Urban BTEX Spatiotemporal Exposure Assessment by Chemometric Expertise

Aleksander Maria Astel · Luigi Giorgini ·
Andrea Mistaro · Italo Pellegrini ·
Sergio Cozzutto · Pierluigi Barbieri

Abstract Normative regulations on benzene in fuels and urban management strategies are expected to improve air quality. The present study deals with the application of self-organizing maps (SOMs) in order to explore the spatiotemporal variations of benzene, toluene, ethylbenzene, and xylene levels in an urban atmosphere. Temperature, wind speed, and concentration values of these four volatile organic compounds were measured after passive sampling at 21 different sampling sites located in the city of Trieste (Italy) in the framework of a multi-year long-term monitoring program. SOM helps in defining pollution patterns and changes in the urban context, showing clear improvements for what concerns benzene, toluene, ethylbenzene, and xylene concentrations in air for the 2001–2008 timeframe.

A. M. Astel (✉)
Biology and Environmental Protection Institute,
Pomeranian University, 22a Arciszewskiego
Str., 76-200 Słupsk, Poland
e-mail: astel@apsl.edu.pl

A. M. Astel
e-mail: AliAst@poczta.fm

L. Giorgini · A. Mistaro · I. Pellegrini
Department and Laboratory of Trieste, A.R.P.A.-F.V.G.,
Via La Marmora 13, 34100 Trieste, Italy

S. Cozzutto · P. Barbieri
D.S.C.F., University of Trieste, Via Giorgieri, 1, 34127
Trieste, Italy

P. Barbieri
e-mail: barbierp@units.it

L. Giorgini · S. Cozzutto
ARCO Solutions srl, Via Giorgieri, 1, Trieste 34127, Italy

1 Introduction

Benzene, after being known as a water and soil pollutant in proximity to crude and refined hydrocarbon storage sites, has been recognized as a relevant environmental issue for urban atmospheres in Europe since the late years of the last century, leading to the issuance of the Second Daughter Directive of the Air Quality Framework 2000/69/EC recently integrated in the 2008/50/EC Directive. In Italy, the Ministry Decree 60/2002 acknowledged the Directive, setting a limit value of 10 μgm^{-3} to be considered until 31 December 2005 and progressive reduction of limits down to 5 μgm^{-3} to be considered from January 2010 onwards. Limits were confirmed by the Legislative Decree 155/2010 that actuates the Directive 2008/50/EC known as "Clean Air For Europe."

Volatile organic compounds usually monitored in urban centers include toluene, ethylbenzene, and xylenes (reported to have harmful effects on the central nervous system (ATSDR 2012)) and benzene, classified by International Agency for Research on Cancer as human carcinogen with pending updates of the assessment (Infante 2011; Cogliano et al. 2011).

For what concerns sources of such atmospheric contaminants, most of the benzene in urban air is

emitted as non-exhausted volatile organic compounds (VOCs) from mobile sources, while a smaller amount comes from evaporative losses. EU Petrol Vapors Recovery Directives 94/63/EC and 2009/126/EC regulate the storage and the refueling of vehicles with petrol, with stricter norms having entered into force from January 2012 to reduce emissions of VOCs by recovering the vapor; the Directive 98/70/EC sets the maximum amount of benzene in gasoline to less than 1 %, and this was confirmed by the Directive 2009/30/EC.

Before the enforcement of Directive 98/70/EC, atmospheric average levels of benzene, toluene, ethylbenzene, and xylene (BTEX) in cities were significantly higher than the current ones, negatively contributing to citizen health. For instance, Brocco et al. (1997) reported the yearly average concentration of benzene in Rome (Italy) of 40 and 47 $\mu g m^{-3}$ during 1992 and 1993, respectively. Despite general benzene decrease in air, Air Quality Guidelines for Europe from WHO (2000) states that no safe level of exposure can be recommended, since benzene is carcinogenic to humans.

Urban atmospheres are still critical due to the high vehicular traffic combined with high population densities; besides fuel specifications, city management plays a fundamental role in determining VOC pollution levels by defining traffic plans and limitations and positioning of gas stations. This kind of pollution depends also on meteorological variables (seasonal and daily variations), socio-economical variables (daily and hourly traffic intensity and fluidity), and geographical variables (main highways, parking places, land morphology, etc.). Different sources, as the presence of industrial plants (e.g., coke ovens) or fuel tank farms, can play a role, but they are not object of the present work.

