Odor control map: Self organizing map built from electronic nose signals and integrated by different instrumental and sensorial data to obtain an assessment tool for real environmental scenarios

S. Licen[a,*], G. Barbieri[b], A. Fabbris[b], S.C. Briguglio[a], A. Pillon[c], F. Stel[c], P. Barbieri[a]

[a] Department of Chemical and Pharmaceutical Sciences, University of Trieste, Via L. Giorgieri 1, 34127, Trieste, Italy
[b] ARCO SolutionS s.r.l., Spin-off Company of the Department of Chemical and Pharmaceutical Sciences, University of Trieste, Via L. Giorgieri 1, 34127, Trieste, Italy
[c] Agenzia Regionale per la Protezione dell'Ambiente del Friuli Venezia Giulia (ARPA-FVG), via Cairoli 14, 33057, Palmanova, UD, Italy

ARTICLE INFO

Accepted 19 February 2018

Keywords:
Electronic nose
Odor
Self organizing map
Pattern recognition
Ambient air
Dynamic olfactometry

ABSTRACT

Olfactory nuisances are an issue of growing concern considering that people's awareness about the effect of pollution on health and environment is increasing and that the perception of odor is related to a possible warning situation. A recent review concluded that odor assessment has to be faced by an integrated multi-tool strategy. In this paper we discuss the building of a model by means of a chemometric approach based on artificial neural networks known as Self Organizing Maps. These are applied to data collected by electronic nose continuous monitoring. The Self Organizing Map output was subjected to a second level clusterization by k-means algorithm. The cluster interpretation (i.e. air types classification in terms of "malodor"/"odor free" attributes) is achieved by crosslinking data produced by different instrumental and sensorial approaches, allowing us to establish the Frequency-Intensity-Duration odor characteristics for every identified air type.

In order to elucidate our approach we focused our work on a four months survey at a residential site close to an integral cycle steel plant in Trieste (Italy). Odor Control Map proved to be a promising tool to provide valuable visualization support for following the dynamic evolution of the system with time. It allowed us to, for example, identify the relationships among sensor responses in different air types; follow the changes of air types with time. identify possible malodor sources; experimentally evaluate the frequency and duration of air types classified as malodorous. Furthermore, this first application highlighted the possible method improvements that have to be tested in different real environmental scenarios to obtain more robust and refined models. Considering that different annoyances (e.g., odor, noise, presence of specific chemical compounds) can cause possible synergistic health effects, Odor Control Map is a suitable tool to integrate data deriving from different and independent analysis/monitoring to obtain a more comprehensive knowledge on complex environmental phenomena involving dwellings close to industrial plants.

1. Introduction

The problem of olfactory nuisances is an issue of growing concern considering that people's awareness about the effect of pollution on health and environment is increasing and that the perception of odor is related to a possibly warning situation. Furthermore, a number of effects on human health, disjoined by toxic levels of pollutants, have been reported [1–3].

Brancher et al. [4] critically reviewed the odor regulations of 28 countries around the world. Their main conclusion was that odor assessment has to be tackled by an integrated multi-tool strategy.

Odor impact is usually described by five characteristics collectively known as FIDOL: frequency (F), intensity (I), duration (D), offensiveness (O), and location (L). A detailed description of each feature can be found elsewhere [4–6]. Information about odor intensity (I) can be gained by the dynamic olfactometry technique, which is a standardized sensorial method [7,8] aimed at evaluating intensity in terms of odor concentration. In Europe, this is expressed

* Corresponding author.
E-mail addresses: slicen@units.it (S. Licen), gbarbieri@arcosolutions.eu (G. Barbieri), a.fabbris@olfattometria.eu (A. Fabbris), saracarmela.briguglio@phd.units.it (S.C. Briguglio), alessandra.pillon@arpa.fvg.it (A. Pillon), fulvio.stel@arpa.fvg.it (F. Stel), barbierp@units.it (P. Barbieri).

as European odor units per cubic meter ($ou_E\ m^{-3}$). One $ou_E\ m^{-3}$ is the amount of odorant(s) that, when evaporated into 1 cubic meter of neutral gas in standard conditions, elicits a physiological response from a panel (detection threshold) equivalent to that elicited by one European Reference Odor Mass (EROM). One EROM is a conventional quantity value equivalent to 123 μg *n*-butanol evaporated in 1 cubic meter of neutral gas in standard conditions, which produces a concentration of 0.040 μmol/mol.

However, it has to be taken into account that complaints increase especially in case of the presence of an intermittent and unpredictable annoyance. In the case of an odorant this is more stressful than an exposition to a persistent, moderate nuisance which usually produces a saturation effect in the olfactory system.

Therefore, a correct evaluation of the frequency (F) and the duration (D) of odor episodes is the key to assess the effects of malodors on the population. Gaining information about the abovementioned odor annoyance features by means of air sampling at the receptor and dynamic olfactometry measures with high frequency would be hardly feasible, very expensive and time-consuming. On the contrary, an electronic nose (e-nose) allows the collection of instrumental data with high temporal resolution.

E-noses have been widely applied in food quality analysis [9], but there is a lower number of papers on the application of e-noses for the measurement of industrial emissions [10]. In both the abovementioned applications the e-nose is usually calibrated with specific mixtures of chemicals or with samples of air collected at the source. A number of different methods for data processing have been already presented and applied [11]. Among the inconveniences enumerated by Marco et al. on the application of previously calibrated e-noses, which can change their responses in terms of sensitivity there are: (i) sensor drifting and aging due to, for example, temperature and humidity parameters or sensor poisoning by chemicals in field experiments, (ii) change in the environmental conditions (from laboratory conditions to field measurements), (iii) change in the background environmental conditions which can affect the target source identification capability of the instrument. When monitoring in a real environmental context it is possible that the source/sources of annoyances have not be already been identified and/or that there is no previous knowledge about them. In such cases the Netherlands Standardization Institute (NEN) suggests to use e-noses to record patterns produced by sensors continuously exposed to ambient air mixtures of compounds and to identify similarities among patterns which can then be linked to citizen complaints and perceptions [12].

In this framework, in the present paper we discuss the building of a model based on artificial neural networks known as Self Organizing Maps (SOM), applied to e-nose data collected by continuous monitoring in ambient air at a site where olfactory annoyances are frequently reported. The Self Organizing Map is a neural network algorithm for exploratory data analysis and pattern recognition [13]. It is based on an unsupervised learning method, thus no previous knowledge on data patterns is needed. There are few examples in the scientific literature of the application of the SOM algorithm to e-nose signals. These mainly concern the food quality field [14,15], while there are also some reports on laboratory e-nose optimization [16,17].

On the other hand, the application of SOM algorithm to data collected in relation to environmental assessment issues has grown over the last decade encompassing different fields: air [18,19]; water [20,21]; soil [22,23] and sediment [24,25].

In order to elucidate our approach we focused our work on a survey on an integral cycle steel plant in Trieste (Italy). The cycle comprises a series of linked processes such as coke distillation, iron ores sintering and smelting in a blast furnace for the production of pig iron [26,27]. This industrial hotspot has already been subject matter of a number of studies regarding sediments [28], water [29], ambient air volatile organic compounds [30] and ambient air particulate matter [31–33]. In the last decade, the municipality of Trieste has registered several hundreds of complaints per year about obnoxious odors by the resident population, thus, in collaboration with the regional environmental protection agency (ARPA-FVG), an e-nose was positioned near apartment buildings in proximity to the steel works.

The chemometric approach we adopted allowed us to operate a second abstraction level by applying a clusterization algorithm, such as k-means clustering or hierarchical clustering [34], to the output of the Self Organizing Map algorithm in order to obtain a distribution of the e-nose data in a small number of clusters (i.e., air types). The cluster interpretation (i.e. air types classification in terms of "malodor"/"odor free" attributes) is achieved by crosslinking data produced by different instrumental and sensorial approaches, allowing the establishment of the Frequency-Intensity-Duration odor characteristics for every identified air type.

With regard to the type of source considered in the case study, few publications focused on odor emissions by steel plants can be found in the scientific literature: in the city of Terni (Italy) Capelli et al. [35,36], using the combination of olfactometric analyses at the sources and dispersion modelling, found that the major source of odorous emissions in the steel plant was the primary emissions from the furnaces. Bootsma et al. [37] used e-noses implemented with a four sensors array to monitor odor emissions from an integrated steel plant in IJmuiden (Netherlands). In the first phase they used a network of five e-noses to collect data coupled with olfactometric analyses. Then, according to the results obtained in the first phase, 25 e-noses were trained to detect the specific odor patterns of the plant emissions (coke oven batteries, blast furnaces, sinter plant) and the e-nose network was installed onsite. Ultimately the e-nose recordings positively matched to the citizens' complaints.

