Assessment and Comparison of Multi-annual Size Profiles of Particulate Matter Monitored at an Urban-industrial Site by an Optical Particle Counter with a Chemometric Approach

Sabina Licen[1]*, Sergio Cozzutto[2], Pierluigi Barbieri[1]

[1] *Dept. of Chemical and Pharmaceutical Sciences, University of Trieste, 34127 Trieste, Italy*
[2] *ARCO SolutionS s.r.l., spin-off company of the Dept. of Chemical and Pharmaceutical Sciences, University of Trieste, 34127 Trieste, Italy*

ABSTRACT

The size of airborne particles is a key air quality parameter that is related to their composition, transport properties and effects on human health and the environment. Optical particle counters (OPCs) are increasingly used to dynamically characterize the size of ambient air particles. Monitoring campaigns lasting several months or even years generate millions of individual data values that must be effectively processed to extract information. Data mining algorithms as Self-Organizing Map (SOM) can support exploratory data analysis and pattern recognition in aerosol science. The use of SOMs, which offer powerful visualization features using 2D maps, allows us to interpret a large amount of data while avoiding any loss of information on variability from pre-treatments, such as compacting data recorded every minute to hourly or daily means. In the present study, we processed the data collected with an OPC during a long-term monitoring campaign (almost 3 years) conducted near residential buildings positioned very close to a steel plant and used them to assess and compare particulate matter (PM) profiles. About 12 million individual recorded values in total were handled. The current approach enabled us to identify four main PM profiles, follow their variation over time, and relate the differences to changes in the plant management and processes. Furthermore, it is potentially broadly applicable in high-frequency, long-term air quality monitoring campaigns employing different types of instruments to characterize the particle size and chemical composition of both PM and gases.

Keywords: Ambient air; Pollution; Self-Organizing Map; Pattern recognition.

INTRODUCTION

Air particulate matter (PM) is a key air quality parameter which can be related to composition, exposure and effects on humans (WHO, 2013; Chen *et al.*, 2016; Yang *et al.*, 2019) and it can be characterized with several different metrics. PM can be vehicle of pollutants such as heavy metals and/or persistent organic pollutants (POPs) that can be transferred to biota (Kodnik *et al.*, 2015; Fellet *et al.*, 2016; Popek *et al.*, 2017; Kłos *et al.*, 2018). It is renown that PM dimensions are a key parameter which is related to the ability of PM to retain and transport specific pollutants (Bernardoni *et al.*, 2017; Barbas *et al.*, 2018; Giungato *et al.*, 2018; Shen *et al.*, 2019).

Optical particle counters (OPCs) are increasingly used for ambient air characterization in the particle size low micrometer range. They are easy to use and can be run with little on-site maintenance for long-term monitoring. In OPCs the light is scattered by sampled PM providing information about particle number and dimensions.

Moreover, nowadays there is a growing interest in assessing the reliability of low-cost OPCs aiming to integrate the existing air quality regulatory networks (Bulot *et al.*, 2019; Feinberg *et al.*, 2019).

The most part of commercially available OPCs are able to record data per minute, or minute fraction. Thus a monitoring campaign lasting several months or years produces millions of single values to be elaborated for extracting information and knowledge to reach the specific aim of the study (city air quality assessment, hot spot industrial monitoring, comparison between different instruments, etc.).

In scientific literature there are several examples of multivariate analysis approaches for elaborating OPC data recorded in ambient air. Khan *et al.* (2015) present a study regarding minutely resolved data recorded in a 25-day monitoring campaign in a semi-urban tropical environment.

---

* Corresponding author.
 Tel.: +39 040-5583915; Fax: +39 040-5583905
 E-mail address: slicen@units.it

In Dominick *et al.* (2015) a study on a 3-month survey recording data per minute during a northeast monsoon in Malaysia is discussed. Agudelo-Castañeda *et al.* (2016) monitored PM size classes near a highway recording data every 15 min for 2 years. All the above mentioned studies perform principal component analysis (PCA) on hourly averages and show the resulting PC loadings by a table, no graphs showing PC scores are present. Etzion and Broady (2018) study concerns a 5-month monitoring in an urban neighborhood in which minutely resolved OPC data are elaborated by wavelet analysis. Ranalli *et al.* (2016) propose the use of functional exploratory data analysis of data collected every 6 seconds for 45 days in an urban site, but with a rather long computing time (several hours on a general-purpose desktop computer).

To our knowledge studies involving ambient air OPC data elaboration by neural networks are not present in literature.

In a recent publication we proposed a chemometric approach based on Self-Organizing Map (SOM) algorithm to elaborate OPC data recorded during a 3-month survey to assess the difference of PM profiles perceived at two different residential sites positioned near an industrial site (Licen *et al.*, 2019). The Self-Organizing Map is an artificial neural network algorithm for unsupervised pattern recognition (Kohonen, 2001). SOMs have been applied to analyze ambient air data collected with monthly, weekly or irregular frequency (Astel *et al.*, 2013; Zhong *et al.*, 2017; Romanić *et al.*, 2018; de Oliveira *et al.*, 2019), handling few thousands of data points. A much challenging task is represented by data with higher frequency, as it is the case of "quasi-real-time"

air monitoring; first case studies have been proposed considering few months for odor nuisance characterization by electronic noses (Licen *et al.*, 2018a, b). In several studies, to simplify data handling, the data are reduced to hourly or daily means, discarding information on intra-hour variability. The use of SOM algorithm allows instead to elaborate a high number of data without omitting information on data variability associated with data pre-treatments. Respect to PCA also possible non-linear relationships between variables can be detected, the data noise is reduced by the intrinsic properties of SOM algorithm and, additionally, SOM provides relevant visualization features using 2D maps (Kohonen, 2001) whereas in the presence of a very high number of samples visualizing PCA scores in a graph with a clear appearance is rather complex. Moreover, the approach permits to relate the model outcomes with data recorded by institutional air quality monitoring stations, providing meteorological data and pollutant concentrations by normed methods.

In the present study we describe the OPC data elaboration for PM profile assessment and comparison of a long monitoring (nearly 3 years) carried out near residential buildings positioned very close to a steel plant. About 12 million recorded single values have been handled on the whole.

