The Journal of Engineering

Special Issue: Industry 4.0: the DIGITAl Transformation in the Engineering Findings (DIGITATE)

IET Journals
The Institution of Engineering and Technology

Iterative complex network approach for chemical gas sensor array characterisation

eISSN 2051-3305
Received on 2nd July 2018
Accepted on 17th December 2018
E-First on 24th April 2019
doi: 10.1049/joe.2018.5125
www.ietdl.org

Angelo Cardellicchio[1], Angela Lombardi[2] ✉, Cataldo Guaragnella[1]

[1]Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, Via Orabona 4, Bari, Italy
[2]Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Via Orabona 4, Bari, Italy
✉ *E-mail: angela.lombardi@ba.infn.it*

Abstract: Gas sensor arrays, also known as e-noses, are used in several heterogeneous fields, ranging from environmental monitoring to food quality control. Often, these measurement systems operate within dynamic environments and are subject to conditions which may dramatically vary over time. Furthermore, the response of an e-nose is influenced by several parameters, whose interactions may be complex and highly non-linear. Therefore, in this study, the authors propose a complex network approach to model the overall interaction pattern of e-noses. They show that this approach can significantly improve the understanding of the overall behaviour of e-noses, and can be used as a basis to optimise the design of these measurement systems.

1 Introduction

Chemical gas sensors compounds analysis has been extensively involved in a broad range of contexts including environmental monitoring [1], food quality control [2, 3], and medical applications [4, 5]. Basically, it requires the setup of chemosensor arrays (i.e. e-nose arrays) to identify a chemical compound and estimate its quantity. In this context, the design of highly sensitive, top-performing and fault-tolerant devices is of paramount importance. However, it should bear in mind, that sensor arrays are placed in environments with characteristics that vary over time. Therefore, dynamic conditions and the responses of the sensor arrays to such variations should be considered, to correctly recognise the chemical analytes.

Recently, a data-driven approach has been proposed to identify gas chemical compounds from pre-recorded measurements collected at different distances from the source and under a range of flow conditions [6]. The work of Vergara and colleagues introduced new approaches for seeking better calibration techniques of e-nose arrays as well as monitoring and maintenance systems for outdoor devices. In particular, they presented a large dataset of experiments from ten chemical agents, recorded by utilising chemoresistive metal–oxide-based chemical sensors located at different positions in the test field. Such measurements are then employed to train a classifier to represent a great variety of possible conditions for better discriminating chemical analytes in a test-set scenario.

In this work, we present a methodology for exploring the interaction of sensor responses to verify whether an optimal device configuration exists to identify different compounds under varying flow conditions. In details, we adopted a complex network approach [7] to model the correlation of the signals' responses to the same input between each couple of sensors of the array. This approach is implicitly multivariate and allows to consider simultaneously the existence of different factors that could affect the sensing of the chemical analytes such as turbulent and environmental conditions (e.g. pressure, temperature, humidity) as well as technical specifications of the device (e.g. a number of sensors, reciprocal position etc.).

The main purpose of the exploratory analysis carried out in this paper is to investigate a method for identifying a minimum set of configuration parameters and flow conditions that maximise the distances among the sensors' responses to different chemical analytes. The remaining of the paper is organised as follows. Section 2 presents a brief review of the main data-driven analysis methods. Section 3 describes the data used in this work together with the adopted methodology. The most relevant results achieved are reported in Section 4. Finally, Section 5 outlines the conclusions of the work and future developments.

2 Related works

Studies concerning the suitability of e-noses to discern among different classes of chemical compounds have been focused on both the *mathematical tools* and the *preprocessing* of data used for classification [8].

Data preprocessing is the first step needed to obtain a signature of the sample [1, 8]. Usually, this step involves several data transformations, such as denoising or standardisation, to properly input signals to the mathematical tools employed in the following steps. Other times, a feature extraction procedure [9, 10] may be needed to extract meaningful features in complementary domains, such as the frequency domain.

Then, *exploratory data analysis* (EDA) [11] is usually performed. EDA is extremely important to assess data quality: i.e. an expert can determine whether data are suitable for the classification task using visual tools, such as polar plots [12]. However, such analysis may involve data mapped into a high-dimensional space, which is difficult (if not impossible) to visually explore. Therefore, dimensionality reduction techniques, such as PCA [13], may be used to perform either an initial cluster analysis [14], to reduce the dimensionality of data [15], or to allow for a simpler data visualisation [16]. We also note that, apart from PCA, feature selection and data visualisation methods such as t-SNE [17] may be well suited to be used throughout the exploratory data analysis.