Scientific literature reports about studies and monitoring campaigns on BTEX in air conducted using passive samplers, highlighting factors affecting spatial and temporal patterns in cities and towns. Pekey and Yilmaz (2011) rationalized BTEX concentrations in the industrial city of Kocaely, Turkey. Caselli et al. (2010) reported about two weekly campaigns in the city of Bari, Italy, while Sturaro et al. (2010) studied correlations between BTEX and phenols during 2007 at two sites in the city of Padua, Italy. Roukos et al. (2009, 2011) studied different meteorological situations that occurred in 2007 in Dunkerque (France) and related VOC concentrations to ozone and long-

range transport of pollutants. Bruno et al. (2008) did daily campaigns in the town of Canosa di Puglia, Italy, while Hoque et al. (2008) described the situation of Delhi (India) with respect to traffic, as it was done by Monteiro Martins et al. (2007) for Rio de Janeiro, Brasil. Khoder (2007) performed a short-time study at Cairo, Egypt. Simon et al. (2004) described spring and summer campaigns near Toulouse (France) in 1999 and 2001, before and after the enforcement of Directive 98/70/EC about benzene content in gasolines. More recent studies focused on the relationship between ambient air and indoor environment (Chen et al. 2011; Hun et al. 2011; Esplugues et al. 2010; Johnson et al. 2010; Zielinska et al. 2012. Some BTEX levels determined during the European monitoring campaigns mentioned above and described in other references are summarized in Table 1.

Despite the literature review, no studies appear to report about multi-year trends of BTEX. In 2001, in consideration of the co-presence of potentially critical sources of VOCs such as traffic, steelworks with coke oven, and harbor activities related to crude oil pumping into a pipeline close to the urban center, the Regional Environmental Protection Agency of Friuli Venezia Giulia (ARPA-FVG) started a diffuse monitoring program in the coastal city of Trieste (Italy), with a quasi-monthly frequency in 21 sites by means of passive samplers. The city counts almost 210,000 inhabitants, and it faces the northernmost part of the Adriatic Sea. The monitoring program in 2012 is still running.

An assessment of the air quality requires consistent and periodical monitoring of pollution indicators as well as other parameters such as meteorological data, which are useful for interpreting pollutant dynamics. This large amount of data can be interpreted using chemometrics and more specifically time series analysis (Astel et al. 2008), multiway principal component analysis (Astel et al. 2010a), and receptor-oriented models (Astel et al. 2010b; Astel 2010). However, most of the techniques have their own requirements as data continuity for time series analysis. To solve data discontinuity problem, as often happens with long-term monitoring projects, the use of unique procedures becomes necessary. One of them could be a self-consistent iterative procedure called expectation maximization (Stanimirova et al. 2007, the other an application of chemometric algorithms which are relatively resistant for missing elements, i.e., the self-organizing maps.

Table 1 BTEX concentrations determined during exemplary monitoring campaigns in European countries

Country	Benzene (µg m^{-3})	Toluene (µg m^{-3})	Ethylbenzene (µg m^{-3})	(m,p)-Xylenes (µg m^{-3})	o-Xylenes (µg m^{-3})	Reference
Turkey	2.26±3.20	35.51±39.55	9.72±9.20	36.87±36.59	12.46±12.46	Pekey and Yilmaz (2011)
	11.6±3.2	26.7±2.9	4.9±2.5	21.1±5.4	21.9±4.1	Muezzinoglu et al. (2001)
Italy	11.1	15.6	3.7	10.3	10.3	Kim et al. (1997)
	0.8–9.0	0.9–15.5	0.2–2.7	0.8-10.0	0.3–3.9	Caselli et al. (2010)
	0.2–10.12	1.1–26.85	E+X, 1.08–34.56	E+X, 1.08–34.56	E+X, 1.08–34.56	Sturaro et al. (2010)
	1.5–9.3	4.4–32.5	1.0–8.5	1.2–22.4	1.0–6.7	Bruno et al. (2008)
France	W, 0.86±0.16	1.60±0.79	1.66±0.61	0.82±0.44	0.75±0.25	Roukos et al. (2009)
	S, 1.08±0.48	0.51±0.19	0.22±0.18	0.66±0.59	0.22±0.20	
	1.1	6.1		4.0	1.3	Simon et al. (2004)
Belgium	BR, 3.07	7.47	1.48	3.89	1.83	Keymeulen et al. (2001)
	PS, 1.78	4.94	1.55	3.29	1.16	
	RA, 0.85	1.81	0.37	0.82	0.46	
Hungary	BR, 2.30	8.22	0.87	2.68	1.32	
	PS, 2.52	6.20	0.74	2.71	1.62	
	RA, 1.27	3.77	0.35	1.62	0.10	
Latvia	BR, 7.80	32.9	5.44	15.1	7.81	
	PS, 7.87	33.4	5.61	15.8	6.96	
	RA, 1.05	4.17	0.64	1.88	0.88	
Greece	5.0	14.3	2.8	15.8	15.8	Kim et al. (1997)
Finland	1.66	5.62	0.99	3.12	1.26	Edwards et al. (2001)
Spain	0.1–27.9	6.4–134.7	1.0–22.1	2.2–63.9	0.9–23.1	Ras-Mallorquí et al. (2007)

