance of gas-sensitive Pd-gate mosfets with SiO_2 and Si_3N_4 gate insulators, Sensors and Actuators 4 (1983) 593–598.
[16] M. Roth, R. Hartinger, R. Faul, H.E. Endres, Drift reduction of organic coated gas-sensors by temperature modulation, Sensors and Actuators B: Chemical 36 (1–3) (1996) 358–362.
[17] A. Vergara, E. Llobet, J. Brezmes, P. Ivanov, C. Cane, I. Gracia, X. Vilanova, X. Correig, Quantitative gas mixture analysis using temperature-modulated micro-hotplate gas sensors: selection and validation of the optimal modulating frequencies, Sensors and Actuators B: Chemical 123 (2007) 1002–1016.
[18] J.-E. Haugen, O. Tomic, K. Kvaal, A calibration method for handling the temporal drift of solid state gas-sensors, Analytica Chimica Acta 407 (1–2) (2000) 23–39.
[19] R. Gutierrez-Osuna, Drift reduction for metal-oxide sensor arrays using canonical correlation regression and partial least squares, in: Proceedings of the 7th International Symposium On Olfaction & Electronic Nose, 2000.
[20] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, A. Perera, Drift compensation of gas sensor array data by common principal component analysis, Sensors and Actuators B: Chemical 146 (2) (2010) 460–465.
[21] B. Schölkopf, A.J. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (5) (1998) 1299–1319.
[22] H. Wang, W. Fan, P.S. Yu, J. Han, Mining concept-drifting data streams using ensemble classifiers, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.
[23] J.Z. Kolter, M.A. Maloof, Dynamic weighted majority: A new ensemble method for tracking concept drift, in: Proceedings of the International IEEE Conference on Data Mining, 2003.
[24] J.Z. Kolter, M.A. Maloof, Using additive expert ensembles to cope with concept drift, in: Proceedings of the Twenty-Second International Conference on Machine Learning, 2005.
[25] J.Z. Kolter, M.A. Maloof, Dynamic weighted majority: An ensemble method for drifting concepts, Journal of Machine Learning Research 8 (Dec) (2007) 2755–2790.
[26] Bronkhorst High-Tech B.V, http://www.bronkhorst.com/.
[27] Airgas, Inc., http://www.airgas.com/.
[28] The LabVIEW Environment, http://www.ni.com/labview/.
[29] A. Vergara, M.K. Muezzinoglu, N. Rulkov, R. Huerta, Information-theoretic optimization of chemical sensors, Sensors and Actuators B: Chemical 148 (1) (2010) 298–306.
[30] A. Pardo, S. Marco, J. Samitier, Nonlinear inverse dynamic models of gas sensing systems based on chemical sensor arrays for quantitative measurements, IEEE Transactions on Instrumentation and Measurement 47 (3) (1998) 644–651.
[31] E. Llobet, J. Brezmes, X. Vilanova, J.E. Sueiras, X. Correig, Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick-film tin oxide gas sensor array, Sensors and Actuators B: Chemical 41 (1–3) (1997) 13–21.
[32] M.K. Muezzinoglu, A. Vergara, R. Huerta, N. Rulkov, M.I. Rabinovich, A. Selverston, H.D. Abarbanel, Acceleration of chemo-sensory information processing using transient features, Sensors and Actuators B: Chemical 137 (2) (2009) 507–512.
[33] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297.
[34] R. Rifkin, A. Klautau, In defense of one-vs-all classification, Journal of Machine Learning Research 5 (2004) 101–141.
[35] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines, Software. Available at http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.
[36] R. Klinkenberg, T. Joachims, Detecting concept drift with support vector machines, in: Proceedings of the International Conference on Machine Learning, 2000.
[37] O. Chapelle, B. Schölkopf, A. Zien (Eds.), Semi-Supervised Learning, MIT Press, Cambridge, MA, 2006.

Biographies

Alexander Vergara (Ph.D., 2006 – Universitat Rovira i Virgili) is a Postdoctoral Scientist at the BioCircuits Institute, UC San Diego. His work mainly focuses on the use of dynamic methods for the optimization of micro gas-sensory systems and on the building of autonomous vehicles that can localize odor sources through a process resembling the biological olfactory processing. His areas of interest also include signal processing, pattern recognition, feature extraction, chemical sensor arrays, and machine olfaction.

Shankar Vembu (Ph.D., 2010 – University of Bonn) is a Postdoctoral Researcher at the BioCircuits Institute, UC San Diego. His research interests are in machine learning algorithms and their applications.

Tuba Ayhan (M.S., 2010 – Technical University of Istanbul) is a Ph.D. student at the Department of Electronics and Communication Engineering, Technical University of Istanbul. Her research interests are embedded systems, FPGA, electronic noses and machine olfaction. Tuba carried out most of her work on this article while visiting the BioCircuits Institute at UCSD.

Margaret A. Ryan (Ph.D., 1987 – University of Massachusetts at Amherst) is an Associate Research Scientist at the Jet Propulsion Laboratory, California Institute of Technology. Before coming to JPL in 1989, she was a research scientist at the Centre National de la Recherche Scientifique in Bellevue, France, and at the Solar Energy Research Institute (now National Renewable Energy Lab) in Golden, Colorado. Her work in both places included modification of metal oxide and semiconductor electrodes for photoelectrochemical energy conversion, photoelectrochemical processing of solids and photochemistry. She came to JPL in 1989, and has worked primarily in the areas of chemical sensing and thermal-to-electric energy conversion. She is the Principal Investigator on the NASA Electronic Nose Technology Development Task. Her work at JPL has concentrated on two primary areas of research: (a) development of chemical sensors

and devices for air quality monitoring, including the JPL ENose, silicon carbide hydrocarbon sensors and colorimetric sensors for ozone and other oxidants, and (b) investigations of metals and metal alloys for use in high temperature energy conversion devices, including the Alkali Metal Thermal to Electric Converter (AMTEC) and thermoelectric devices. Amy is now working at the U.S. Department of Energy.

Margie L. Homer (Ph.D., 1993 – University of California Los Angeles) is an Associate Research Scientist at the Jet Propulsion Laboratory, California Institute of Technology. Physical Chemist Margie Homer is the Co-Investigator of the Electronic Nose project at JPL. Her research interest includes chemical sensors.

Ramón Huerta (Ph.D., 1994 – Universidad Autónoma de Madrid) is an Associate Research Scientist at the BioCircuits Institute, UC San Diego. His areas of expertise include dynamic systems, artificial intelligence, and Neuroscience. His work deals with the development algorithms for the discrimination and quantification of complex multidimensional times series, model building to understand the information processing in the brain, and chemical sensing and machine olfaction applications based on bio-inspired technology. Ramón has co-authored over 80 articles in peer-reviewed journals at the intersection of computer science, physics, and biology.