2. Materials and methods

2.1. Site description

Trieste is a city located in NE Italy by the Adriatic Sea and has about 200,000 inhabitants. In one of its districts, which has about 12,000 inhabitants for a population density of 8300 inh/km^2, there is a fully integrated steel plant. The dwellings of the district are positioned in close proximity to the plant: the sampling point (SLS) was positioned at apartment buildings which are located only 180 m from the plant. At the boundary of the works, 200 m from the SLS there is a regional EPA sampling point (RFI) which is used for monitoring the steel plant performance [38] (Fig. 1).

2.2. Continuous monitoring by use of e-nose

The electronic nose used for this study was a MSEM-32 Environmental Monitor System device purchased from Sensigent (Baldwin Park CA –U.S.A.) In the selected configuration, the MSEM-32 multi sensor environmental monitor implemented 19 chemoresistive sensors with four MOS sensors, from now on named S3-S4-S20-S21 and fifteen polymer/black carbon nano composite sensors [39] (named from S5 to S19), plus two more electrochemical sensors (named SULFUR and NITROGEN). The e-nose was placed at the SLS (Fig. 1). The monitoring period started on Friday, 5th June 2015 and ended on Wednesday, 30th September 2015. We chose the period from June to September because it is characterized by the presence of sea breezes blowing from the sea to the inland, i.e., from the steel plant to the city districts/civil dwellings [30]. The e-nose registered data-per-minute for every sensor thereby obtaining a

Fig. 1. Map of the Trieste district which hosts the integrated steel plant. The sampling site (SLS), the regional environmental protection agency sampling site (RFI), the boundary of the steel plant, the position of the coke oven batteries (C.O.) and the position of the blast furnace (B.F.) are highlighted.

vector of 21 values for every minute. Sensor signals are reported as relative resistance changes ($\Delta R/R$).

2.3. Air sampling

The air samples were collected in 8L bags prepared in-house using Nalophan^{TM} for the bag and Teflon^{TM} for the pipes. Both automatic and manual sampling systems were used. The core of the sampling systems is a vacuum pump which allows collection of the sample without any contact with the pump parts except for the bag [8], in compliance with EN 13725:2003 [7]. A number of citizens who live close to the sampling point were asked to inform (by phone call or SMS) the operators about the presence of an obnoxious odor as soon as it had been perceived. Immediately after a complaint reception, the automatic sampling system (OdorPrep, by Lab Service Analytica S.r.L., Italy) was remotely activated. As a general rule, when the operators went to the site to withdraw the sample, they collected another air sample by use of the manual sampler (Vacuum pump Sampler, by Osmotech Srl, Italy). The sampling lasted two minutes both for remote sampling and the manual one. The automatic sampling system and the e-nose were placed very close in space at the sampling site, thus they contemporarily collected the "same" air. The manual sampling was also carried out in close proximity to the e-nose.

2.4. Odor concentration measurements

The air samples were analyzed in compliance with EN 13725:2003 [7] within 6 h after sampling as suggested by German VDI 3880 norm [40], which is stricter than EN 13725:2003 on this issue. The analysis was carried out by use of a dynamic olfactometer (WOLF by ArcoSolutions s.r.l., Italy) [41].

The instrument presents the odor samples diluted with odor free air, at precise ratios, to a panel of human assessors, selected according to a reference gas (*n*-butanol) [7]. WOLF olfactometer works using the yes/no method, in which each panelist sniffs from a single port and communicates if the odor is detected or not. The odor concentration is then calculated as the geometric mean of the odor threshold values of each panelist and it is expressed in European odor units per cubic meter ($ou_E\ m^{-3}$).

2.5. Citizens' complaints addressed to the municipality

In addition to the complaints of the citizens voluntarily involved in the present study (see par 2.3) we examined the list of complaints communicated by the resident population to the municipality. We selected only the complaints which were classified as odor annoyance and that came from inhabitants living in the same street in which the SLS sampling point was positioned.

2.6. Wind speed and direction data

The hourly wind speed and direction data collected at a site (Molo Fratelli Bandiera, a synoptic weather station – in an open position by the sea 2 km from the plant – considered representative of the meteorological conditions of the city), were obtained from the regional environmental protection agency (OSMER-ARPA) website [42].

2.7. Pollutant data

The hourly benzene, CO and H_2S data collected at a site (RFI – see Fig. x) positioned at the boundary of the steel plant (at about 200 m from SLS) were obtained from the regional environmental protection agency (ARPA-FVG) website [43]. The site is used by ARPA-FVG for monitoring the steel plant performance [38].

2.8. Self organizing map features

A Self Organizing Map is constituted by a two-dimensional array of neurons (also called units), which are vectors of scalars related to the sensor responses. Neurons are usually represented as squares or hexagons in a bi-dimensional map. The array dimensions have to be set a priori: some heuristic rules were proposed by Vesanto et al. [44] The choices for the present SOM evaluation will be discussed in detail in par. 3.4. The algorithm is then fed with the experimental data, the model training is based on the "winner-take-all" selection rule: basically each neuron of the grid is initialized as a random unit vector of *n*-values (with "n" equal to the number of variables) and all the experimental vectors are presented to each neuron. The algorithm selects the neuron which best matches the input vector in terms of similarity (usually measured by the Euclidean distance). The winner neuron is also called the best matching unit (BMU). Once found, the vector weights of the BMU and its neighbors (included in a given *radius* from the BMU) are adjusted to reduce their difference to the input vector. In this way each neuron can represent similar vectors and it adjusts itself during the training process. Once each input vector has been presented to the neurons, the first run (epoch) of the training process is finished. This presentation and adjustment is repeated several times so that the SOM describes progressively in a more accurate way the experimental vector variability. The output is a Self Organizing Map in which similar vectors are mapped close together on the grid. Each final unit is represented by an *n*-dimensional vector, in which "n" is equal to the original number of variables and each experimental vector can be associated to a unit in terms of similarity (Euclidean distance). According to Vesanto [45] and Himberg et al. [46] the SOM can be explored to observe:

- the Unified Distance Matrix (U-matrix) representation [47], which visualizes the distances between the neurons and highlights borders between map regions with different charac-

teristics, thus allowing identification of map areas with relatively homogeneous neurons;
- the number (*hits*) of experimental vectors which are represented by every single map unit,
- the distribution of the values of every single experimental variable on the map (*heatmap*), showing how each one of the n original variables relates to the others in the SOM.

The Self Organizing Map can also be explored to identify possible outliers [48]. Outlier detection can be achieved by inspecting the quantization errors (QEs) distribution on the map. Every experimental vector is associated to a unit and the measure of how well the unit represents that vector is expressed in terms of the QE, i.e. the QE accounts for the accuracy of the match of each experimental vector to the corresponding unit. A high QE value for an experimental vector means that it can be an outlier observation according to the SOM model.

2.9. Calculations

SOM calculations, clustering classification and Sammon's projection were performed in the Matlab 6.5 (MathWorks, Inc.) computing environment, implementing the SOM toolbox [49]. SOM outputs exploration and SOM visualization were performed using in-house scripts in R software environment [50] implemented by the "openair" package [51]. In addition, wind speed and direction data, pollutant data and citizens' complaints data were elaborated in R environment. The site map was prepared in R environment implemented by the "ggplot2" package [52] and "ggmap" packages [53]. The trajectory animation was prepared in R environment implemented by the "animation" package [54].

3. Results and discussion

3.1. Building an odor control map: general idea

The core of the model is the unsupervised classification of the e-nose data. This data mining approach can be very effective for a first characterization of the dynamics of the site when no previous detailed characterization about odor impacts at the receptors, as well as about odor emissions data at the source, are available. We chose the Self Organizing Map algorithm (see par. 2.8 for details) for elaborating the e-nose data and obtaining a Self Organizing Map.

With the intention to classify the map regions in terms of malodorous air/"odor free" air and to identify possible odor sources, we integrated the SOM model with data collected by means of other approaches or devices. Ancillary information is provided as odor concentration of air samples, wind speed and direction data and pollutant data collected by the regional environmental protection agency monitoring network. Citizens' complaints records were also used.