E+X ethylbenzene+xylene, *W* winter, *S* summer, *BR* busy road, *PS* petrol station, *RA* remote area

The aim of the study was to model a discontinuous long-term monitoring data set on BTEX level using self-organizing maps to explore spatiotemporal variations of these pollutants after the lowering of benzene levels in gasoline and the removal of fuel stations from city centers. Additional aims were to determine whether the levels of BTEXs depend on the geographical location of sampling stations (being a direct consequence of traffic intensity) and to investigate the relations between basic meteorological parameters (wind speed and temperature) and BTEX levels measured in the atmosphere of the city of Trieste.

2 Experimental

2.1 Sampling Sites and Sampling Procedure

Air samples were collected at 21 fixed stations spread across the city of Trieste, Italy, forming an irregular grid of approximately 5-km long from north to south and 1.5-km wide from east to west. This rectangular disposition reflects the geographical features of the city, which develops longitudinally between the Adriatic Sea coast (west) and the highlands (east). Fixed stations were located both in the city center and in its northern and southern peripherals, dividing the entire city area into several subareas according to their location (Fig. 1).

The urban area located in the west of the city center has not been included in the sampling protocol since it is characterized by inhomogeneous features (generally low traffic, hilly residential zone located immediately in the western direction with respect to the city center, fluid traffic highways located along the coast line, and harbor activities near the sea). The explanations of the district abbreviations and their urban classification presented in Fig. 1 are depicted in Table 2.

Pollution sources in the area of interest are mainly due to vehicular traffic and, during cold seasons, domestic
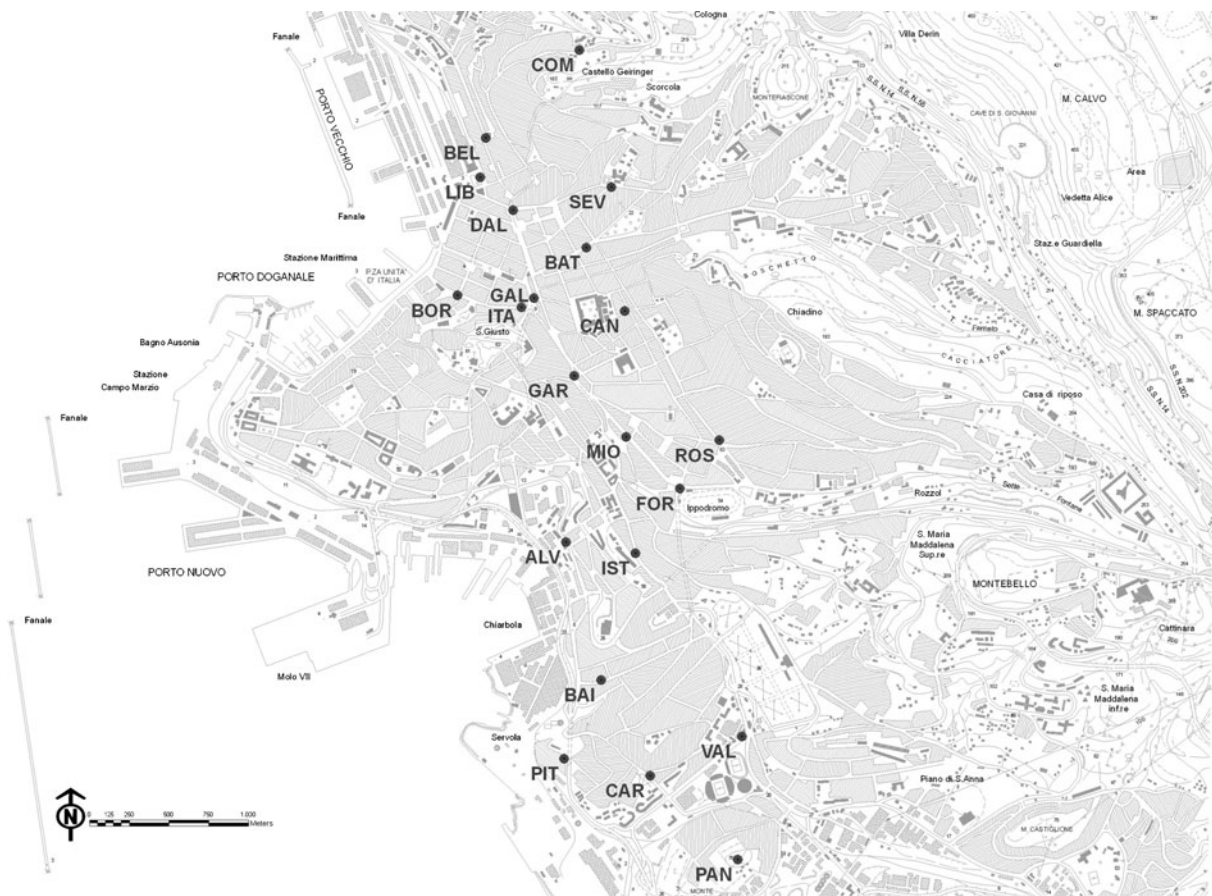
Fig. 1 Passive sample location map