The following step is to use classification techniques to discern from data. Often, data are considered as independent and identically distributed and, therefore, traditional classifiers can be used. Less often, however, data present dependencies over time, and therefore time series clustering or forecasting techniques may apply.

As for the first type of classifier, simpler classification techniques rely on the notion of *distance* between data points. That is, as each data point can be mapped to an n-dimensional space, where n is the number of features, the distance between data points can be evaluated as an index of similarity, as in [18]. However, we must argue that such approaches are more similar to ranking procedures, where each data point can be ranked as closer or farther from another one; to provide for a more traditional

classification, some boundaries within the feature space must be introduced. These boundaries are strictly related to the behaviour of more advanced techniques.

The simplest of these techniques is a *k-nearest neighbour* (k-NN), which is widely used thanks to its simplicity and good performances [19]. k-NN simply computes Euclidean distance between data points, assigning each unlabelled sample to one of the classes on which the classifier has been trained. However, as noted in [20], we note that, especially when the number of features is high, Euclidean distance may not be the best choice to compute meaningful distances between points. Traditional statistical approaches are also used, such as *discriminant function analysis* [21, 22] and *partial least squares* (PLS) [23].

Artificial neural networks (ANNs) have also been used for classification tasks [14, 24]. More advanced approaches have also used ensemble learning [25], with [6] which proposes the use of an *inhibitory SVM* (ISVM) [26]. Specifically, an ISVM trains one classifier f_i for each class i available within the dataset and compares its output to the average output of the ensemble of classifiers.

The approaches which consider data coming from e-noses as time series focus on different mathematical tools. A popular tool which has been used for such tasks are *Time-Delay Neural Networks* (TDNNs) [27–29], and also *recurrent neural networks* have been employed for such tasks [30]. Another perspective has been given by Schleif *et al.* [31], which proposes the use of *generative topographic mapping trough time* as an unsupervised model for time series inspection. Another interesting approach deals with the need to perform forecasting and prediction in real-time, starting from data acquired by the e-nose. Specifically, Fonollosa *et al.* [32] provide predictions starting from data acquired in real time by using *reservoir computing* [33].

3 Experimental section

3.1 Dataset description

In our analysis, we used the dataset defined in [6], freely available at the UCI Machine Learning Repository [34]. This dataset contains data acquired by a set of nine identical e-noses, each composed of eight different MO-X sensors. Specifically, the selected sensors belong to the TGS26XX family and are sensitive to hydrocarbons, hydrogen, nitrogen, sulphur compounds, and carbon monoxide.

The dataset was acquired in a wind tunnel test-bed facility, and the e-noses were positioned at six different locations, normal to the wind direction, uniformly distributed throughout the tunnel. At each trial, the authors injected within the tunnel a selection of chemical compounds at a predefined concentration. Chemical compounds are acetaldehyde, acetone, ammonia, benzene, butanol, CO with two different concentrations, ethylene, methane, methanol and toluene. All these substances are characterised by high volatility. The chosen sensors appear to be suited to respond to the compounds. However, a proper chemical characterisation is beyond the scope of this paper.

Trials were performed following the classic methodology [32], and the entire reference-measurement-cleaning cycle lasts about 260 s. Throughout the trials, two different conditioning parameters were considered. The first is the *heater voltage* V_h, directly related to the temperature of the active surface of each sensor, while the second is the *airflow speed S*, i.e. the speed of the fan within the wind tunnel. For V_h, five values (i.e. {4.0, 4.5, 5.0, 5.5, 6.0} V) were considered, while for S only three values (i.e. {0.10, 0.21, 0.34} m/s).

Within their experiments, the authors found out that chemical sensors had weaker responses if exposed to higher values of S, and central measurement lines showed better robustness if compared to the rest of the wind tunnel facility. We can see in Section 3.3 how these results are confirmed, and therefore reinforced, by our methodology.

3.2 Methodology

We use the mathematical framework of complex networks, which can be employed to model complex systems, such as a wind tunnel facility. In our case, the e-nose is modelled as a complex network $G = (V, E)$, whose nodes $V = v_1, \ldots, v_n$ are represented by the sensors embedded into the array, and edges $E = e_{12}, \ldots, e_{nn}$ are defined according to a certain principle.

In our case, $n = 72$, as there is one node for each sensor within the e-nose. Therefore, there will be $(n \times (n-1))/2$ edges, since our network is supposed to be dense.