The only common variable between the abovementioned data and e-nose signals is date/time. When in the following text we will refer to a map unit which is associated to an odor concentration value, it means that we identified the e-nose data contemporarily registered during the air sampling, then we identified the map unit which best represented that e-nose data and we eventually associated the odor concentration value to the map unit.

Since the olfactometric measurement on air samples is the sole European accepted method to measure odor concentration, before the SOM training the whole e-nose data set was split in training and test data, so that e-nose data corresponding (date/time) to the olfactometric measures referring to odor nuisance episodes were present in both the training set and what we called an "odor test set" (see details in par. 3.3). The latter has been used to assess the recall ability of the trained Self Organizing Map [55] i.e. to check the ability of the model to map similar objects in the same SOM map area. Thus, the training produces a map where independent experimentally determined odor concentrations can be associated to map units, allowing us (a) to achieve a first, rough classification of different areas of the map corresponding to malodorous and odor free ambient air and (b) to check the consistency of the classification by projecting the odor test set vectors onto the map.

A more refined classification was obtained by performing a k-means cluster analysis [34] on SOM neurons data set. Khedairia et al. [56,57] reported the advantages derived from a two-stage clusterization of experimental data (1st- SOM, 2nd- k-means) instead of the application of only one of the two algorithms. The SOM algorithm allows identification of recurrent e-nose data patterns with reduced noise and visualization of the data in a 2D map which can be explored as described in par. 2.8. Subsequently, a reduced number of map regions with similar properties can be achieved by k-means clustering. The application of k-means clustering directly to the experimental data would lead to a list of values (cluster numbers) associated to the data with poor visualization possibilities when compared to SOM.

The clusters were characterized both using the olfactometric analyses (as described previously in this paragraph) and identifying the map units corresponding to the date/time of the citizens' complaints. The sensor patterns in clusters were calculated and depicted using box-plots to inspect possible differences in the sensor ranges among the clusters.

With the aim to corroborate the cluster classification between odorous/non odorous episodes, the wind speed and direction data and pollutant data collected by the regional protection agency in the same period were used as follows. Each cluster represents a number of e-nose experimental data collected at specific date/time values. The abovementioned devices outputs were grouped according to the same date/time values (i.e. according to the clusters) and the groups obtained were represented in box-plots. This step led to a more detailed classification which allowed evaluation of the malodor duration and frequency in the monitored period, and to hypothesize possible different sources.

3.2. Odor measurements

In Table below odor measurements results are reported (16 samples). The values measured ranged from 118 $ou_E\ m^{-3}$ to 500 $ou_E\ m^{-3}$ with a mean of 326 $ou_E\ m^{-3}$ and a median of 360 $ou_E\ m^{-3}$.

These relatively high values of $ou_E\ m^{-3}$ were related to samples collected at the SLS at about 180 m from coke oven batteries, during the revamping of the steel plant after a change in the ownership and in a season when the sea breeze is usually more intense, pushing air-masses toward the monitoring site. The evidence of these values was then taken into account for a dedicated monitoring provision in the IPCC Environmental Integrate Authorization of the steel plant. The odor measurements were less than the number of the signaling volunteers due to two different principal factors: (a) panel test members were not always available within 6 h (see par. 2.4) and (b) lack of power supply on site for the remotely activated odor bag sampler due to rather frequent and random black-outs that happen locally close to the industrial plant.

3.3. Data pre-treatment

The e-nose data were cleaned by removing the vectors-per-minute which occurred in proximity of instrumental zero calibration cycles or occasional black-outs, which caused spikes in the sensor signals. The data removed were the 0.3% of the total available, producing 150 072 vectors, for a total number of sensor values above 3 million. Noise filtering algorithms were not

Table 1
list of air sampling times and related outcome of the olfactometric analysis.

Sample	Date/hour	Sampling technique	OUe m^{-3}	Data subset
1	28/07/2015 12:21	RS	430	TrS
2	*28/07/2015 13:06*	*RS*	*360*	*OtS*
3	03/08/2015 13:24	RS	380	TrS
4	*03/08/2015 14:56*	*MS*	*290*	*OtS*
5	04/08/2015 12:18	RS	380	TrS
6	*04/08/2015 13:05*	*MS*	*300*	*OtS*
7	*02/09/2015 15:23*	*RS*	*170*	*OtS*
8	02/09/2015 15:34	RS	118	TrS
9	03/09/2015 10:36	RS	204	TrS
10	03/09/2015 15:15	MS	360	TrS
11	*03/09/2015 15:37*	*MS*	*360*	*OtS*
12	15/09/2015 11:34	MS	380	TrS
13	*15/09/2015 11:45*	*MS*	*360*	*OtS*
14	22/09/2015 12:36	RS	360	TrS
15	*22/09/2015 14:32*	*MS*	*500*	*OtS*
16	22/09/2015 15:00	MS	270	TrS

RS = remotely activated sampling; MS = manual sampling; TrS = training set: *OtS = odor test set.*

applied since the way in which the SOM algorithm works already reduces noise [34]. As explained in par. 3.1 the olfactometric measurements on air samples carried "objective information" about malodor therefore we decided to extract a small "odor test set" from the whole data to verify the ability of the SOM map to coherently classify malodor events confirmed by an independent analysis method. The odor test set was built by selecting an interval of ten minutes centered on the collection time of the samplings highlighted in italics in Table 1 (7 odor samplings), resulting in a matrix containing 77 vectors. The remaining 9 odor samplings (corresponding to 99 vectors) were contained in the training set containing a total of 149 995 vectors. The training set was normalized across the variables to have zero mean and unit standard deviation.

3.4. Self organizing map of e-nose data

While building the SOM, we chose an hexagonal lattice as it provides more neighbors within a given radius of a node with respect to a rectangular one, as well as a smoother map with a better representation of the topology of the data set [34]. The SOM parameters were set accordingly to Vesanto et al. [44]. The number of neurons was selected as five times the square root of the number of experimental data (in the present study represented by 149 995 vectors of 21 variables, which are the sensor responses) while the *ratio* between the map side lengths was set equal to the *ratio* between the two largest eigenvalues of the data set. The *ratio* for our data set was 2.3. The actual side lengths were set so that their product was as close as possible to the number of map units evaluated by the abovementioned heuristic formula. The result was a 67×29 lattice. The SOM was trained by feeding the algorithm with the training data set and using the batch training algorithm. Linear initialization started along the two greatest eigenvectors. After initialization, the SOM was trained in two phases: first a rough training and then a fine-tuning. The "rough neighborhood radius" was set equal to 3 and the "fine neighborhood radius" to 1. The neighborhood function was Gaussian. The learning rate function type was set as "reciprocally decreasing", starting from a learning rate of 0.5 for the rough phase and of 0.05 for the fine-tuning phase. The number of epochs was set according to Sinha et al. [58], resulting in 1 (rough phase) + 1(fine tuning) epochs. From now on the SOM obtained with the described set of parameters will be named "regular" SOM.

The SOMToolbox allows also us to train a "small" SOM with the number of units equal to the 0.25 times the number of units evaluated with the abovementioned Vesanto heuristic formula. In this way we obtained a SOM with a 32×15 lattice. The number of epochs was selected as for the "regular" one, obtaining 1 + 1 epochs.

Table 2
ordinal statistics of the quantization errors obtained for the training and odor test sets.

	Training set	Odor test set
min	0.20	0.65
1st quantile	0.95	1.31
median	1.42	1.81
3rd quantile	2.09	2.11
max	11.96	3.80

To evaluate the SOM quality two parameters are typically used: the average quantization error q_e (i.e. the average distance between each data vector and its BMU) which accounts for map resolution; and the topographic error t_e (i.e. the proportion of all data vectors for which first and second choice BMUs are not adjacent units) which measures the topology preservation. There are no reference values for these parameters, nevertheless they can be used to compare SOM maps resulting from different starting run parameters. The values for the "regular" SOM were $q_e = 1.659$ and $t_e = 0.095$, while for the "small" one they were $q_e = 1.625$ and a $t_e = 0.090$. Considering that the smaller the q_e and t_e values, the better the SOM quality, we chose the "small" SOM as the model for our data.

From now on the outputs related to the numerical elaboration of the "small" SOM will be discussed. In Fig. 2a the unified distance matrix is presented. The distance is reported in grayscale from dark gray (short distance) to light gray (long distance). The map shows boundaries (lighter color) between short distance regions (darker color – highlighting portions of the map characterized by highly similar neurons).