heating. Moreover, a steelworks with coke oven is located on the sea shore near the monitoring station of Baiamonti Str., while the proper industrial zone develops in the southern direction (out of the map) with respect to the monitored area. According to vehicular traffic, the city center is mainly impacted by small private vehicles (cars and motorcycles) and public transportation (busses and taxis). During the last few years, the access to most part of the downtown and semi-peripheral areas has been progressively restricted to vehicles featuring more advanced anti-pollution devices, meeting more demanding standards of emission. The effect of this strategy has been a significant decrease of air pollution levels (e.g., benzene concentration decrease (Fig. 2)), in spite of the general trend of increasing number of circulating vehicles and their increasing use, due to the reduced prices of fuel established by local administrators according to the proximity of the Slovenian border, where fuel is cheaper than in Italy since less taxed, as a rule.

Local worsening of the air quality reflected by an increase of benzene concentration in 2006 could be attributed to extended road works in the street connecting Garibaldi Square and Corso Italia, and consequent difficulties for vehicle circulation in this area and the neighboring ones were due to alternative traffic paths.

Air samples were collected in the period between January 2001 and December 2008 with Radiello RAD 130 (Sigma-Aldrich, USA) for BTEX sampling, according to the procedure given by the supplier of Radiello (Radiello 2003; Cocheo et al. 2009). Samplers have been mounted on appropriate standings located at 2.5–3.0 m above the ground level. Samplers that have been exposed for 3 weeks, from the 1st to the 21th day of every month, were then stored in the laboratory at +4 °C. Two milliliters of carbon disulfide (free from benzene) was added to the Radiello glass tube. Tubes were mixed manually and left for 30 min;

Table 2 An explanation of the codes of the streets according to the sampling grid located in the city of Trieste

Downtown (city center)	Widespread downtown	Northern urban area	Southern urban area
BAT: Battisti Str.	ALV: d'Alviano Str.	BEL: Belvedere square	BAI: Baiamonti Str.
BOR: Borsa square	DAL: Dalmazia square	COM: Commerciale Str.	CAR: Carpineto Str.
CAN: Canova Str.	FOR: Foraggi square	LIB: Libertà square	PAN: Monte S. Pantaleone
GAL: Gallina Str.	IST: dell'Istria Str.		PIT: Pitacco Str.
GAR: Garibaldi Str.	MIO: Mioni Ave.		VAL: Valmaura Str.
ITA: Corso Italia	ROS: Rossetti Str.		
	SEV: F. Severo Str.		

then, they were mixed again and left for another 30 min. The solution of CS_2 containing the extracted hydrocarbons was transferred to a crimp-capped 2-mL glass vial mounted with a PTFE cap.

2.2 Determination of Analytes

Analysis was performed by means of a Clarus 500 gas chromatograph (Perkin-Elmer, USA) provided with an autosampler. The columns used were a SGE BPX35 (length 30 m, diameter 0.25 mm, film thickness 0.25 μm) or equivalent. One microliter of the sample was injected with split technique through a 235 °C traditional injector, while the oven temperature has been ramped from 35 to 265 °C with an 8 °C min^{-1} rate. A FID detector was used. Calibration has been performed by analyzing five standard solutions each containing 1, 2, 5, 10, and 20 mg L^{-1} of benzene, ethylbenzene, o-xylene, m-xylene, p-xylene, and 4, 8, 20, 40, and 80 mg L^{-1} of toluene. Method performances were controlled by injecting periodically (every 20 injections) a blank and two standard solutions. The method had been previously validated through an intercalibration test among different laboratories (Villalta et al. 2004). The detection limit was lower than 0.05 mg L^{-1}. Collected data were converted to air concentration (μg m^{-3}) according to the instructions given by the supplier of Radiello (Radiello 2003), considering time and average temperature of exposition.

2.3 Meteorological Data

Trieste climate is Mediterranean, characterized by episodes of a strong katabatic ENE wind known as Bora; during the coldest month (January), average temperature approaches 6 °C, while in the warmest month (July), it is slightly higher than 24 °C. Meteorological data are provided by the meteorological service of ARPA-FVG, managing a synoptical station and several micrometeorological stations. Average data for temperature (°C) and wind speed (m/s) registered by the seven automatic air quality and meteorological stations located inside the sampling area are reported in Table 3.