We defined edges in a way which is similar to the one adopted to define functional connectivity in brain networks [35], i.e. the edge e_{ij} between nodes v_i and v_j is defined using the correlation coefficients between the time series y_{t_i} and y_{t_j} acquired for nodes v_i and v_j, respectively. Formally

$$e_{ij} = \tau\left(y_{t_i}, y_{t_j}\right) \tag{1}$$

In (1), τ is the Kendall correlation coefficient. We choose to use τ over other correlation coefficients (e.g. Pearson r or Spearman ρ), because we could suppose neither linearity nor monotonicity between y_{t_i} and y_{t_j}.

Intuitively, the configuration of the network G gives an idea about the overall response of the e-nose to a specific substance. As an example, if the responses of the e-nose to two different substances are uncorrelated, these substances will be easily discriminated by the respective signatures.

To evaluate the discriminative capabilities of such an approach, we proceed as follows:

(i) Firstly, we identified how many degrees of freedom are available within data. In our case, there are three of them. The first two are determined by the conditioning parameters, i.e. the values for V_h and S. The third one is given by the layer l at which the measurement is performed. In our experiments, we tried to cover a large part of possible cases, determining a set of networks G_{vsl}, with $v \in V_\mathrm{h}, s \in S, l \in L$. Each network models the response of the signal to various experiments for each chemical compound when values for voltage heater, fan speed and level are fixed.
(ii) Once the networks are determined, a signature was obtained for each layer of each substance, represented by the edge set E. Then, we evaluated the distance among these signatures by using cosine distance $D = (1 - C)$, where C is the cosine similarity. Following our intuition, high values of cosine distance allow to easily discern between the overall responses of the e-nose to each couple of chemical substances.

3.3 Discussion

Figs. 1–3 show the count of maximum distances between trials varying V_h and L for each of the three fixed $s = 0.10$, 0.21 and 0.34 m/s. Interestingly, as it can be seen, the behaviour of the system appears to be consistent with the value fixed for V_h. Specifically, it can be noted that, for a fixed value of $s = 0.10$ m/s, the value of V_h which results in the most discriminative is $v = 4.0$ V, while for $s = 0.21$ and 0.34 m/s the most discriminative v is 4.5 V.

Therefore, we can observe in Figs. 4–6, the values of the cosine distance fixing V_h at the previously defined most discriminant values. Visually, the heat maps with 'warmer' colours are more discriminative, as the distance values D are on average higher. These findings are better summarised in Table 1 which shows the cosine distance values for each of the three fan speeds for the most discriminative V_h and varying L.

From the above considerations, we can confirm some of the results given in [6]; in particular, S results in the most relevant conditioning parameter, and we can also give some more hints on the relevance of L and V_h. Specifically, as expected, L seems to be relevant for higher values of S due to the fact that the initial layers are mostly affected by the aspects already depicted in [6].