The basic statistics for the hits represented by a single unit (see par. 2.8) were 1st quantile = 114, median = 220, 3rd quantile = 394. The hits distribution on the map is shown in Fig. 2b. The maximum number of hits (4001) is represented by the unit symbolized as a black filled hexagon. That unit represents the most recurrent pattern of sensor data recorded at the e-nose site.

Boundaries of the map were present near the edges (Fig. 2a) but the hits (Fig. 2b) were not gathered at the edges, thus the present SOM did not show a "Map edge distortion" as defined by Fujita et al. [59]. Moreover, in order to check the reliability of the map in terms of topological relations between the neurons, Himberg et al. [46] suggest to compute and plot the Sammon's projection of the neuron space into a three dimensional space. If the projection is highly twisted or folded the map is unreliable. The Sammon's projection of our SOM trained as described before in this paragraph was neither twisted nor folded (see Supplementary Material).

The heatmaps (see par. 2.8) for each of the 21 variables are presented in Fig. 3, where four tones of gray characterize quartiles of a single sensor values. The sequence of heatmaps in the figure derives from the hierarchical clustering of the variables. The hierarchical clustering of the variables of the SOM was done by applying the complete linkage method, considering euclidean distances. Visual inspection of the dendrogram (see Supplementary material) and heatmaps (Fig. 3) led to group the e-nose sensors as follows: S20-MOS (singleton); S3-MOS, SULFUR-electrochemical, NITROGEN-electrochemical (group 1); S4-MOS (singleton); S21-MOS (singleton); S7,S5,S9,S19,S17,S10,S14,S11,S13,S15,S18-nanocomposites (group 2); S12,S6,S8,S16-nanocomposites (group 3) [60].

The quantization errors (QEs – see par. 2.8) for the training data set were calculated, obtaining the ordinal statistics values reported in Table 2, in which it can be seen that there was a considerable difference between the 3rd quantile and the maximum value. Thus we computed, for every map unit, the average QEs of the experimental vectors assigned to that unit. Fig. 2d reports hexagons with

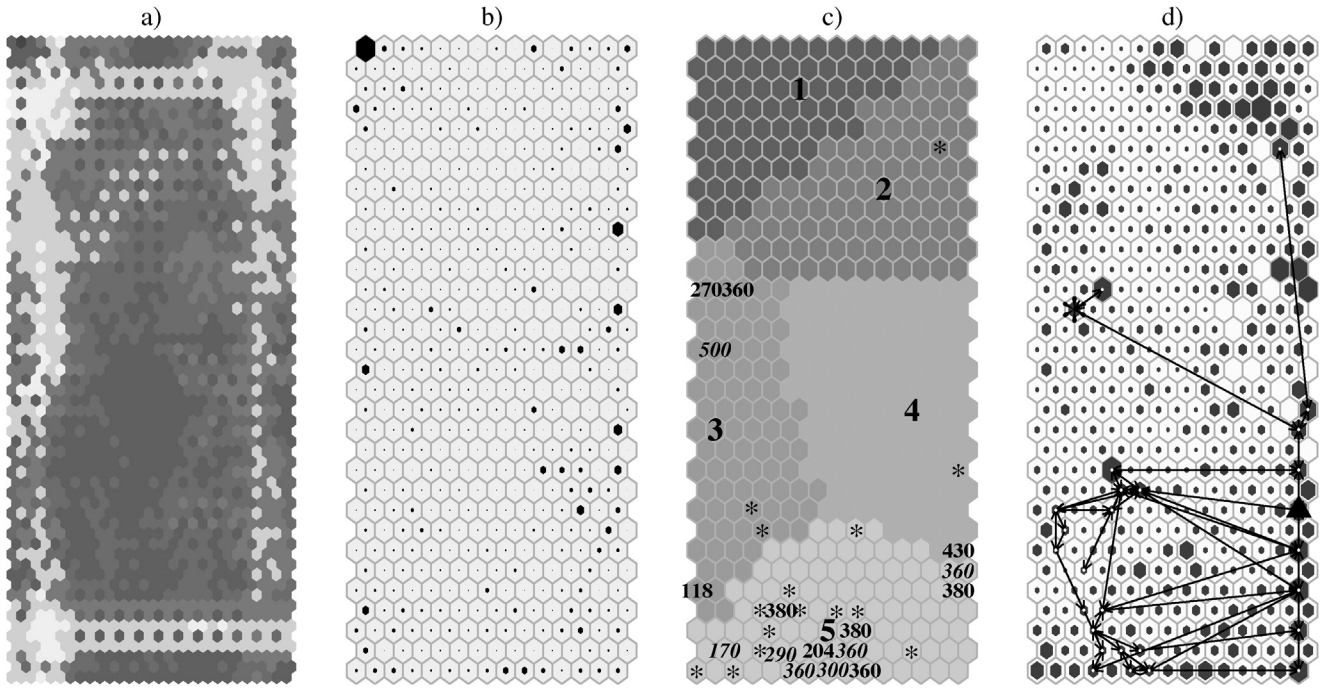

Fig. 2. (2a) Unified distance matrix (U-matrix) map, the distance is reported in grayscale from dark gray (short distance) to light gray (long distance). (2b) Hits map. The filling of the hexagons is proportional to the number of hits. The black hexagon highlights the unit which represented the maximum number of hits (4001). (2c) SOM map, the filling of the hexagons represents the cluster split of the units. The cluster numbers are positioned at the units which are the centroids of the clusters respectively. The units which correspond with the air samplings are labelled with the obtained odor concentration in $ou_E\ m^{-3}$ (roman stile for the training set, *italics* for the odor test set). The star symbols represent the complaints of the citizens which live in close proximity of the sampling point. (2d) SOM map, the filling of the hexagons represents the basic statistics (lower outliers = white, quantiles = gray scale, upper outliers = black) of the average value of QEs. The arrows represent the trajectory of the day (24/07/2015) which was the main responsible of high QEs in some units (day start = star; day end = triangle).

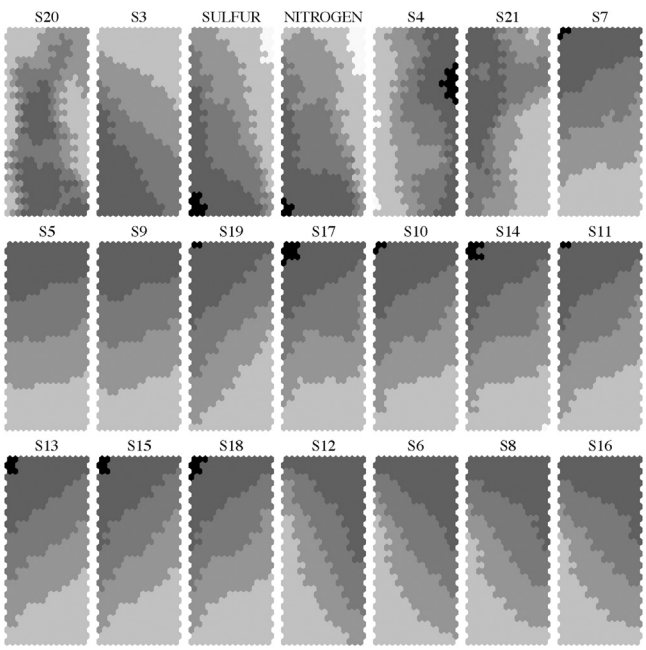

Fig. 3. SOM heatmaps ordered according to the hierarchical clustering of the experimental variables, the filling of the hexagons represents the basic statistics (lower outliers = white, quantiles = gray scale, upper outliers = black) of each sensor.

a dimension proportional to these average QEs. 32 units showed average QE values higher than the 3rd quantile value. Outlier detection can be used to discard outlying e-nose data and build a more robust model, however, in the application of the present study it is also useful to detect important variations due, for example, to a sensor fault or to air dynamics changes following a variation in the industrial process of the monitored factory. For example, a detailed examination of experimental vectors represented by the units with higher average QEs helped in the identification of a few experimental vectors that were not sufficient in number to characterize those map units but that were sufficient to enhance their average QE. We found that one day in particular was responsible for the high QE average values for 12 units.

The trajectory of that day (24/07/2015) across the map is represented by arrows in Fig. 2b. A trajectory shows the BMUs assigned to the experimental vectors (minutes) of the day, plotting a sequence of arrows which follows the time series. In that day the signals of the sensors named "SULFUR" and "NITROGEN" were found to have very low values respect to those expected from the BMU assignment, suggesting (i) that the sensors with low values did not work well or (ii) that in that day there was an "air typology" very different from the others which have been classified by the SOM. The sensor patterns in clusters for the experimental data of the day 24/7/2015 – overlaid on the cluster patterns of the SOM algorithm output depicted by box-plots – are reported in the Supplementary material (point 4). For an example of a daily trajectory animated visualization see Supplementary material (point 7).