2.4 Data Analysis Procedure

Various statistical multivariate techniques are commonly used for air pollution modeling and assessment (Astel et al. 2008, 2010; Astel 2010a, b). In the present study, the self-organizing map (SOM) algorithm, as one of the most efficient neural network architectures for solving problems in the field of exploratory data analysis, clustering, and data visualization, was applied. The SOM approach was originally proposed by Kohonen (1984); it is a model that implements a characteristic nonlinear projection from the high-dimensional space of objects onto a low-dimensional array of neurons. Theoretical background of the SOM approach can be found elsewhere (Kohonen 1984, 1995; Kohonen et al. 1996; Vesanto 2000; Vesanto et al. 2000; Giraudel and Lek 2001; Fort 2006; Cottrell et al. 1998; Ultsch and Siemon 1990) and this is why only major advantages related to both exceptional visualization, dimension reduction, and exploration abilities are briefly presented below.

2.4.1 General Idea

The SOM algorithm, much like conventional ordination methods, shares the basic idea of displaying a high-dimensional signal manifold onto a much lower
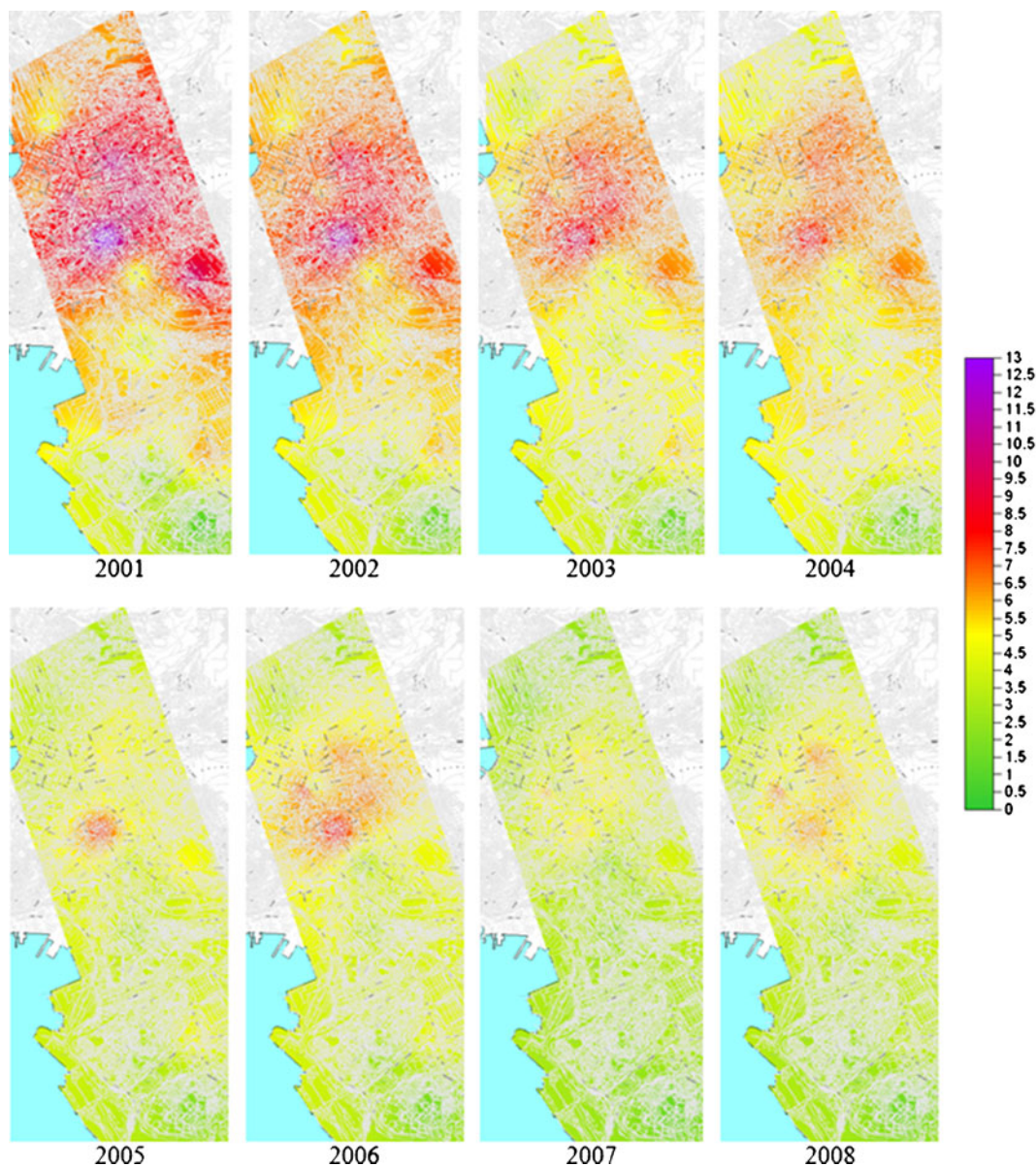
Fig. 2 An annual average benzene concentration in air ($\mu g m^{-3}$) in the city of Trieste in the period between 2001 and 2008 (maps acquired with Golden Surfer v.8.0; interpolation algorithm used inverse distance to a power (power=2)