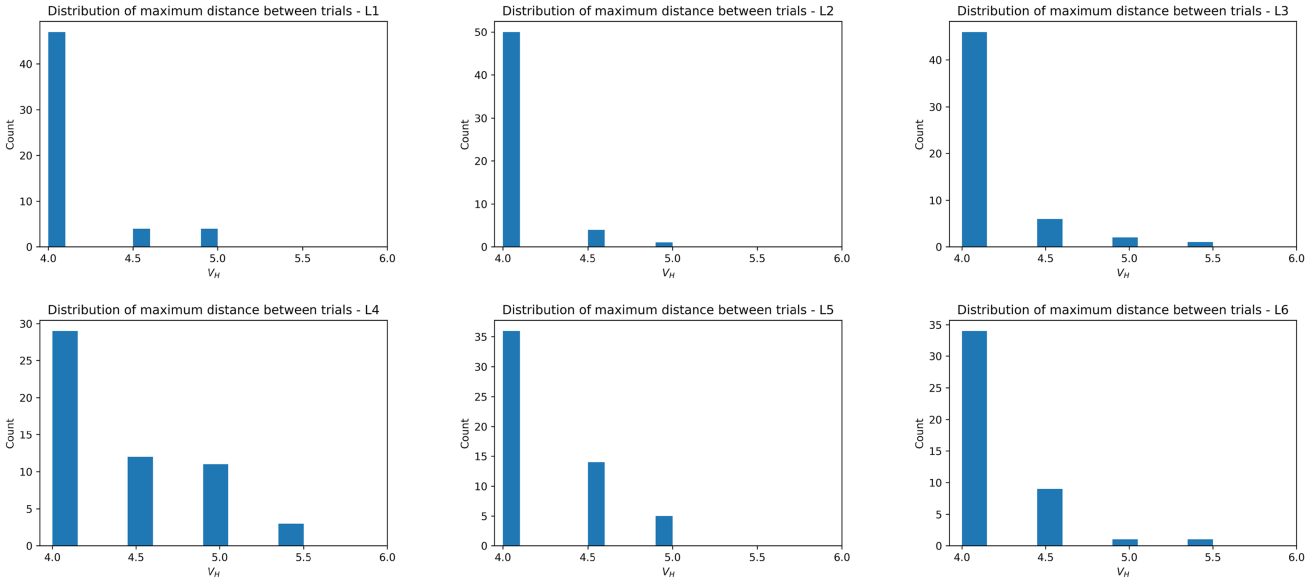

Fig. 1 *Count of maximum distances varying V_h and L with fixed $S = 0.10$ m/s*

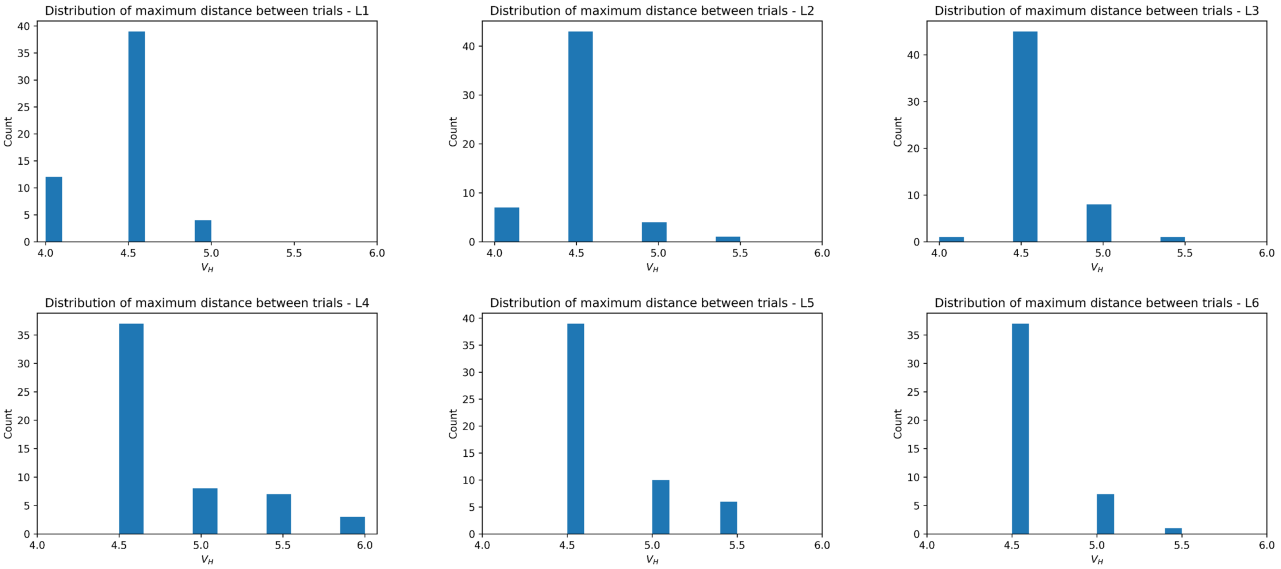

Fig. 2 *Count of maximum distances varying V_h and L with fixed $S = 0.21$ m/s*