3.5. SOM regions characterization

A further step in model building has been the identification of regions of the map discriminating between different types of ambient air, in terms of malodorous or odor-free characteristics. We have identified on the map the units representing the air samplings related to dynamic olfactometry analyses providing evidence of high odor events, i.e., the units representing the experimental e-nose data vectors recorded contemporarily to those samplings. In Fig. 2c, the units which correspond with the air samplings are labelled with the obtained odor concentration in $ou_E\ m^{-3}$ (roman

Table 3
distribution among clusters of complaints coming from volunteers and addressed from citizens to the municipality.

Cluster	Volunteers	Citizens	Overall
1	0	2	**2**
2	1	1	**2**
3	4	3	**7**
4	1	5	**6**
5	15	27	**42**

style). The BMUs are spread with high density mainly in the lower part of the map and partially in the middle left part of the map, thus we have an indication that neurons defining these regions of the SOM are related to malodor episodes. Moreover, considering (i) that opposite sides of the map usually represent opposite behavior of the variables [45], (ii) that odor events are usually transient and (iii) that the unit in the upper left corner (black hexagon in Fig. 2b) accounted for the highest number of experimental vectors (hits), we assumed that the upper left region of the map represented odor-free air.

The odor test data set was rescaled accordingly to the training set normalization parameters (see par. 3.3) and projected onto the map. The QEs were calculated obtaining the ordinal statistics values reported in Table 2. The QEs dispersion of the odor test set, describing odor nuisance events, was comparable with the training set one.

As it was done for the training set and the corresponding set of air samplings, we located on the map the units representing the air samplings which were in the odor test set and labelled them (*italic style* Fig. 2c). The scattering was found to be congruent (respectful of the malodorous and odor free regions identification) with that of the training set. Moreover we located the units representing the experimental vectors recorded contemporarily with the list of available citizens' complaints, which are labelled by a star in Fig. 2c. The complaints were gathered in the lower part of the map, confirming the attribution of "malodorous air" to that map region. The total number of complaints registered was 59, 21 coming from volunteers (see par. 2.3) and 38 addressed from citizens to the municipality (see par. 2.5). In Fig. 2c, for the sake of the readability of the figure, stars were not plotted on a unit if it was already labelled with a number. A table with details is reported in the Supplementary material.

3.6. SOM clusterization

In order to group neurons representing sensor data patterns on the basis of their similarity, the k-mean algorithm requires the selection of the desired number of clusters: the Davis-Bouldin index [61] is function of the ratio of the within-cluster scatter to the between-cluster separation, and it has to be the low for good clusterization. The DB index was calculated for a range of clusters from 2 to 10, iterating the k-means algorithm for 200 epochs. The number of clusters which showed the lowest DB-index was five. The cluster separation of the map is reported in Fig. 2c.

The air samplings appeared to be grouped into two clusters separately: 3 and 5. The complaint distribution among clusters is reported in Table 3, showing how the complaints were mainly in relation to minutes mapped on cluster 5.

The variability of values of each sensor for each cluster can be represented by box-plots, as illustrated in Fig. 4. For the sake of the readability of the figure, the sensor values were normalized across the variables to have zero mean and unit standard deviation and the box-plot whiskers are omitted. The box-plots with whiskers, as well as the sensor patterns in clusters calculated using the experimental data, are reported in the Supporting material. The sequence of presentation derives from the hierarchical clustering of the variables,

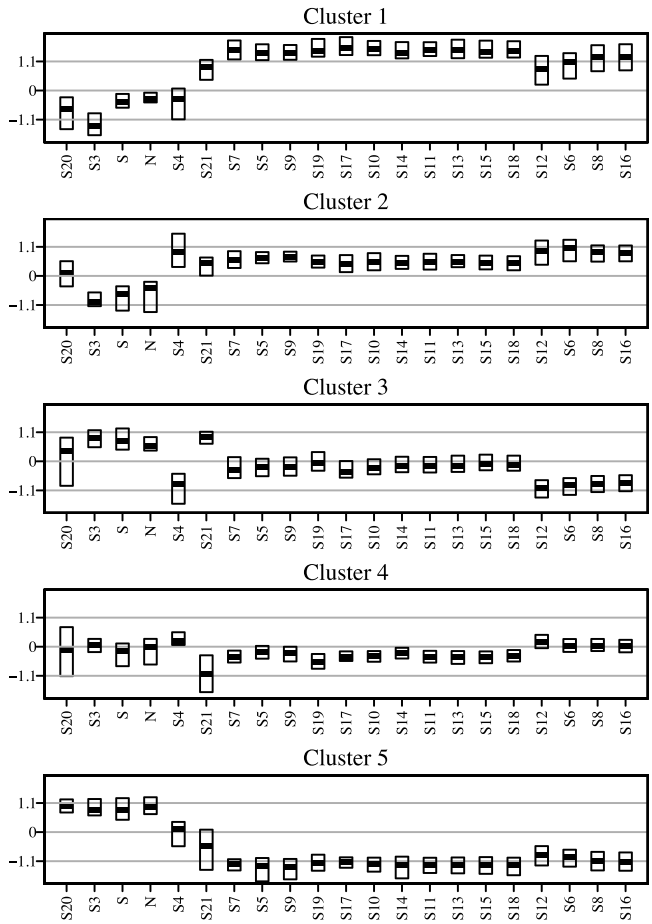

Fig. 4. Sensor patterns in clusters depicted by box-plots. S = SULFUR (sensor), N = NITROGEN (sensor). The sequence of presentation derives from the hierarchical clustering of the variables, the SOM data set has been normalized by variables for the sake of the readability of the figure.

introduced in chapter 3.5. It can be noticed that sensor patterns in cluster 1 and 5 show an opposite behavior.

Integrating the outcomes of the olfactometric measurements, the citizens' complaints records and the sensor patterns in clusters, we classified the clusters as: 1-odor free, 2-possibly odor free, 3- malodorous (source 1), 4-possibly malodorous, 5-malodorous (source 2).

With the aim of improving the characterization of the clusters we retrieved wind speed and direction data, as well as pollutant data recorded by the regional environmental protection agency in a monitoring station located nearby (see par. 2.1), representing them in box-plots according to the clusters, as shown in Fig. 5.

Considering that benzene and carbon monoxide are significantly present in fugitive emissions from coke distillation in steel plants [27,62], finding high values of these parameters in hours mapped on the Self Organizing Map regions characterized by both cluster 4 and 5 suggested that air identified by these grouping originated from the same source. Also the characteristic wind direction for the time being mapped on the aforementioned clusters confirmed a very similar origin for the two. On the basis of the aforesaid cases, clusters 4 and 5 were both characterized as malodorous (a lower frequency of citizens' complaints indicated that cluster 4 was probably less critical or malodorous than 5). Cluster 1 and 2 were characterized by high winds from ENE direction ("Bora" wind) and thus could both be classified as odor free for the SLS site, in relation to the industrial source. According to wind data (light wind from E direction, see median value of the box-plot) cluster 3 could

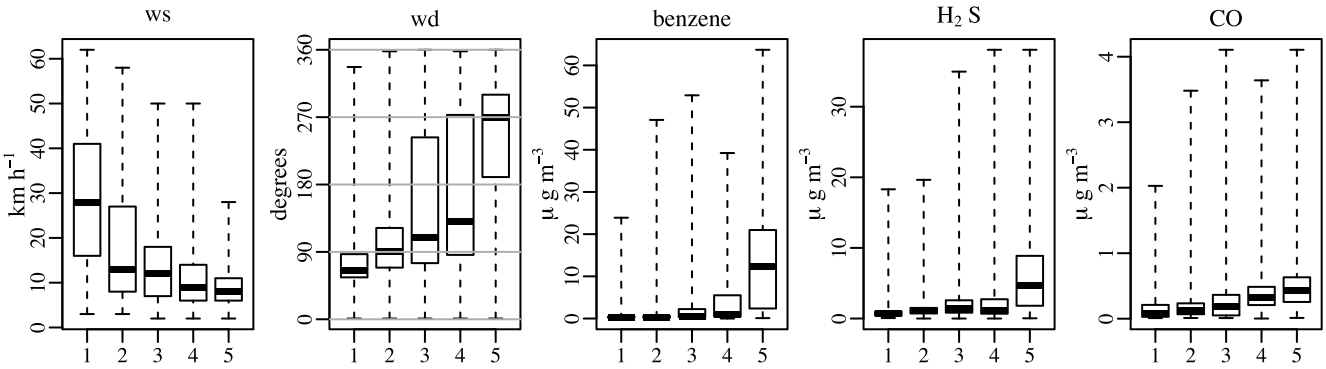

Fig. 5. Box-plots of wind speed and direction data and pollutant data, registered by the regional environmental protection agency at Molo Fratelli Bandiera site and at RFI site respectively, split by cluster.