dimensional network in an orderly fashion (usually 2D space). An SOM produces virtual communities in a low-dimensional lattice through an unsupervised learning process. The term "self-organizing" refers to the ability to learn and organize information without being given the associated dependent output values for the input pattern (Mukherjee 1997). An SOM is noise tolerant; this property is highly desirable when site-measured data are used. The advantages of the SOM in relation to conventional

ordination methods as principal component analysis and cluster analysis are discussed elsewhere (Astel et al. 2007; Zhang et al. 2008).

2.4.2 Kohonen's Neural Network Map Description

The SOM consists of neurons organized on a regular, usually two-dimensional grid. The neurons are connected to adjacent neurons by a neighborhood

Table 3 Basic statistics concerning BTEX concentrations ($\mu g\,m^{-3}$) and meteorological parameters according to sampling site

	Site	N	Mean	Min	Max	S.D.	VC	SC	Site	N	Mean	Min	Max	S.D.	VC	SC	Site	N	Mean	Min	Max	S.D.	VC	SC
Benzene	FOR	87	4.4	1.8	12.2	1.7	0.398	1.5	LIB	87	3.7	0.6	7.4	1.5	0.388	0.3	BAI	86	4.4	2.1	9.0	1.7	0.373	0.6
Toluene		87	21.5	5.9	203.8	22.9	1.066	6.3		87	15.3	2.3	47.6	8.3	0.539	1.4		86	18.6	5.3	52.8	9.0	0.487	1.1
Ethylbenzene		87	5.1	2.3	21.3	2.8	0.539	3.1		87	3.9	0.5	9.1	1.7	0.430	0.5		86	4.9	1.8	11.6	2.1	0.420	0.6
Xylene		87	25.2	6.1	441.5	46.4	1.838	8.6		87	15.4	3.1	42.1	6.9	0.444	1.0		86	19.6	7.1	44.7	8.2	0.420	0.5
Temp		87	15.3	3.3	29.8	7.1	0.467	0.0		88	15.3	3.3	29.8	7.2	0.469	0.0		86	15.4	3.3	29.8	7.1	0.461	0.0
Wind		87	1.9	0.4	17.1	1.8	0.969	7.2		88	1.9	0.4	17.1	1.8	0.970	7.2		86	1.8	0.0	17.1	1.8	0.985	7.0
Benzene	MIO	76	4.2	1.5	8.2	1.6	0.390	0.6	BEL	79	3.7	1.5	8.0	1.5	0.398	0.8	PIT	85	3.7	1.4	7.8	1.4	0.373	1.1
Toluene		76	20.2	5.4	164.5	19.2	0.950	5.9		79	15.2	3.6	41.2	7.5	0.494	0.9		84	7.5	1.8	42.1	5.7	0.757	3.4
Ethylbenzene		76	4.8	1.2	10.8	2.3	0.468	0.6		79	3.9	1.0	10.7	1.8	0.455	0.9		82	2.0	0.3	10.4	1.5	0.729	2.6
Xylene		76	19.2	4.7	49.0	9.4	0.489	0.8		79	15.8	3.2	44.7	7.3	0.463	1.0		85	7.4	1.1	52.6	6.9	0.928	4.2
Temp		76	15.3	3.3	29.8	7.2	0.473	0.0		79	15.1	3.3	29.8	7.1	0.471	0.1		85	15.5	3.3	29.8	7.0	0.450	0.0