METHODS

Site

The optical particle counter was positioned near dwellings (Fig. 1, *A*) which are distant approximatively 200 m from an integral cycle steel plant in the city of Trieste (NE Italy). The
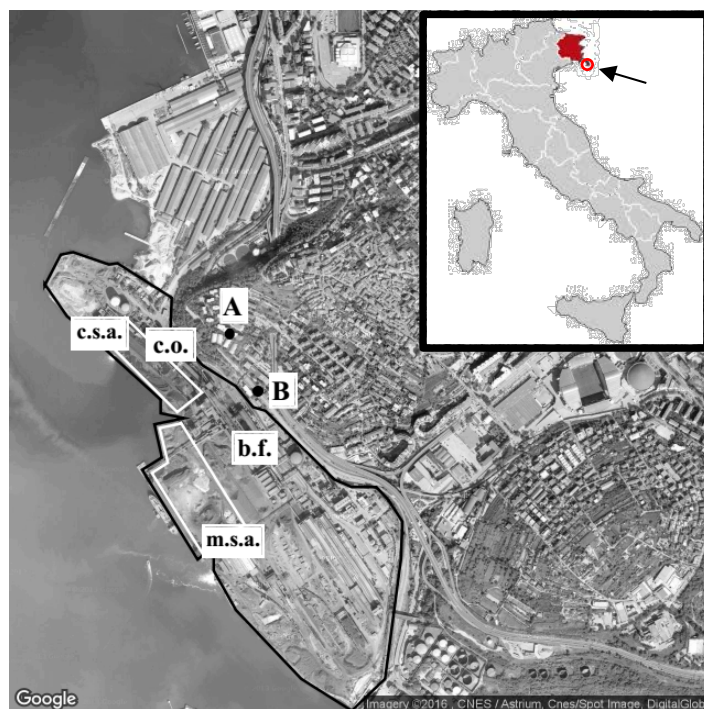
Fig. 1. Map of the area of interest. The location of Trieste in NE Italy is shown on the upper right corner of the figure. The sampling site (*A*), the regional environmental protection agency sampling site (*B*), the boundary of the steel plant (black line), the position of the coke oven batteries (*c.o.*), the position of the blast furnace (*b.f.*), the raw material storage areas (white line polygons; *c.s.a.* = carbon storage area, *m.s.a.* = mineral storage area) are highlighted.

main renown sources of particulate matter of the integrated cycle steel plant are the blast furnace, the raw material storage areas, and the coke oven batteries (Aries *et al.*, 2007; Boscolo and Padoano, 2011; Boscolo *et al.*, 2019). In the past years, studies were conducted on the site to assess impacts on civil dwellings by carcinogenic pollutants and odor emissions (Cozzi *et al.*, 2010; Licen *et al.*, 2016; Licen *et al.*, 2018c).

Optical Particle Counter Monitoring

The time- and size-resolved monitoring of particulate matter has been performed by an OPC with eight channels (Model 212 Eight Channel Particle Counter; Met One Instruments, Inc., Rowlett, Texas, USA). The instrument permits to discriminate eight size ranges ($< 0.3, > 0.3$–$0.5, > 0.5$–$0.7, > 0.7$–$1.0, > 1.0$–$2.0, > 2.0$–$3.0, > 3.0$–$5.0, > 5.0$–10.0 µm). Channels will be named from now on in the text respectively as *PM03*, *PM05*, *PM07*, *PM1*, *PM2*, *PM3*, *PM5* and *PM10*. The monitoring period lasted nearly 3 years starting from January 2014 and ending in October 2016. Data count per minute for each channel has been provided with air sampling at 1 L min^{-1}.

Wind Direction Data

Hourly wind direction data representative of the meteorological conditions of the city have been collected from a synoptic weather station (Molo Fratelli Bandiera (MFB)) in an open position by the sea 2 km from the plant, and they are available from the regional environmental protection agency (ARPA-FVG) website (OSMER). These data have not been used to build the models described in the study, thus we will refer to them in Sec. 3 as wind "external data."

PM$_{10}$ Hourly Data

The hourly PM$_{10}$ concentration data (µg m^{-3}; measured by normalized method EN 12341) collected at Site B (Fig. 1) positioned at the fence of the steel plant about 200 m from Sampling Site A, were obtained from the ARPA-FVG website (ARPA; ARPA-FVG, 2016). These data have not been used to build the models described in the study, thus we will refer to them in Sec. 3 as PM$_{10}$ "external data."

Self-Organizing Map

The SOM unsupervised algorithm works with no need of *a priori* data grouping or classification. A Self-Organizing Map is organized as a bi-dimensional array of neurons (or units), which are vectors of scalars related to the experimental variable data, that in the present study are the particle count responses of the channels. The array dimensions and starting values of the neurons are set according to heuristic rules proposed by Vesanto *et al.* (2000). During the training, the model "learns" from the data, adjusting the neuron values according to the Euclidean distance from the experimental vectors. In this way a new set of vectors (neurons) which still represent the variability of the processed data is obtained. The neuron vectors show simple geometric relationships (distances) and have the same number of variables as the experimental data. This algorithm is able to handle also non-linear relationships between variables and reduces data noise.

Details on the SOM algorithm and its general analytical features can be found elsewhere (Zupan *et al.*, 1997; Himberg *et al.*, 2001; Kohonen, 2001; Ballabio and Vasighi, 2012).

The neurons can be represented as hexagons stuck together in a bi-dimensional map allowing visual exploration of the data (Vesanto, 1999). In the present specific application, the neuron vectors represent recurrent PM profiles perceived at the monitoring site. According to Vesanto (1999) and Himberg *et al.* (2001) the Self-Organizing Map can be explored to observe the distribution of the values of every single modeled experimental variable on the map (*heatmap*), showing how each one of the original variables relates to the others in the SOM. Moreover, the number (*hits*) of experimental vectors which are represented by every single map unit can be evaluated.

The recurrent profiles can be further grouped using, e.g., a *k*-means clustering algorithm (Vesanto and Alhoniemi, 2000) on SOM neurons data set. The second-level clusterization can be represented on the SOM map by a color code to identify regions with relatively homogeneous properties. In this application the clusters represent air type profiles observed at the receptor.

Moreover, as shown in Licen *et al.* (2019) the results of the model and the cluster assignment can be coupled with other data collected on site by other instruments matching them by date/time, allowing cluster characterization.