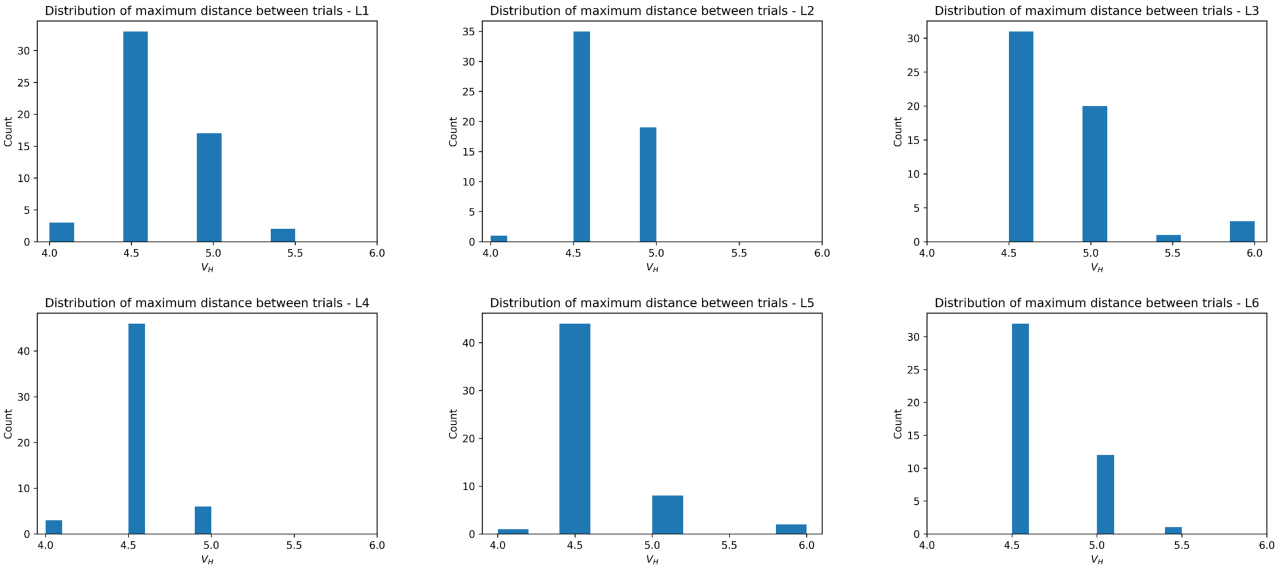

Fig. 3 *Count of maximum distances varying V_h and L with fixed $S = 0.34$ m/s*

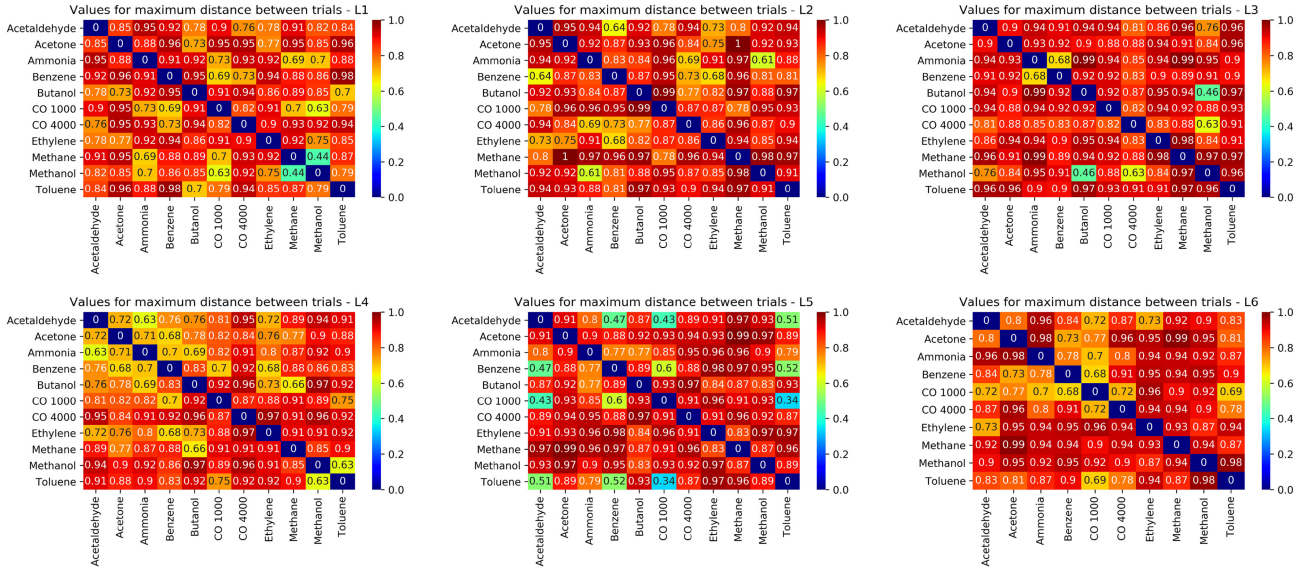

Fig. 4 *Maximum distances for the most discriminative V_h varying L with fixed $S = 0.10$ m/s*