Wind		76	1.8	0.4	17.1	1.9	1.035	6.9		79	1.9	0.4	17.1	1.9	0.972	7.0		85	1.9	0.4	17.1	1.8	0.989	7.0
Benzene	GAR	86	8.6	3.8	17.9	3.5	0.410	0.7	COM	77	3.8	1.3	8.2	1.6	0.436	0.6	CAR	70	2.9	1.3	5.9	1.1	0.366	0.8
Toluene		86	47.7	9.9	209.3	28.2	0.591	2.5		76	18.4	4.0	242.2	28.9	1.573	6.6		70	7.0	2.1	21.9	3.8	0.534	1.4
Ethylbenzene		86	12.4	5.0	28.0	5.7	0.461	1.0		77	3.9	1.1	10.3	2.1	0.541	0.9		67	2.0	0.6	5.7	1.2	0.592	1.2
Xylene		86	51.4	18.8	108.3	22.4	0.436	0.9		77	15.1	5.0	36.5	8.5	0.563	0.9		70	7.2	1.7	25.8	4.5	0.620	1.8
Temp		86	15.5	3.3	29.8	7.0	0.452	0.0		77	15.7	3.3	29.8	7.3	0.463	−0.1		70	15.5	3.7	27.1	7.0	0.449	0.0
Wind		86	2.0	0.4	17.1	2.3	1.142	5.6		77	1.9	0.4	17.1	1.9	0.989	6.8		70	1.9	0.4	17.1	2.0	1.035	6.7
Benzene	CAN	86	7.0	2.3	14.0	2.6	0.372	0.9	BOR	79	4.6	0.8	11.7	1.9	0.406	1.1	VAL	88	4.8	1.4	11.1	2.0	0.410	0.7
Toluene		86	38.2	10.8	242.8	27.3	0.715	5.2		78	24.5	6.1	270.4	30.0	1.228	7.3		88	22.1	2.4	184.3	20.1	0.909	6.2
Ethylbenzene		86	9.4	1.3	19.5	3.7	0.390	0.6		78	5.6	2.0	17.9	2.6	0.464	1.8		88	5.5	0.8	12.9	2.7	0.481	0.6
Xylene		86	39.1	6.6	89.5	15.2	0.388	0.8		78	22.4	7.8	70.8	10.2	0.456	1.7		88	21.6	2.7	63.0	11.0	0.510	1.0
Temp		86	15.3	3.3	29.8	7.1	0.465	0.0		79	15.9	3.7	29.8	7.1	0.445	−0.1		88	15.2	3.3	29.8	7.1	0.470	0.1
Wind		86	2.0	0.4	17.1	1.8	0.981	7.2		79	1.9	0.4	17.1	1.9	0.996	7.0		88	1.8	0.4	17.1	1.8	0.973	7.2
Benzene	BAT	90	7.3	2.5	14.3	2.6	0.360	0.7	GAL	83	4.9	1.5	11.2	2.0	0.406	1.0	PAN	81	1.6	0.2	4.2	0.8	0.479	1.2
Toluene		90	41.8	12.9	256.6	28.0	0.671	5.3		82	25.6	6.2	263.8	29.5	1.153	6.8		81	4.0	0.9	62.3	6.9	1.699	7.9
Ethylbenzene		90	10.3	4.7	19.7	3.6	0.352	0.7		83	5.9	1.8	13.1	2.6	0.437	0.8		70	1.0	0.3	4.7	0.7	0.760	2.6
Xylene		90	42.2	19.1	83.2	14.2	0.338	0.7		83	23.6	7.7	48.3	10.0	0.422	0.6		79	3.4	0.3	38.0	4.6	1.353	6.0
Temp		90	15.2	3.3	29.8	7.0	0.464	0.1		83	15.5	3.3	29.8	6.8	0.442	0.0		81	15.6	3.3	29.8	7.1	0.457	−0.1
Wind		90	1.8	0.3	17.1	1.8	0.982	7.2		83	1.9	0.3	17.1	1.8	0.991	7.0		81	1.9	0.4	17.1	1.9	1.002	7.0
Benzene	SEV	80	5.7	2.3	12.1	2.1	0.363	0.8	ITA	67	7.6	3.9	14.7	2.5	0.324	1.1	ALV	89	4.7	1.8	8.9	1.7	0.359	0.6
Toluene		80	29.3	8.6	154.5	18.8	0.640	4.1		67	37.2	7.9	81.8	13.8	0.371	0.9		89	21.5	5.7	171.3	18.1	0.843	6.6

Table 3 (continued)