Calculations

SOM building and *k*-means clustering were performed in the Matlab 6.5 (MathWorks, Inc.) computing environment, implementing the SOM toolbox (2000). Algorithm results exploration and 2D map visualization of the outputs were performed using in-house scripts in R software environment (R Core Team, 2016) implemented by the openair package (Carslaw and Ropkins, 2012). In addition, wind direction data and pollutant data were elaborated in R environment. The site map was prepared in R environment implemented by the ggplot2 package (Wickham, 2016) and ggmap package (Kahle and Wickham, 2013). Circlize package (Gu *et al.*, 2014) has been used to produce Fig. 4. The trajectory animation presented in the supplementary material was prepared in R environment implemented by the animation package (Xie, 2013).

RESULTS AND DISCUSSION

Monitoring Results Overview

In the present paragraph we illustrate an overview of results obtained during the 3 years' monitoring, using basic statistics applied to OPC channel counts. The results are shown in Table 1 where mean, median, max and min for each PM dimension split by year are reported.

The variation of PM channel counts for each year by hour of the day, day of the week and month, obtained using timeVariation function present in openair package are reported in the supplementary material.

Self-Organizing Maps Evaluation

An independent SOM model has been evaluated for each year. For all the three models the algorithm parameters were

Table 1. Basic statistics for each PM dimension split by year and expressed in counts min^{-1} L^{-1}.

		PM03	PM05	PM07	PM1	PM2	PM3	PM5	PM10
2014	min	$2.94\ 10^{+2}$	6	0	0	0	0	0	0
	median	$4.77\ 10^{+4}$	$2.70\ 10^{+3}$	$6.69\ 10^{+2}$	$3.56\ 10^{+2}$	$1.63\ 10^{+2}$	$3.20\ 10^{+1}$	7	0
	mean	$7.89\ 10^{+4}$	$7.47\ 10^{+3}$	$1.45\ 10^{+3}$	$6.53\ 10^{+2}$	$2.78\ 10^{+2}$	$5.50\ 10^{+1}$	$1.30\ 10^{+1}$	1
	max	$9.12\ 10^{+5}$	$1.82\ 10^{+5}$	$6.00\ 10^{+4}$	$4.62\ 10^{+4}$	$2.80\ 10^{+4}$	$8.84\ 10^{+3}$	$2.77\ 10^{+3}$	$3.18\ 10^{+2}$
2015	min	0	0	0	0	0	0	0	0
	median	$5.42\ 10^{+4}$	$2.89\ 10^{+3}$	$6.77\ 10^{+2}$	$3.32\ 10^{+2}$	$1.28\ 10^{+2}$	$1.80\ 10^{+1}$	3	0
	mean	$8.62\ 10^{+4}$	$6.89\ 10^{+3}$	$1.15\ 10^{+3}$	$5.28\ 10^{+2}$	$2.10\ 10^{+2}$	$3.20\ 10^{+1}$	6	0
	max	$8.36\ 10^{+5}$	$3.37\ 10^{+5}$	$1.40\ 10^{+5}$	$1.05\ 10^{+5}$	$5.70\ 10^{+4}$	$1.01\ 10^{+4}$	$2.38\ 10^{+3}$	$1.74\ 10^{+2}$
2016	min	$5.04\ 10^{+2}$	$4.30\ 10^{+1}$	8	2	0	0	0	0
	median	$2.62\ 10^{+4}$	$1.48\ 10^{+3}$	$4.05\ 10^{+2}$	$2.26\ 10^{+2}$	$7.70\ 10^{+1}$	$1.40\ 10^{+1}$	3	0
	mean	$5.17\ 10^{+4}$	$4.07\ 10^{+3}$	$7.55\ 10^{+2}$	$3.79\ 10^{+2}$	$1.29\ 10^{+2}$	$2.60\ 10^{+1}$	7	1
	max	$9.56\ 10^{+5}$	$1.44\ 10^{+5}$	$4.46\ 10^{+4}$	$2.05\ 10^{+4}$	$9.17\ 10^{+3}$	$4.29\ 10^{+3}$	$2.04\ 10^{+3}$	$3.59\ 10^{+2}$

set accordingly to Vesanto *et al.* (2000). The number of neurons was initially determined as 5 times the square root of the number of experimental vectors and divided by 4. The square root of the ratio between the two largest eigenvalues of the data set was evaluated and it was used to set the ratio between the map side lengths. Finally, the actual side lengths of the map were set to have their product as close as possible to the initially evaluated number of map units. We obtained SOM maps with a 39 × 21, a 37 × 22 and a 37 × 21 lattice for year 2014, 2015 and 2016 respectively.

In the left part of Fig. 3 the distribution of the modeled experimental variables on the map (*heatmaps*) are shown for each year. Four gray hues characterize quartiles of a single variable. Black color represents values higher than the 95th percentile.

It can be seen that for all years the upper part of the heatmaps shows lower number of counts for all the variables. For year 2014 higher number of counts for small PM dimensions are present in the lower left corner of the map while higher counts for large PM dimensions are present in the lower right corner. Heatmaps for year 2016 show a similar behavior while heatmaps for year 2015 seem to be a "mirror" with respect to the other years (Fig. 2).

SOM Maps Clusterization

SOM algorithm allowed to reduce the starting data sets to approximatively 800 recurrent profiles (units) for each year. In order to group the units representing akin PM size profiles, a *k*-means clustering algorithm (Vesanto and Alhoniemi, 2000) was operated on the matrix composed by the vectors characterizing the SOM units. The algorithm was iterated for 200 epochs for a range of clusters from 2 to 10. In order to select the best clusterization we used the Davies-Bouldin index (Davies and Bouldin, 1979) which is function of the ratio of the within-cluster scatter to the between-cluster separation. The number of clusters for which the corresponding DB index was lower was 4 for each SOM model. The grouping of the units by cluster is reported in the right part of Fig. 2. It can be seen that the larger part of the maps is enclosed in Cluster 1, while Clusters 2–4 are located in the lower part of the maps.

Cluster Characterization with External Data

Aiming to relate the clusters to wind direction and PM_{10}

concentration recorded by the regional EPA (see Sec. 2.3 and 2.4), we built a series of box plots grouping the external data variables by clusters. Every experimental vector recorded by OPC was assigned to the more similar unit, in terms of Euclidean distance, which is called "best matching unit" (BMU). As every unit is assigned to a specific cluster, the cluster value is in this way associated to all the experimental vectors assigned to that specific unit. Recording "date-time" was the only common variable between the OPC data and the external data, thus we were able to assign a cluster value to the "external data" recorded at the same "date-time" to OPC data. At the end of this process the "external-data" were grouped by cluster and represented as box plots, which are shown in Fig. 3 for PM_{10} and in Fig. 4 for wind direction. In Fig. 4 the interquartile ranges of box plots are laid on a wind rose. Each wind rose represents the extent of the interquartile range separately by year and corresponding to the same cluster.