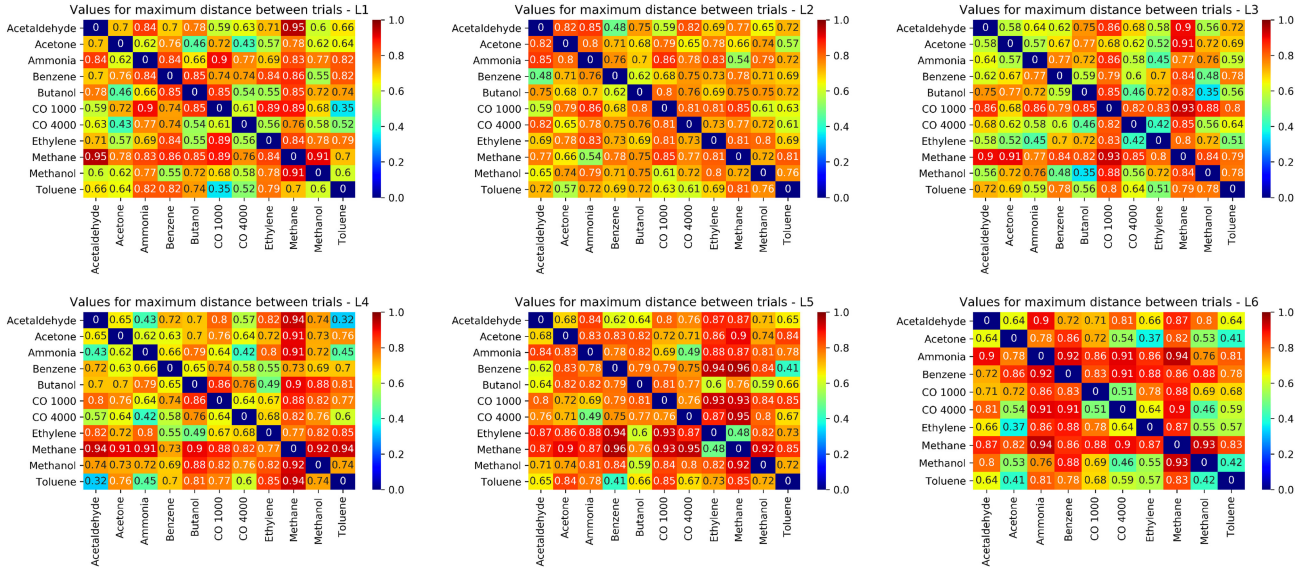

Fig. 5 *Maximum distances for the most discriminative V_h varying L with fixed $S = 0.21$ m/s*

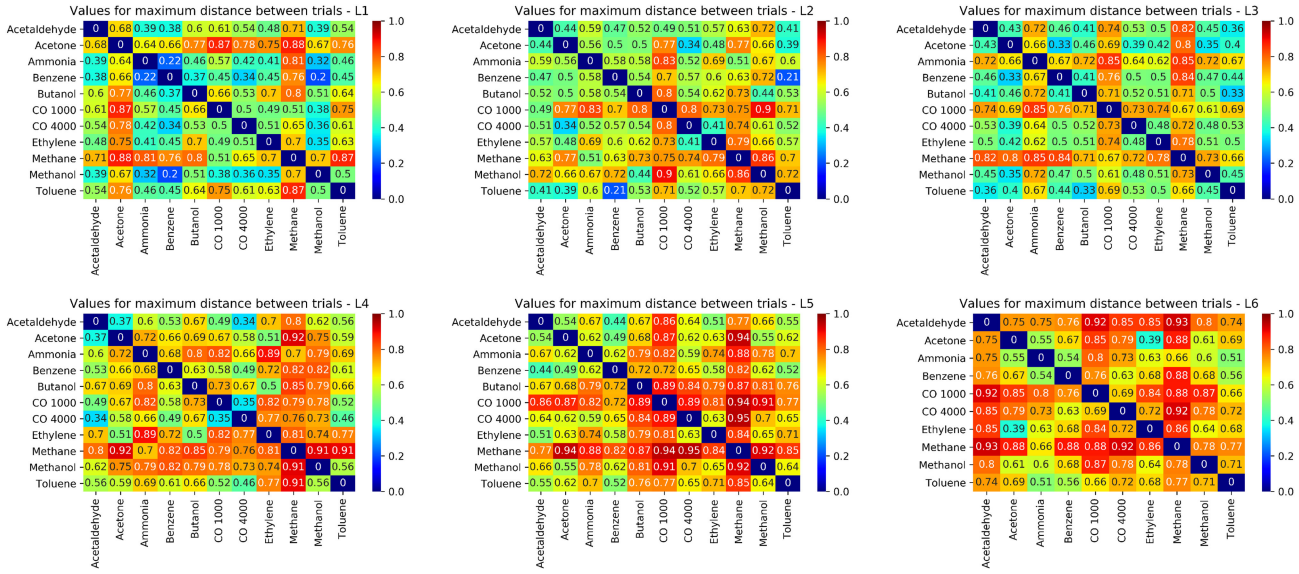

Fig. 6 *Maximum distances for the most discriminative V_h varying L with fixed $S = 0.34$ m/s*

Table 1 Average cosine distance value for the most discriminative V_h varying L

Fan speed, m/s	Optimal V_h	L_1	L_2	L_3	L_4	L_5	L_6
0.10	4.0	0.85	0.88	0.89	0.83	0.86	0.87
0.21	4.5	0.71	0.73	0.70	0.72	0.78	0.74
0.34	4.5	0.57	0.61	0.59	0.68	0.72	0.73

Furthermore, from the bar plots, it clearly appears that the value of the voltage heater, for this set of sensors, is strongly relevant, and dependent upon the specific value of the fan speed. It can also be argued that at low speeds, the maximum discrimination among the chemical agents is achieved, as the system is less affected by turbulent phenomena.