Table 4
overall and monthly percentage distribution (frequency) of the experimental data among the clusters.

Cluster	June	July	August	September	Overall
1	2	1	1	13	**17**
2	1	5	8	4	**18**
3	0	3	9	4	**16**
4	12	8	3	1	**24**
5	4	11	6	2	**23**

Table 5
duration presented as the percentage of experimental data distribution among the clusters in selected time intervals.

Cluster	Duration (hours)						
	0–1	1–2	2–4	4–8	8–12	12–24	24–50
1	70	11	5	7	2	4	1
2	80	6	6	5	2	1	0
3	85	6	5	3	1	0	0
4	84	6	6	2	1	0	0
5	86	7	4	2	1	1	0

be classified as influenced by a malodorous source different to the steel plant, possibly a water treatment plant nearby, that it is subject to significant on-going renewal. The H_2S data supported the assignment of the malodorous classification to Cluster 5. Cluster 3 showed a median of H_2S slightly higher than Cluster 4.

As an example, citizens reported of an entire malodorous day (7/8/2015). In that day all the sensor vectors lied in cluster 3 (81%) and cluster 5 (19%). The trajectory of the day on the SOM map and the time plot of the sensor behaviors are reported in the Supplementary Material.

Following on from the cluster assignment, we evaluated the frequency and the duration of malodorous events (F and D in the FIDOL model for odor impact assessment). In Table 4 the frequency is presented as the percentage of the overall and monthly experimental data distribution among the clusters.

Clusters 3 and 5, for which the malodorous assignment is confirmed by the results obtained by olfactometric analysis, contained Best Matching Units for 47% of all the experimental e-nose vectors per minute in the four-month period. Interestingly, the ranking in frequency among the clusters changed throughout the months. Cluster 4 and 5 switched their ranking on moving from June to August. In August Cluster 3 had the first place and in September cluster 1 (odor free air) had the first place. The result is consistent with the fact that the sea breeze phenomena transporting coastal industrial emissions to the inland civil dwellings usually, fade away in late summer.

In Table 5 the duration (D) is presented as the percentage of experimental data distribution among the clusters in selected time ranges (0–1 h; 1–2 h; 2–4 h; 4–8 h; 8–12 h; 12–24 h; 24–50 h). It is expected that "odor-free air" events last usually longer than malodorous ones. In fact,the duration of cluster 1 and 2 were more spread among intervals longer than 1 h than cluster 3, 4 and 5.

4. Conclusions

In this paper we propose the first application of a method where we apply an unsupervised data analysis approach for handling e-nose monitoring data collected by a "general purpose" e-nose suitable to detect unspecified odorants. A refined model of odor impact can be achieved by tailoring a more specific e-nose configuration to assess that particular odor typologies identified with a first explorative application.

Ancillary information (olfactometric measurements on air samples after citizen calls, wind data, pollutant data and citizens' complaints records collected by the municipality) were used for map interpretation. In this application we trained the SOM using nearly the whole data, which contained a (small) number of e-nose signals registered in correspondence with olfactometric measurements (which carry "objective" odor information). Another small number of e-nose signals registered in correspondence with other olfactometric measurements was used to assess the recall ability of the map. This method proved that the SOM map was able to classify malodor events in a coherent way. The application can be improved by sampling air and evaluating its odor concentration by EN 13725:2003 in the absence of odor or in the presence of a weakly annoying odor to allow the identification of false positive results (e-nose data suggesting malodor, while in reality there was no odor) and to finely characterize map regions that in this application has been defined as "possibly odor-free" or "possibly malodorous". Another way to improve the model is to recruit volunteers who live in proximity of the sampling point and train them to fill (during e-nose sampling period) a daily questionnaire according to German VDI 3883 which standardizes a way to report odor Duration-Intensity-Frequency. The data collected by the questionnaire method can be used to refine the SOM interpretation.

Being aware that the abovementioned possible improvements have to be tested in different real environmental scenarios to obtain more robust and refined models, Odor Control Map proved to be a promising tool to be used as a diagnostic chart. It provides a valuable visualization support for following the dynamic evolution of the system with time: (i) relationships among sensor responses in different air types can be easily visualized and studied; (ii) dynamic changes of air types in time – e.g., during a day – can be followed by plotting a trajectory on the map and identifying the regions (clusters) involved; (iii) possible malodor sources can be identified;

(iv) emerging novel profile types, significantly different from those already classified, can be detected by means of outlier detection methods; (v) possible sensor failures can be detected by comparing cluster sensor profiles with profiles identified by means of outlier detection methods. Moreover, the air type classification allows the attainment, among the odor features (FIDOL), an experimental evaluation of the frequency and duration of air types classified as malodorous. The described approach does not need previous knowledge on odor sources or odor sources characterization to calibrate the e-nose, nevertheless air samples collected at the emission sources can be used for further training of the SOM algorithm and enrichment of the information provided by the map. Furthermore, considering that different annoyances (e.g., odor, noise, presence of specific chemical compounds) can cause possible synergistic health effects [1], Odor Control Map is a suitable tool to integrate data deriving from different and independent analysis/monitoring to obtain a more comprehensive knowledge on complex environmental phenomena involving civil dwellings close to industrial plants.

Conflict of interest

The authors declare no conflict of interest.

Funding sources

The present study was partly funded by the regional environmental protection agency (Agenzia Regionale per la Protezione dell'Ambiente del Friuli Venezia Giulia, ARPA-FVG).

Acknowledgments

The authors gratefully thank "No Smog Association Trieste" for the volunteers recruitment and for the information about the complaints rising from the citizens living in the district. The authors acknowledge Lab Service Analytica S.r.L. for providing the OdorPrep automatic sampling system. The authors gratefully thank Professor Neil Hickey of the University of Trieste for improving the language.