In Fig. 3 it can be noticed that Cluster 1 is neatly related to low values of PM_{10}, whereas the other clusters are related to higher values of PM_{10}. In particular, Cluster 3 is related with the highest median values in year 2014 and 2016, while in year 2015 an inversion in the behavior of Cluster 3 and 4 can be observed.

Considering the interquartile ranges of wind direction reported in Fig. 4, a similar extent can be observed along the years for each cluster. Moreover, comparing the wind direction with the relative positioning of the monitoring site and the plant reported in Fig. 1, it can be noticed that Cluster 1 is related with wind blowing from the city, conversely the other clusters are associated with wind blowing from the plant.

It is possible to explore and compare the cluster particulate matter profiles representing the variables, as modeled in the SOM units, by box plots. Each box plot is generated gathering together all the values of a variable assumed by the units classified in the specific cluster. The PM profiles are shown in Fig. 5, the values are normalized by variable and the box plot whiskers are omitted for the sake of the readability of the figure. In the supplementary material it is reported another possible representation of Fig. 5 in which a box plot of each modeled variable split by cluster is shown without normalization of the channel counts.

The Cluster 1 to 3 profiles show a similar behavior among
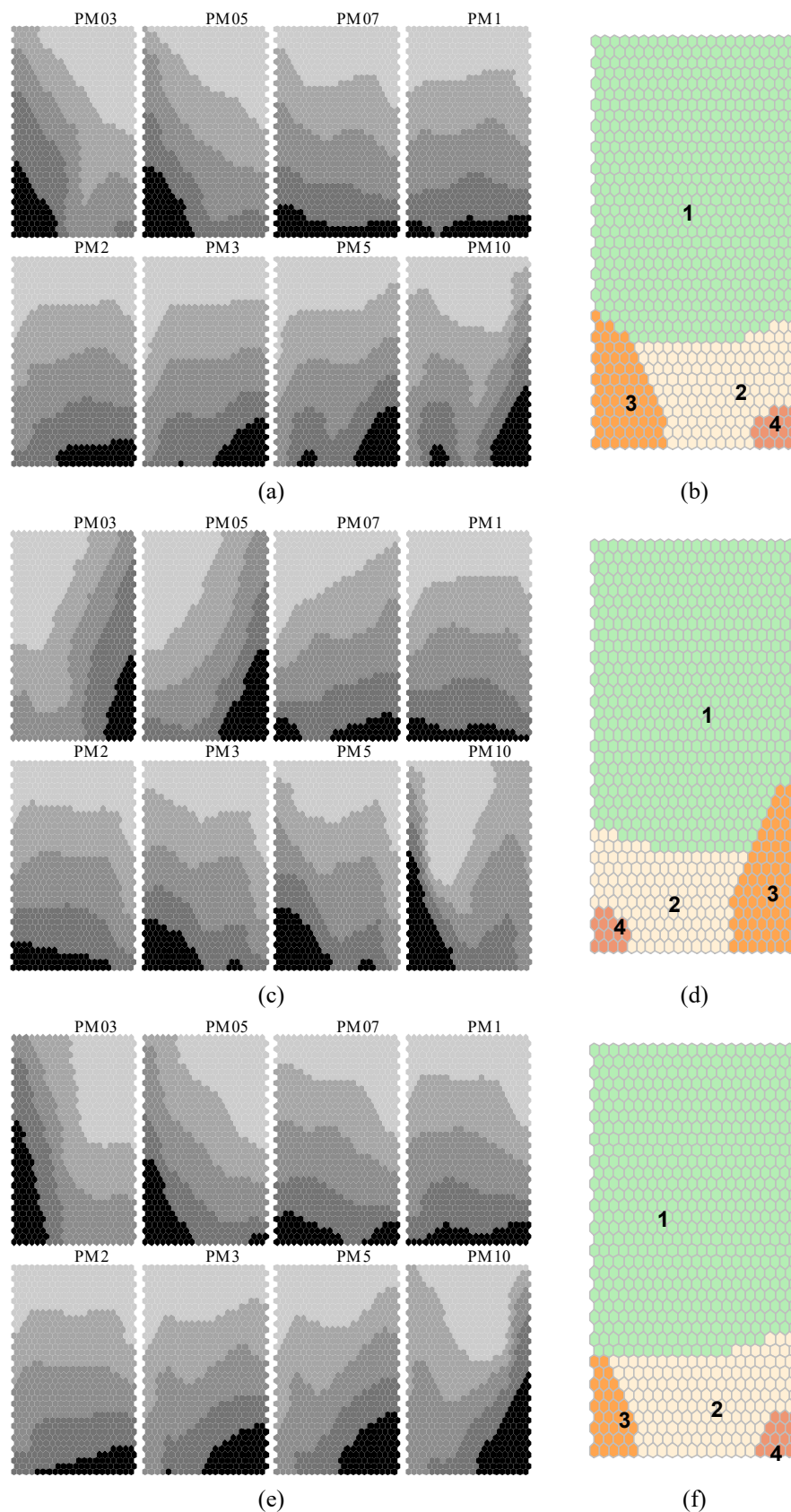
Fig. 2. Heatmaps for each SOM model: (a) year 2014, (c) year 2015, (e) year 2016. The filling of the hexagons represents the basic statistics (quartiles = gray scale, upper outliers > 95th percentile = black) of each modeled variable. SOM map for each model: (b) year 2014, (d) year 2015, (f) year 2016. 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.
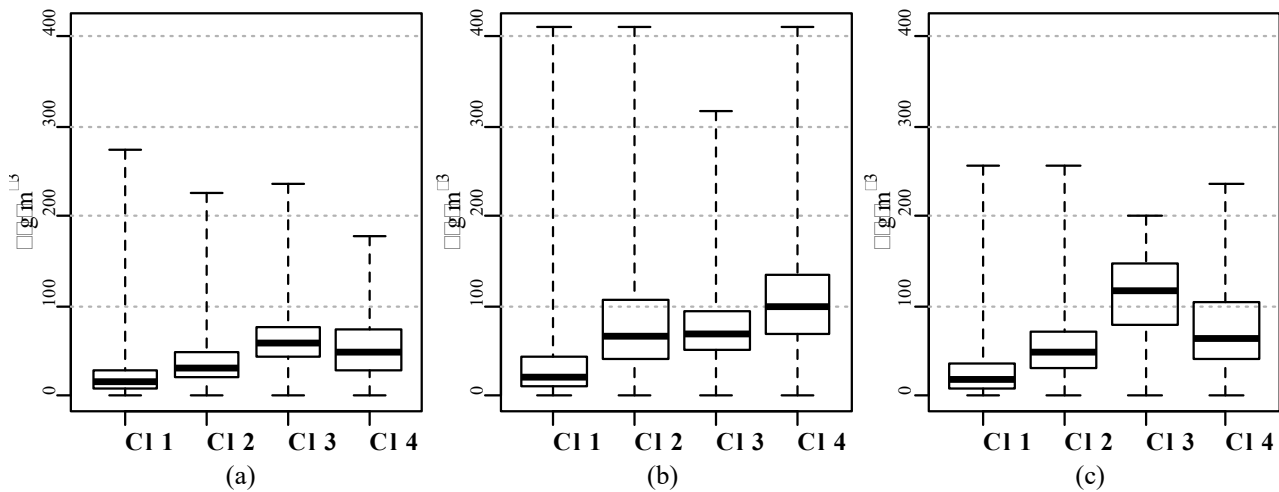
Fig. 3. Box plots of PM$_{10}$ data, recorded by the regional environmental protection agency at B site, split by cluster for each year: (a) year 2014, (b) year 2015, (c) year 2016. The box represents the interquartile range; the thick black line represents the median value; the dashed whiskers extend to the data extremes.
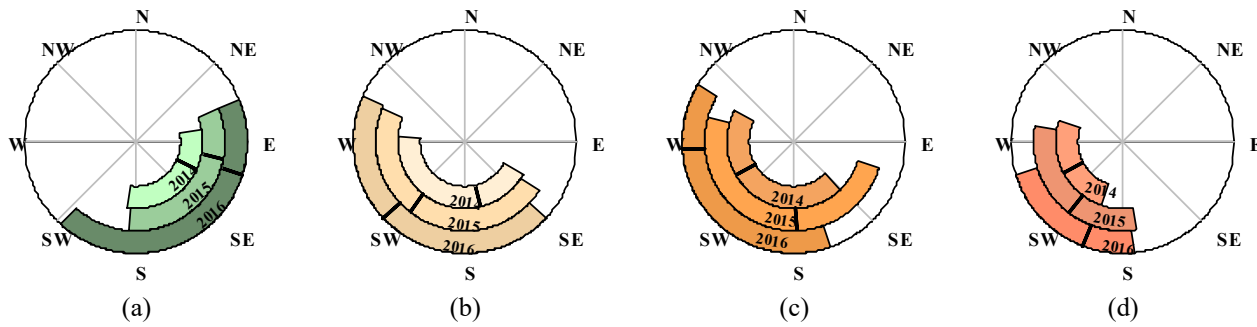
Fig. 4. Interquartile range of wind direction data, recorded by the regional environmental protection agency at MFB site, split by cluster and year: (a) Cluster 1, (b) Cluster 2, (c) Cluster 3, (d) Cluster 4. The thick black line represents the median value.
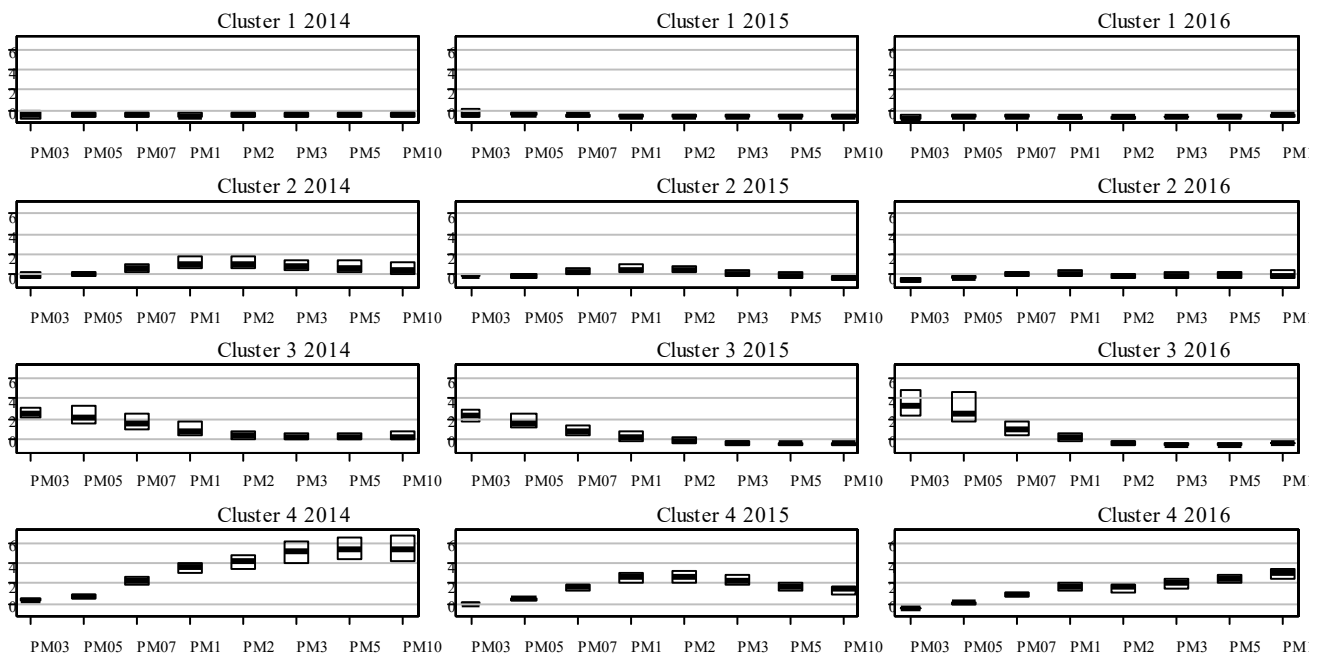
Fig. 5. Particulate matter profiles of units split by clusters for each year represented by box plots.

the 3 years, with Cluster 1 displaying low values of counts for all the PM sizes, Cluster 2 moderately higher values from PM07 to PM3 and Cluster 3 high values from PM03 to PM07.