4 Conclusions and future works

In this paper, we gave some insights on how complex networks can be used to analyse the behaviour of an e-nose when several parameters vary. On the one side, our experiments confirm previous results; on the other side, we give some other perspective on how the other conditioning parameters condition the discrimination results, therefore allowing for a proper design of the measurement experiment.

However, several improvements can be made in our approach. First, in future developments, we will consider the whole network as a multiplex network [36], with the various layers given either by the variation of each one of the conditioning parameters or by the definition of the behaviour of the networks over time. Specifically, in the second case, the behaviour of the network over time can lead to a better characterisation of the reference-measurement-cleaning steps. Furthermore, data coming from our analysis can feed a classifier, whose performances can be evaluated for substance classification.

The code used in this work is publicly available on GitHub at the following address: https://github.com./anhelus/envlab.

5 References

[1] Capelli, L., Sironi, S., Del Rosso, R.: 'Electronic noses for environmental monitoring applications', *Sensors*, 2014, **14**, (11), pp. 19979–20007

[2] Loutfi, A., Coradeschi, S., Mani, G.K., *et al.*: 'Electronic noses for food quality: A review', *J. Food Eng.*, 2015, **144**, pp. 103–111

[3] Baietto, M., Wilson, A.D.: 'Electronic-nose applications for fruit identification, ripeness and quality grading', *Sensors*, 2015, **15**, (1), pp. 899–931

[4] Wilson, D.: 'Recent progress in the design and clinical development of electronic-nose technologies', *Nanobiosensors Disease Diagn.*, 2016, **5**, pp. 15–27

[5] Dragonieri, S., Pennazza, G., Carratu, P., *et al.*: 'Electronic nose technology in respiratory diseases', *Lung*, 2017, **195**, (2), pp. 157–165

[6] Vergara, A., Fonollosa, J., Mahiques, J., *et al.*: 'On the performance of gas sensor arrays in open sampling systems using inhibitory support vector machines', *Sens. Actuators, B*, 2013, **185**, pp. 462–477

[7] Boccaletti, S., Latora, V., Moreno, Y., *et al.*: 'Complex networks: structure and dynamics', *Phys. Rep.*, 2006, **424**, (4–5), pp. 175–308

[8] Srivastava, A.K.: 'Detection of volatile organic compounds (VOCs) using SnO2 gas-sensor array and artificial neural network', *Sens. Actuators B, Chem.*, 2003, **96**, (1–2), pp. 24–37

[9] Ehret, B., Safenreiter, K., Lorenz, F., *et al.*: 'A new feature extraction method for odour classification', *Sens. Actuators B, Chem.*, 2011, **158**, (1), pp. 75–88

[10] Distante, C., Leo, M., Siciliano, P., *et al.*: 'On the study of feature extraction methods for an electronic nose', *Sens. Actuators B, Chem.*, 2002, **87**, (2), pp. 274–288

[11] Tukey, J.W.: 'Exploratory data analysis', vol. 2, 1977

[12] Brezmes, J., Ferreras, B., Llobet, E., *et al.*: 'Neural network based electronic nose for the classification of aromatic species', *Anal. Chim. Acta*, 1997, **348**, (1–3), pp. 503–509

[13] Gardner, J.W.: 'Detection of vapours and odours from a multisensor array using pattern recognition part 1. Principal component and cluster analysis', *Sens. Actuators B, Chem.*, 1991, **4**, (1–2), pp. 109–115

[14] Zhang, Q., Xie, C., Zhang, S., *et al.*: 'Identification and pattern recognition analysis of Chinese liquors by doped nano ZnO gas sensor array', *Sens. Actuators B, Chem.*, 2005, **110**, (2), pp. 370–376

[15] Tomchenko, A.A., Harmer, G.P., Marquis, B.T., *et al.*: 'Semiconducting metal oxide sensor array for the selective detection of combustion gases', *Sens. Actuators B, Chem.*, 2003, **93**, (1–3), pp. 126–134

[16] Choi, N.J., Kwak, J.H., Lim, Y.T., *et al.*: 'Classification of chemical warfare agents using thick film gas sensor array', *Sens. Actuators B, Chem.*, 2005, **108**, (1–2), pp. 298–304