References

[1] T.H. Oiamo, I.N. Luginaah, J. Baxter, Cumulative effects of noise and odor annoyances on environmental and health related quality of life, Soc. Sci. Med. 146 (2015) 191–203, http://dx.doi.org/10.1016/j.socscimed.2015.10.043.
[2] S. Nimmermark, Odor influence on well-being and health with specific focus on animal production emissions, Ann. Agric. Environ. Med. 11 (2) (2004) 163–173.
[3] M.D. Shusterman, Critical review: the health significance of environmental odor pollution, Arch. Environ. Health: An Int. J. 47 (1992) 76–87, http://dx.doi.org/10.1080/00039896.1992.9935948.
[4] M. Brancher, K.D. Griffiths, D. Franco, H. de Melo Lisboa, A review of odor impact criteria in selected countries around the world, Chemosphere 168 (2017) 1531–1570, http://dx.doi.org/10.1016/j.chemosphere.2016.11.160.
[5] J.A. Nicell, Assessment and regulation of odor impacts, Atmos. Environ. 43 (2009) 196–206, http://dx.doi.org/10.1016/j.atmosenv.2008.09.033.
[6] T. Freeman, R. Cudmore, Review of Odor Management in New Zealand. Air Quality Technical Report No. 24, New Zealand Ministry of Environment, Wellington, New Zealand, 2002, Available at: www.mfe.govt.nz/sites/default/files/odor-tr-aug02.pdf.
[7] CEN. EN 13725:2003, Air Quality –Determination of Odor Concentration by Dynamic Olfactometry, Brussels, 2003.
[8] M. Brattoli, G. De Gennaro, V. De Pinto, A. Demarinis Loiotile, S. Lovascio, M. Penza, Odor detection methods: olfactometry and chemical sensors, Sensors 11 (2011) 5290–5322, http://dx.doi.org/10.3390/s110505290.
[9] M. Śliwińska, P. Wiśniewska, T. Dymerski, J. Namieśnik, W. Wardencki, Food analysis using artificial senses, J. Agric. Food Chem. 62 (2014) 1423–1448, http://dx.doi.org/10.1021/jf403215y.
[10] S. Deshmukh, R. Bandyopadhyay, N. Bhattacharyya, R.A. Pandey, A. Jana, Application of e-nose for industrial odors and gaseous emissions measurement and monitoring –an overview, Talanta 144 (2015) 329–340, http://dx.doi.org/10.1016/j.talanta.2015.06.050.
[11] S. Marco, The need for external validation in machine olfaction: emphasis on health-related applications, Anal. Bioanal. Chem. 406 (2014) 3941–3956, http://dx.doi.org/10.1007/s00216-014-7807-7.
[12] Nederlands Normalisatie-Instituut, Netherlands Technical Agreement: NTA 9055 Air Quality –Electronic Air Monitoring –Odor (nuisance) and Safety Appendix D, 2012.
[13] T. Kohonen, Self-organizing Maps, third ed., Springer, Berlin, 2001.
[14] A. Blanco-Rodriguez, F. Campo, O.M. Morales, R. Valiente, B. Lambert, L. Becher'n, A. Garcia-Ramirez, M.L. De, A. Dur'n, Development of an e-nose to identify and classify odours from spirits beverages, Chem. Eng. Trans. 54 (2016) 337–342, http://dx.doi.org/10.3303/CET1654057.
[15] U. Siripatrawan, B.R. Harte, Data visualization of Salmonella Typhimurium contamination in packaged fresh alfalfa sprouts using a Kohonen network, Talanta 136 (2015) 128–135, http://dx.doi.org/10.1016/j.talanta.2014.11.070.
[16] T. Liu, K. Chaibou, Z. Huang, A Novel retraining method of multiple self-organizing maps for gas sensor drift compensation, Sens. Mater. 25 (2013) 109–120.
[17] R. Bataller, I. Campos, M. Alcañiz, L. Gil-Sánchez, E. García-Breijo, R. Martínez-Máñez, L. Pascual, J. Soto, J.-L. Vivancos, A humid e-nose based on pulse voltammetry: a proof-of-concept design, Sens. Actuators B 186 (2013) 666–673, http://dx.doi.org/10.1016/j.snb.2013.06.033.
[18] A.M. Astel, L. Giorgini, A. Mistaro, I. Pellegrini, S. Cozzutto, P. Barbieri, Urban BTEX spatiotemporal exposure assessment by chemometric expertise, Water Air Soil Pollut. 224 (2013) 1503, http://dx.doi.org/10.1007/s11270-013-1503-7.
[19] S. Khedairia, M.T. Khadir, Impact of clustered meteorological parameters on air pollutants concentrations in the region of Annaba, Algeria, Atmos. Res. 113 (2012) 89–101, http://dx.doi.org/10.1016/j.atmosres.2012.05.002.
[20] M. Wu, Y. Wang, J. Dong, F. Sun, Y. Wang, Y. Hong, Spatial assessment of water quality using chemometrics in the Pearl River Estuary, China, Front. Earth Sci. 11 (2017) 114–126, http://dx.doi.org/10.1007/s11707-016-0585-0.
[21] S. Tsakovski, A. Astel, V. Simeonov, Assessment of the water quality of a river catchment by chemometric expertise, J. Chemom. 24 (2010) 694–702, http://dx.doi.org/10.1002/cem.1333.
[22] D. Rivera, M. Sandoval, A. Godoy, Exploring soil databases: a self-organizing map approach, Soil Use Manage. 31 (2015) 121–131, http://dx.doi.org/10.1111/sum.12169.
[23] M.P. Gómez-Carracedo, J.M. Andrade, G.V.S.M. Carrera, J. Aires-de-Sousa, A. Carlosena, D. Prada, Combining Kohonen neural networks and variable selection by classification trees to cluster road soil samples, Chemom. Intell. Lab. Syst. 102 (2010) 20–34, http://dx.doi.org/10.1016/j.chemolab.2010.03.002.
[24] D.M. Crnković, D.Z. Antanasijević, V.V. Pocajt, A.A. Perić-Grujić, D. Antonović, M.Đ. Ristić, Unsupervised classification and multi-criteria decision analysis as chemometric tools for the assessment of sediment quality: a case study of the Danube and Sava River, CATENA 144 (2016) 11–22, http://dx.doi.org/10.1016/j.catena.2016.04.025.
[25] H. Chen, J. Wang, J. Chen, H. Lin, C. Lin, Assessment of heavy metal contamination in the surface sediments: a reexamination into the offshore environment in China, Mar. Pollut. Bull. 113 (2016) 132–140, http://dx.doi.org/10.1016/j.marpolbul.2016.08.079.
[26] M. Boscolo, E. Padoano, Monitoring of particulate emissions to assess the outcomes of retrofitting measures at an ironmaking plant, ISIJ Int. 51 (2011) 1553–1560, http://dx.doi.org/10.2355/isijinternational.51.1553.
[27] E. Aries, D. Ciaparra, M.J. Schofield, D.R. Anderson, N. Schofield, R. Fisher, The Year Book of the Coke Oven Managers Association, Chapter: Fugitive and Stationary Source Emissions from Coke Plants and Impact on Local Ambient Air Quality, COMA, 2007, pp. 136–197 (Editors).
[28] G. Adami, P. Barbieri, S. Piselli, S. Predonzani, E. Reisenhofer, Detecting and characterizing sources of persistent organic pollutants (PAHs AND PCBs) in surface sediments of an industrialized area (Harbour of Trieste, Northern Adriatic), J. Environ. Monit. 2 (3) (2000) 261–265, http://dx.doi.org/10.1039/B000995O.
[29] P. Barbieri, G. Adami, S. Predonzani, E. Reisenhofer, D.L. Massart, Survey of environmental complex systems: pattern recognition of physicochemical data describing coastal water quality in the Gulf of Trieste, J. Environ. Monit. 1 (1999) 69–74, http://dx.doi.org/10.1039/A807528J.
[30] S. Licen, A. Tolloi, S. Briguglio, A. Piazzalunga, G. Adami, P. Barbieri, Small scale spatial gradients of outdoor and indoor benzene in proximity of an integrated steel plant, Sci. Total Environ. 553 (2016) 524–531, http://dx.doi.org/10.1016/j.scitotenv.2016.02.071.
[31] A. Astel, S. Cozzutto, F. Cozzi, G. Adami, P. Barbieri, S. Tsakovski, V. Simeonov, Seasonal apportionment of the sources of ambient air particulates in the city of Trieste, Int. J. Environ. Pollut. 41 (2010) 70–89, http://dx.doi.org/10.1504/IJEP.2010.032246.
[32] F. Cozzi, G. Adami, P. Barbieri, E. Reisenhofer, P. Apostoli, M. Bovenzi, Toxic elements content of PM 10 in a coastal area of the Northern Adriatic Sea, Cent. Eur. J. Chem. (2010) 2014–2026, http://dx.doi.org/10.2478/s11532-010-0074-3.
[33] S. Licen, A. Tolloi, G. Barbieri, A. Fabbris, S. Briguglio, P. Barbieri, Ambient PM10 concentration reconstruction in an inhabited area close to an industrial