Cluster 4 shows a similar behavior for years 2014 and 2016, with the latter displaying absolute lower values with respect to the former year. The behavior of Cluster 4 in year 2015 shows not very high values of PM3 to PM10 with respect to the trend of the other years. This evidence can be explained considering that in 2015 a change in the plant management and process occurred. In particular, a revamping of the blast furnace has been operated thus the casting activity of the plant has been stopped during some periods of the year to allow makeover operations. As the blast furnace is a renowned source of coarse PM fraction (Boscolo and Padoano, 2011) the effect of the abovementioned changes can be seen in year 2015 were the PM coarse fraction in Cluster 4 is lower and in year 2016, during which, after the setting up of the blast furnace activity, the profile of Cluster 4 shows a similar behavior as in year 2014 but with lower absolute values.

Cluster Frequency Evaluation

Every experimental vector is associated to a best matching unit as described in Sec. 3.3, thus, considering the units grouped into the same cluster, we can count the experimental vectors associated to that cluster and eventually evaluate the recurrence of that cluster during the monitoring period in terms of percentage frequency. In Table 2 the frequency for each cluster for each year is shown.

Clusters show similar frequencies among the year, except

for Cluster 3. Going into detail and exploring the frequency distribution *per* cluster for every month of the year (see tables reported in Appendix A) it can be noticed that Cluster 3 shows higher frequency in autumn and winter for years 2014 and 2015. Thus, considering that in 2016 the monitoring ended in October, it is possible that difference in frequency with respect to the previous years is due to the lower number of experimental data recorded in the autumn season.

PM Profiles' Temporal Evolution

The temporal evolution of the PM cluster profiles can be represented by stacked bar graphs in which each day is represented by a stacked bar built according to the percentage of each cluster assigned to the experimental data. An example regarding the period January–March 2014 is reported in Fig. 6.

The temporal evolution can be followed at a finer detail representing the trajectory (sequence of the BMU assigned to each experimental sample). In the supplementary material we present a graphical animation of 1-day trajectory as well as the minutely normalized PM profiles plotted in the same graph as the profile of the assigned cluster. In this way the cluster profiles can be used as a "reference" to interpret the single recorded experimental sample.

CONCLUSIONS

The approach we presented allowed us to extract variable profiles from millions of individual recorded data and to support model outcomes using data recorded with other

Table 2. Percentage frequency distribution among the clusters of the experimental data for each year.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
2014	82.4	8.8	7.2	1.5
2015	80.4	8.8	9.0	1.8
2016	85.0	10.4	2.9	1.7
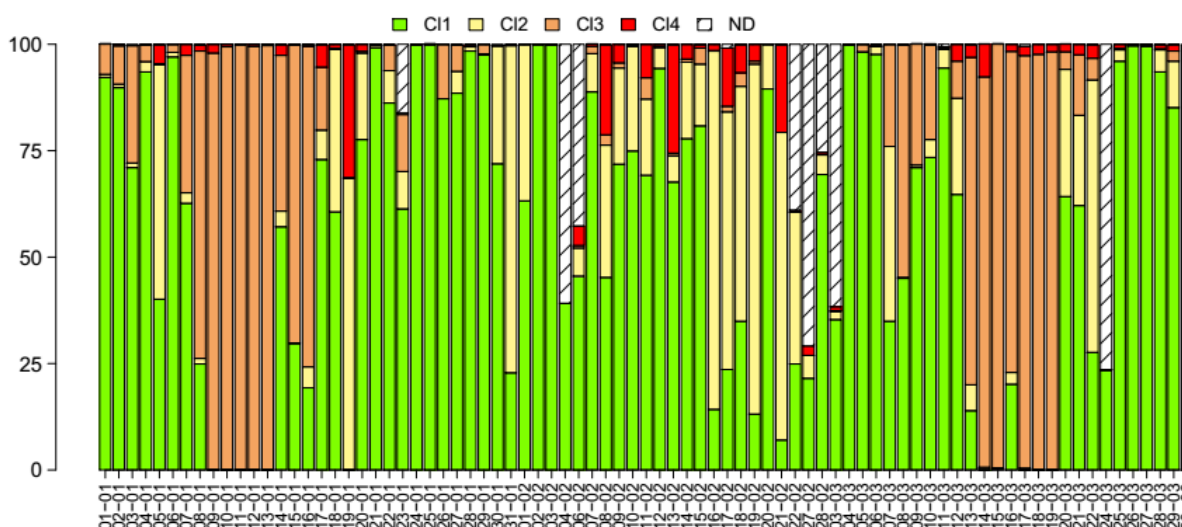
Fig. 6. Example of daily stacked bar graph (Jan–Mar 2014) built according to the percentage of each cluster assigned to the experimental data for each day (Y-axis = percentage, ND = not determined (experimental data not recorded)).

instruments. Three independent Self-Organizing Maps constructed from data collected during years 2014, 2015 and 2016 showed overall coherence, highlighting typical PM size patterns identified as clusters of recurrent profiles recorded annually at the monitoring site. The variability between the 3-year models was associated with i) the wind direction, which identifies the sources of PM affecting the site, and ii) independent PM_{10} data measured with a beta gauge PM monitor. Slight changes in the PM size profiles between the years were attributed to changes in the management of the nearby steel plant. Our method is potentially broadly applicable in high-frequency, long-term air quality monitoring campaigns employing different types of instruments to characterize the particle size and chemical composition of both PM and gases.

SUPPLEMENTARY MATERIAL

Supplementary data associated with this article can be found in the online version at http://www.aaqr.org.

REFERENCES

Agudelo-Castañeda, D.M., Teixeira, E.C., Schneider, I.L., Pereira, F.N., Oliveira, M.L.S., Taffarel, S.R., Sehn, J.L., Ramos, C.G. and Silva, L.F.O. (2016). Potential utilization for the evaluation of particulate and gaseous pollutants at an urban site near a major highway *Sci. Total Environ.* 543: 161–170. doi: 10.1016/j.scitotenv.2015.11.030.

Aries, E., Ciaparra, D., Schofield, M.J., Anderson, D.R., Schofield, N. and Fisher, R. (2007). The 2007 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, pp.136–197.

ARPA – Air pollutant data of Friuli Venezia Giulia Region (Italy), http://www.arpaweb.fvg.it/qagis/ariastor.asp.

ARPA-FVG (2016). Valutazione dell'impatto dell'impianto siderurgico sulla qualità dell'aria nel comprensorio di Servola nel periodo gennaio 2015 – 31 marzo 2016. Concentrazione di PM_{10}, C_6H_6 e B(a)P, deposizione di polveri, As, Ni, Cd e Pb. http://www.arpa.fvg.it/export/sites/default/ufficio_stampa/allegati/Puntuale_TS_servola_2015_31mar2016_dd16mag2016.pdf

Astel, A.M., Giorgini, L., Mistaro, A., Pellegrini, I., Cozzutto, S. and Barbieri, P. (2013). Urban BTEX spatiotemporal exposure assessment by chemometric expertise. *Water Air Soil Pollut.* 224: 1503. doi: 10.1007/s11270-013-1503-7.