[17] Maaten, L.V.D., Hinton, G.: 'Visualizing data using t-SNE', *J. Mach. Learn. Res.*, 2008, **9**, pp. 2579–2605

[18] Nicolas, J., Romain, A.C., Wiertz, V., *et al.*: 'Using the classification model of an electronic nose to assign unknown malodours to environmental sources and to monitor them continuously', *Sens. Actuators B, Chem.*, 2000, **69**, (3), pp. 366–371

[19] Carmel, L., Levy, S., Lancet, D., *et al.*: 'A feature extraction method for chemical sensors in electronic noses', *Sens. Actuators B, Chem.*, 2003, **93**, (1–3), pp. 67–76

[20] Aggarwal, C.C., Hinneburg, A., Keim, D.A.: 'On the surprising behavior of distance metrics in high dimensional space'. Int. Conf. on Database Theory, Berlin, Heidelberg, 2001, pp. 420–434

[21] Martın, Y.G., Oliveros, M.C.C., Pavón, J.L.P., *et al.*: 'Electronic nose based on metal oxide semiconductor sensors and pattern recognition techniques: characterisation of vegetable oils', *Anal. Chim. Acta*, 2001, **449**, (1–2), pp. 69–80

[22] Oliveros, M.C.C., Pavón, J.L.P., Pinto, C.G., *et al.*: 'Electronic nose based on metal oxide semiconductor sensors as a fast alternative for the detection of adulteration of virgin olive oils', *Anal. Chim. Acta*, 2002, **459**, (2), pp. 219–228

[23] Sohn, J.H., Hudson, N., Gallagher, E., *et al.*: 'Implementation of an electronic nose for continuous odour monitoring in a poultry shed', *Sens. Actuators B, Chem.*, 2008, **133**, (1), pp. 60–69

[24] Lee, D.S., Jung, J.K., Lim, J.W., *et al.*: 'Recognition of volatile organic compounds using SnO2 sensor array and pattern recognition analysis', *Sens. Actuators B, Chem.*, 2001, **77**, (1–2), pp. 228–236

[25] Vergara, A., Vembu, S., Ayhan, T., *et al.*: 'Chemical gas sensor drift compensation using classifier ensembles', *Sens. Actuators B, Chem.*, 2012, **166**, pp. 320–329

[26] Huerta, R., Vembu, S., Amigó, J.M., *et al.*: 'Inhibition in multiclass classification', *Neural Comput.*, 2012, **24**, (9), pp. 2473–2507

[27] Zhang, H., Balaban, M.Ö., Principe, J.C.: 'Improving pattern recognition of electronic nose data with time-delay neural networks', *Sens. Actuators B, Chem.*, 2003, **96**, (1-2), pp. 385–389

[28] Yamazaki, A., Ludermir, T.B., De Souto, M.C.P.: 'Classification of vintages of wine by artificial nose using time delay neural networks', *Electron. Lett.*, 2001, **37**, (24), pp. 1466–1467

[29] De Vito, S., Castaldo, A., Loffredo, F., *et al.*: 'Gas concentration estimation in ternary mixtures with room temperature operating sensor array using tapped delay architectures', *Sens. Actuators B, Chem.*, 2007, **124**, (2), pp. 309–316

[30] Duckett, T., Axelsson, M., Saffiotti, A.: 'Learning to locate an odour source with a mobile robot'. Proc. 2001 ICRA, IEEE Int. Conf. on Robotics and Automation, 2001, 2001, vol. 4, pp. 4017–4022

[31] Schleif, F.M., Hammer, B., Monroy, J.G., *et al.*: 'Odor recognition in robotics applications by discriminative time-series modeling', *Pattern Anal. Appl.*, 2016, **19**, (1), pp. 207–220

[32] Fonollosa, J., Sheik, S., Huerta, R., *et al.*: 'Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring', *Sens. Actuators B, Chem.*, 2015, **215**, pp. 618–629

[33] Lukosevicius, M., Jaeger, H.: 'Reservoir computing approaches to recurrent neural network training', *Comput. Sci. Rev.*, 2009, **3**, (3), pp. 127–149

[34] Dua, D., Karra Taniskidou, E.: '*UCI machine learning repository*' (University of California, School of Information and Computer Science, Irvine, CA, 2017). Available at http://archive.ics.uci.edu/ml

[35] Bullmore, E., Sporns, O.: 'Complex brain networks: graph theoretical analysis of structural and functional systems', *Nat. Rev. Neurosci.*, 2009, **10**, (3), p. 186

[36] Kivela, M., Arenas, A., Barthelemy, M., *et al.*: 'Multilayer networks', *J. Complex. Netw.*, 2014, **2**, (3), pp. 203–271