hot spot by using particle density and optical particle counting values, JSM Environ. Sci. Ecol. 4 (1) (2016) 1026.
[34] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Trans. Neural. Networks 11 (3) (2000) 586–600, http://dx.doi.org/10.1109/72.846731.
[35] L. Capelli, S. Sironi, R. Del Rosso, P. Céntola, A. Rossi, C. Austeri, Odor impact assessment in urban areas: case study of the city of Terni, Proc. Environ. Sci. Urban Environ. Pollut. 2010 (4) (2011) 151–157, http://dx.doi.org/10.1016/j.proenv.2011.03.018.
[36] L. Capelli, S. Sironi, R. Del Rosso, P. Céntola, A. Rossi, C. Austeri, Olfactometric approach for the evaluation of citizens' exposure to industrial emissions in the city of Terni, Italy, Sci. Total Environ. 409 (2011) 595–603, http://dx.doi.org/10.1016/j.scitotenv.2010.10.054.
[37] S. Bootsma, T. Leuwerink, I. Bilsen, Online air monitoring with eNoses at the TATA steel plant in the Netherlands, Chem. Eng. Trans. 40 (2014) 79–84, http://dx.doi.org/10.3303/CET1440014.
[38] ARPA Valutazione dell'impatto dell'impianto siderurgico sulla qualità dell'aria nel comprensorio di Servola nel periodo gennaio 2015-31 marzo, 2016. Concentrazione di PM10, C6H6 e B(a)P, deposizione di polveri, As, Ni, Cd e Pb. Available at http://www.arpa.fvg.it/export/sites/default/ufficio_stampa/allegati/Puntuale_TS_servola_2015_31mar2016_dd16mag2016.pdf.
[39] J. Gebicki, Application of electrochemical sensors and sensor matrixes for measurement of odorous chemical compounds, TrAC Trends Anal. Chem. 77 (2016) 1–13, http://dx.doi.org/10.1016/j.trac.2015.10.005.
[40] VDI 3880: 2011/2017 Olfactometry – Static Sampling, The Association of German Engineers (VDI), 2018.
[41] M. Brattoli, G. Barbieri, P. Barbieri, S. Cozzutto, G. De Gennaro, A. Fabbris, R. Spaccavento, Development and technology assessment of the analytical performance of an eight position dynamic olfactometer, Chem. Eng. Trans. 40 (2014) 115–120, http://dx.doi.org/10.3303/CET1440020.
[42] OSMER-Meteorological Data of Friuli Venezia Giulia Region (Italy), 2018, Available at: http://www.osmer.fvg.it/archivio.php?p=stazioni&ln=&m=0.
[43] ARPA-Air Pollutant Data of Friuli Venezia Giulia Region (Italy), 2018, Available at: http://www.arpaweb.fvg.it/qagis/ariastor.asp.
[44] J. Vesanto, J. Himberg, E. Alhoniemi, J. Parhankagas, SOM Toolbox for Matlab 5, Report A57, 2000, Available at: www.cis.hut.fi/projects/somtoolbox/package/papers/techrep.pdf.
[45] J. Vesanto, SOM-based data visualization methods, Intell. Data Anal. 3 (1999) 111–126, http://dx.doi.org/10.1016/S1088-467X(99)00013-X.
[46] J. Himberg, J. Ahola, E. Alhoniemi, J. Vesanto, O. Simula, The Self-Organizing Map as a Tool in Knowledge Engineering, in: Pattern Recognition in Soft Computing Paradigm, Fuzzy Logic Systems Institute (FLSI) Soft Computing Series, World Scientific, 2001, pp. 38–65, http://dx.doi.org/10.1142/9789812811691_0002.
[47] A. Ultsch, H.P. Siemon, Kohonen's self organizing feature maps for exploratory data analysis, in: Proceedings of International Neural Network Conference (INNC'90), Kluwer academic Publishers, Dordrecht, 1990, pp. 305–308.
[48] A. Muñoz, J. Muruzábal, Self-organizing maps for outlier detection, Neurocomputing 18 (1998) 33–60, http://dx.doi.org/10.1016/S0925-2312(97)00068-4.
[49] SOM Toolbox-Copyright(C), 2000–2005, by Esa Alhoniemi, Johan Himberg, Juha Parhankangas and Juha Vesanto. Available at: http://www.cis.hut.fi/projects/somtoolbox/.
[50] R. Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, R Core Team, Vienna, Austria, 2016 https://www.R-project.org/.
[51] D.C. Carslaw, K. Ropkins, Openair – an R package for air quality data analysis, Environ. Model. Softw. 27–28 (2012) 52–61.
[52] H. Wickham, Ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2009.
[53] D. Kahle, H. Wickham, Ggmap: spatial Visualization with ggplot2, R J. 5 (1) (2018) 144–161 http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.
[54] Y. Xie, Animation: an r package for creating animations and demonstrating statistical methods, J. Stat. Softw. 53 (1) (2013) 1–27 http://www.jstatsoft.org/v53/i01/.
[55] T. Kohonen, Adaptive, associative, and self-organizing functions in neural computing, Appl. Opt. 26 (23) (1987) 4910–4918, http://dx.doi.org/10.1364/AO.26.004910.
[56] S. Khedairia, M.T. Khadir, L'analyse en Composantes Principales (ACP) et Carte Auto-organisatrice de Kohonen pour L'identification des types de jours météorologiques de la région d'Annaba, in: Proc of the 10th Maghrebian Conference on Information Technologies (MCSEAI'08), Oran, Alegria, 2008, pp. 18–23.
[57] S. Khedairia, M.T. Khadir, Self-organizing map and K-means for meteorological day type identification for the region of annaba –Algeria-, in: Proceedings of the IEEE Conference on Computer Information Systems and Industrial Management Applications (CISIM), Ostrava, The Czech Republic, 2008, pp. 91–96.
[58] S. Sinha, T.N. Singh, V.K. Singh, A.K. Verma, Epoch determination for neural network by self-organized map (SOM), Comput. Geosci. 14 (2010) 199–206, http://dx.doi.org/10.1007/s10596-009-9143-0.
[59] H. Fujita, A. Selamat, S. Omatu, New trends in intelligent software methodologies, tools and techniques, in: Proceedings of the 16th International Conference SoMeT_17, IOS Press BV, Netherlands, 2017, ISBN 978–1-61499–800-6 (on-line).
[60] J. Gutiérrez, M.C. Horrillo, Advances in artificial olfaction, Sens. App. Talanta 124 (2014) 95–105, http://dx.doi.org/10.1016/j.talanta.2014.02.016.
[61] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. 1 (2) (1979) 224–227, http://dx.doi.org/10.1109/TPAMI.1979.4766909.
[62] EMEP, B.1. B Fugitive Emissions from Solid Fuels – Solid Fueltransformation EMEP Guidebook, 2016, Available at: www.eea.europa.eu/publications/emep-eea-guidebook-2016/part-b-sectoral-guidance-chapters/1-energy/1-b-fugitives/1-b-1-b-fugitive/at_download/file.

Biographies

S. Licen obtained her M.Sc. in Chemistry from the University of Trieste (Italy) and her Ph.D. degree in Molecular Sciences from the same institution. Now she is a lecturer of Analytical Chemistry Laboratory at the Dept. of Chemical and Pharmaceutical Sciences, University of Trieste, Italy. Her current research interests are focused on the volatile and semi-volatile organic compounds outdoor and indoor assessment and on chemometric approaches for pattern recognition.

G. Barbieri obtained his M.Sc. in Territorial Management from the University of Trieste (Italy). He is co-founder of ARCO SolutionS (Environment Research Consulting and Sustainable Solutions ltd), spin off company of the Dept. of Chemical and Pharmaceutical Sciences of the University of Trieste, which core-business are odor concerning issues. He is member of the working group "Odors" at AIDIC (Italian Association of Chemical Engineering).

A. Fabbris is a civil engineer, she is part of the staff of ARCO SolutionS (Environment Research Consulting and Sustainable Solutions ltd), spin off company of the Dept. of Chemical and Pharmaceutical Sciences of the University of Trieste. She is the Panel leader of the olfactometric laboratory. She is member of the working group "Odors" at AIDIC (Italian Association of Chemical Engineering).

S. C. Briguglio obtained her M.Sc. in Chemistry from the University of Messina (Italy) and her Ph.D. degree in Chemistry from the University of Trieste (Italy). Her current research interests are focused on the volatile organic compounds in ambient air and in the human volatilome of exposed subjects.

A. Pillon obtained her M.Sc. in Physics from the University of Trieste (Italy). Since 2004 she is part of the staff of the Regional Environmental Protection Agency of Friuli-Venezia Giulia (Italy). Her current position is technical assistant. Her research interests include odor impact assessment and electronic nose applications.

F. Stel obtained his M.Sc. in Physics from the University of Trieste (Italy). Since 2000 he is part of the staff of the Regional Environmental Protection Agency, where he currently heads the Air Quality Sector. His research interests include air quality assessment and data mining methods while in the past his activity was devoted to the study of atmospheric hazards in connection with deep moist convection. He is member of the European Meteorological Society Awards Committee and chairs the Atmospheric Hazards Session in the European Meteorological Society annual meeting. He is lecturer of Atmospheric Physics at the Dept. of Physics of the University of Trieste, Italy.

P. Barbieri is associate professor at the Dept. of Chemical and Pharmaceutical Sciences, University of Trieste, Italy, where he coordinates the environmental chemistry research group. As a national expert, he is member of the working groups WG2 "Determination of odor concentration by dynamic olfactometry' and WG 41 "Electronic sensors for odorant monitoring' of the technical committee TC264 'Air Quality' at the European Committee of Normalization (CEN). He has been a co-founder of the spin off company ARCO SolutionS (Environment Research Consulting and Sustainable Solutions ltd), spin off company of the Dept. of Chemical and Pharmaceutical Sciences of the University of Trieste. He is author of more than 60 publications in analytical and environmental chemistry.