Ballabio, D. and Vasighi, M. (2012). A MATLAB toolbox for Self-Organizing Maps and supervised neural network learning strategies. *Chemom. Intell. Lab. Syst.* 118: 24–32. doi: 10.1016/j.chemolab.2012.07.005.

Barbas, B., de la Torre, A., Sanz, P., Navarro, I., Artíñano, B. and Martínez, M.A. (2018). Gas/particle partitioning and particle size distribution of PCDD/Fs and PCBs in urban ambient air *Sci. Total Environ.* 624: 170–179. doi: 10.1016/j.scitotenv.2017.12.114.

Bernardoni, V., Elser, M., Valli, G., Valentini, S., Bigi, A., Fermo, P., Piazzalunga, A. and Vecchi, R. (2017). Size-segregated aerosol in a hot-spot pollution urban area: Chemical composition and three-way source apportionment. *Environ. Pollut.* 231: 601–611. doi: 10.1016/j.envpol.2017.08.040.

Boscolo, M. and Padoano, E. (2011). Monitoring of particulate emissions to assess the outcomes of retrofitting measures at an ironmaking plant. *ISIJ Int.* 51: 1553–15600. doi: 10.2355/isijinternational.51.1553.

Boscolo, M., Padoano E., Bregant, L., Dimastromatteo, V., Piller, M. and Bellini, M. (2019). A novel system for the abatement of fugitive emissions during charging in a by-product coke oven. *Metall. Res. Technol.* 116: 406. doi: 10.1051/metal/2019018.

Bulot, F.M.J., Johnston, S.J., Basford, P.J., Easton, N.H.C., Apetroaie-Cristea, M., Foster, G.L., Morris, A.K.R., Cox, S.J. and Loxham, M. (2019). Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. *Sci. Rep.* 9: 7497. doi: 10.1038/s41598-019-43716-3.

Carslaw, D.C. and Ropkins, K. (2012). *openair* - An R package for air quality data analysis. *Environ. Modell. Software* 27–28: 52–61. doi: 10.1016/j.envsoft.2011.09.008.

Chen, R., Hu, B., Liu, Y., Xu, J., Yang, G., Xu, D. and Chen, C. (2016). Beyond $PM_{2.5}$: The role of ultrafine particles on adverse health effects of air pollution. *Biochim. Biophys. Acta, Gen. Subj.* 1860: 2844–2855. doi: 10.1016/j.bbagen.2016.03.019.

Cozzi, F., Adami, G., Barbieri, P., Reisenhofer, E., Apostoli, P. and Bovenzi, M. (2010). Toxic elements content of PM_{10} in a coastal area of the Northern Adriatic Sea. *Cent. Eur. J. Chem.* 8: 1014–1026. doi: 10.2478/s11532-010-0074-3.

Davies, D.L. and Bouldin, D.W. (1979). A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* 1: 224–227. doi: 10.1109/TPAMI.1979.4766909.

de Oliveira, R.H., Carneiro, C.d.C., de Almeida, F.G.V., de Oliveira, B.M., Nunes, E.H.M. and dos Santos, A.S. (2019). Multivariate air pollution classification in urban areas using mobile sensors and self-organizing maps International. *Int. J. Environ. Sci. Technol.* 16: 5475–5488. doi: 10.1007/s13762-018-2060-9.

Dominick, D., Latif, M.T., Juneng, L., Khan, M.F., Amil, N., Mead, M.I., Nadzir, M.S.M., Moi, P.S., Samah, A.A., Ashfold, M.J., Sturges, W.T., Harris, N.R.P., Robinson, A.D. and Pyle, J.A. (2015). Characterisation of particle mass and number concentration on the east coast of the Malaysian Peninsula during the northeast monsoon. *Atmos. Environ.* 117: 187–199. doi: 10.1016/j.atmosenv.2015.07.018.

Etzion, Y. and Broday, D.M. (2018). Highly resolved spatiotemporal variability of fine particle number concentrations in an urban neighborhood. *J. Aerosol Sci.* 117: 118–126. doi: 10.1016/j.jaerosci.2018.01.004.

Feinberg, S.N., Williams, R., Hagler, G., Low, J., Smith, L., Brown, R., Garver, D., Davis, M., Morton, M., Schaefer, J. and Campbell, J. (2019). Examining spatiotemporal variability of urban particulate matter and application of high-time resolution data from a network of low-cost air pollution sensors *Atmos. Environ.* 213: 579–584. doi: 10.1016/j.atmosenv.2019.06.026.

Fellet, G., Pošćić, F., Licen, S., Marchiol, L., Musetti, R., Tolloi, A., Barbieri, P. and Zerbi, G. (2016). PAHs accumulation on leaves of six evergreen urban shrubs: A field experiment *Atmos. Pollut. Res.* 7: 915–924. doi: 10.1016/j.apr.2016.05.007

Giungato, P., Barbieri, P., Cozzutto, S. and Licen, S. (2018). Sustainable domestic burning of residual biomasses from the Friuli Venezia Giulia region. *J. Cleaner Prod.* 172: 3841–3850. doi: 10.1016/j.jclepro.2017.05.190.

Gu, Z., Gu, L., Eils, R., Schlesner, M. and Brors, B. (2014). *circlize* implements and enhances circular visualization in R. *Bioinformatics* 30: 2811–2812. doi: 10.1093/bioinform atics/btu393.

Himberg, J., Ahola, J., Alhoniemi, E., Vesanto, J. and Simula, O. (2001). The Self-Organizing Map as a Tool in Knowledge Engineering, In *Pattern Recognition in Soft Computing Paradigm*, Pal, N.R. (Ed.), Fuzzy Logic Systems Institute (FLSI) Soft Computing Series, World Scientific, pp. 38–65. doi: 10.1142/9789812811691_0002.

Kahle, D. and Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *R J.* 5: 144–161.

Khan, M.F., Latif, M.T., Amil, N., Juneng, L., Mohamad, N., Nadzir, M.S.M. and Hoque, H.M.S. (2015). Characterization and source apportionment of particle number concentration at a semi-urban tropical environment *Environ. Sci. Pollut. Res.* 22: 13111–13126. doi: 10.1007/s11356-015-4541-4.

Kłos, A., Ziembik, Z., Rajfur, M., Dołhańczuk-Śródka, A., Bochenek, Z., Bjerke, J.W., Tømmervik, H., Zagajewski, B., Ziółkowski, D., Jerz, D., Zielińska, M., Krems, P., Godyń, P., Marciniak, M. and Świsłowski, P. (2018). Using moss and lichens in biomonitoring of heavy-metal contamination of forest areas in southern and north-eastern Poland. *Sci. Total Environ.* 627: 438–449. doi: 10.1016/j.scitotenv.2018.01.211.

Kodnik, D., Candotto Carniel, F., Licen, S., Tolloi, A., Barbieri, P. and Tretiach, M. (2015). Seasonal variations of PAHs content and distribution patterns in a mixed land use area: A case study in NE Italy with the transplanted lichen. *Pseudevernia furfuracea. Atmos. Environ.* 113: 255–263. doi: 10.1016/j.atmosenv.2015.04.067.

Kohonen, T. (2001). *Self-Organizing Maps*, third ed. Springer, Berlin.

Licen, S., Tolloi, A., Briguglio, S., Piazzalunga, A., Adami, G. and Barbieri, P. (2016). Small scale spatial gradients of outdoor and indoor benzene in proximity of an integrated steel plant. *Sci. Total Environ.* 553: 524–531. doi: 10.1016/j.scitotenv.2016.02.071.

Licen, S., Barbieri, G., Fabbris, A., Briguglio, S.C., Pillon, A., Stel, F. and Barbieri, P. (2018a). 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. *Sens. Actuators, B* 263: 476–485. doi: 10.1016/j.snb.2018.02.144.

Licen, S., Cozzutto, S., Angelucci, M. and Barbieri, P. (2018b). Self-organizing map algorithm as a tool for analysis, visualization and interpretation of electronic nose high dimensional raw data. *Chem. Eng. Trans.* 68: 313–318. doi: 10.3303/CET1868053.

Licen, S., Pettarin, N., Barbieri, G. and Barbieri, P. (2018c). On odour tolerability criteria from odorant instrumental monitoring. *Chem. Eng. Trans.* 68: 157–162. doi: 10.3303/CET1868027.

Licen, S., Cozzutto, S., Barbieri, G., Crosera, M., Adami, G. and Barbieri, P. (2019). Characterization of variability of air particulate matter size profiles recorded by optical particle counters near a complex emissive source by use of Self-Organizing Map algorithm. *Chemom. Intell. Lab. Syst.* 190: 48–54. doi: 10.1016/j.chemolab.2019.05.008.

OSMER - Meteorological data of Friuli Venezia Giulia Region (Italy). http://www.osmer.fvg.it/archivio.php?p=s tazioi&ln=&m=0.

Popek, R.; Łukowski, A.; Bates, C. and Oleksyn, J. (2017). Accumulation of particulate matter, heavy metals, and polycyclic aromatic hydrocarbons on the leaves of *Tilia cordata* Mill. in five Polish cities with different levels of air pollution. *Int. J. Phytorem.* 19: 1134–1141. doi: 10.1080/15226514.2017.1328394.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ranalli, M.G., Rocco, G., Jona Lasinio, G., Moroni, B., Castellini, S., Crocchianti, S. and Cappelletti, D. (2016). Functional exploratory data analysis for high-resolution measurements of urban particulate matter. *Biom. J.* 58: 1229–1247. doi: 10.1002/bimj.201400251.

Romanić, S.H., Vuković, G., Klinčić, D. and Antanasijević, D. (2018). Self-organizing maps for indications of airborne polychlorinated biphenyl (PCBs) and organochlorine pesticide (OCPs) dependence on spatial and meteorological parameters. *Sci. Total Environ.* 628–629: 198–205. doi: 10.1016/j.scitotenv.2018.02.012.

Shen, R., Wang, Y., Gao, W., Cong, X., Cheng, L. and Li, X. (2019). Size-segregated particulate matter bound polycyclic aromatic hydrocarbons (PAHs) over China: Size distribution, characteristics and health risk assessment. *Sci. Total Environ.* 685: 116–123. doi: 10.1016/j.scitotenv. 2019.05.436.

SOM Toolbox (2000). SOM Toolbox - Copyright (C) 2000-2005 by Esa Alhoniemi, Johan Himberg, Juha Parhankangas and Juha Vesanto. http://www.cis.hut.fi/projects/somtoo lbox/.

Vesanto, J. (1999). SOM-based data visualization methods. *Intell. Data Anal.* 3: 111–126. doi: 10.1016/S1088-467X (99)00013-X.

Vesanto, J. and Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Trans. Neural Networks* 11: 586–600. doi: 10.1109/72.846731

Vesanto, J., Himberg, J., Alhoniemi, E. and Parhankagas, J. (2000). SOM Toolbox for Matlab 5, Report A57. www.cis.hut.fi/projects/somtoolbox/package/papers/tech rep.pdf.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag, New York.

World Health Organization (WHO) (2013). Review of evidence on health aspects of air pollution – REVIHAAP Project: Technical Report. WHO Regional Office for

Europe Publication, Copenhagen http://www.euro.who.int/__data/assets/pdf_file/0004/193108/REVIHAAP-Final-technical-report.pdf.

Xie, Y. (2013). animation: An R package for creating animations and demonstrating statistical methods. *J. Stat. Software* 53: 1–27. doi: 10.18637/jss.v053.i01.

Yang, Y., Ruan, Z., Wang, X., Yang, Y., Mason, T.G., Lin, H. and Tian, L. (2019). Short-term and long-term exposures to fine particulate matter constituents and health: A systematic review and meta-analysis. *Environ. Pollut.* 247: 874–882. doi: 10.1016/j.envpol.2018.12.060.

Zhong, B., Wang, L., Liang, T. and Xing, B. (2017). Pollution level and inhalation exposure of ambient aerosol fluoride as affected by polymetallic rare earth mining and smelting in Baotou, north China. *Atmos. Environ.* 167: 40–48. doi: 10.1016/j.atmosenv.2017.08.014.

Zupan, J., Novič, M. and Ruisánchez, I. (1997). Kohonen and counterpropagation artificial neural networks in analytical chemistry. *Chemom. Intell. Lab. Syst.* 38: 1–23. doi: 10.1016/S0169-7439(97)00030-0.