	Site	N	Mean	Min	Max	S.D.	VC	SC	Site	N	Mean	Min	Max	S.D.	VC	SC	Site	N	Mean	Min	Max	S.D.	VC	SC
Ethylbenzene		80	7.3	2.3	18.5	3.0	0.418	0.9		67	10.1	2.3	19.9	3.5	0.341	0.8		89	6.2	2.0	13.6	2.4	0.386	0.8
Xylene		80	30.0	9.1	64.4	12.1	0.402	0.8		67	41.8	8.8	83.8	14.2	0.339	0.8		89	25.1	9.0	59.7	10.2	0.404	1.0
Temp		80	15.5	3.3	27.1	7.0	0.452	−0.1		67	16.3	3.3	29.8	7.1	0.436	−0.2		89	15.2	3.3	29.8	7.1	0.466	0.0
Wind		80	1.9	0.4	17.1	1.9	0.974	6.9		67	1.9	0.3	17.1	2.0	1.067	6.7		89	1.9	0.4	17.1	1.8	0.961	7.3
Benzene	DAL	83	5.6	2.1	12.5	2.1	0.377	0.8	IST	89	3.7	0.9	7.7	1.4	0.388	0.6	ROS	80	6.3	1.7	13.7	2.5	0.405	0.6
Toluene		83	29.0	9.3	265.9	28.3	0.976	7.3		89	17.3	4.6	272.7	28.1	1.623	8.7		80	28.7	5.6	68.3	12.8	0.447	0.6
Ethylbenzene		83	6.7	2.2	13.5	2.5	0.380	0.6		89	4.2	1.7	9.6	1.7	0.407	0.9		80	8.0	1.5	20.1	3.7	0.462	0.8
Xylene		83	27.0	7.9	58.7	10.7	0.397	1.0		89	16.4	6.1	43.6	6.9	0.418	1.1		80	31.3	5.9	74.4	13.6	0.435	0.6
Temp		83	15.6	3.7	29.8	6.9	0.442	0.0		89	15.2	3.3	29.8	7.0	0.464	0.0		80	15.3	3.3	29.8	7.3	0.480	0.0
Wind		83	1.8	0.4	17.1	1.8	1.006	7.3		89	1.9	0.4	17.1	1.8	0.957	7.3		80	1.8	0.4	17.1	1.9	1.012	7.0

S.D. standard deviation, *VC* variability coefficient, *SC* skewness coefficient

relation, which dictates the topology, or structure of the map and thus similar, from the chemical point of view, air samples should be mapped close together on a grid. In time iterative training, an algorithm constructs grid nodes in the SOM in order to represent the whole data set, and their weights are optimized. In our study, the batch training algorithm was applied (Ripley 1996; Vesanto et al. 2000; Sarkisian 2010). When an SOM net has been trained, the map needs to be evaluated to find out if it has been optimally trained or if further training is required. The number of grid nodes to be used in an SOM analysis may vary from a few dozen up to several thousand and can be considered as a trade-off between representation accuracy. A small number of grid nodes will result in a high quantization error (QE, scaled as 1–0) and well-defined clusters, while a large number of grid nodes result in a low quantization error and, in the most extreme case, a cluster for each data sample. In other words, the lower the quantization error, the higher the number of clusters to be interpreted. Quantization error being traditionally related to all forms of vector quantization and clustering algorithms is a measure which completely disregards map topology and alignment. The QE is computed by determining the averaged distance of the sample vector to the cluster centroids by which they are represented. In case of the SOM, the cluster centroids are the prototype vectors. In order to assess the SOM topology, preservation is quantified and represented by a topographic error (TE) (Kohonen 2001). It is the most simple of the topology preservation measures. The lower the topographic error (close to 0), the better the SOM which preserves the topology (Arsuaga and Díaz 2005). TE is computed as follows: for all data samples, the respective best and second best matching units are determined. If these are not adjacent on the map lattice, this is considered an error. The total error is then normalized to a range from 0 to 1, where 0, as mentioned above, means perfect topology preservation. Unfortunately, there are no reference values of QE and TE. They strongly depend on the map resolution. According to Peeters et al. (2006), quantization error rapidly decreases with increasing number of grid nodes, while topographic error converges to a stable value. As suggested by Vesanto et al. (2000), the most satisfactory compromise between clustering ability, quantization error, and topographic error could be achieved when the number

of nodes (n) is determined using the following formula: $n = 5 \cdot \sqrt{\text{number of samples}}$.

2.4.3 Clustering

Once the SOM has converged, the weight vectors of the chemical and meteorological variables were fed into a non-hierarchical K-means algorithm in order to find clusters in the neurons of the SOM. Partitioning by K-means implies the user to decide the final number of K groups (or clusters) the algorithm will converge into. Different numbers of predefined clusters were tried, and finally, the seven cluster pattern with the lowest Davies–Bouldin (DB) index (a function of the ratio of the sum of within-cluster scatter and between-cluster separation) (Davies and Bouldin 1979) was chosen. In order to assess the significance of the clustering pattern, non-parametric tests such as the Kruskal–Wallis and Dunn's tests were used; however, the Kruskal–Wallis test results were not discussed since it only confirms or denies an existence of a difference with respect to a given variable among more than two populations (in our case clusters). When adequate confirmation was obtained, Dunn's test was consecutively calculated. All calculations in this study were performed using Matlab 2007 (MathWorks, Inc.), Prism 5 Trial Version (GraphPad Software, Inc.), and Statistica 10.0 (Statsoft, Inc.) software running on a Windows VISTA platform.

3 Results and Discussion

Monthly average concentrations of benzene, toluene, ethylbenzene, xylene, as well as wind speed and temperature values are shown as basic statistics in Table 2. Besides the mean value, minimum (Min), and maximum (Max), standard deviation (S.D.), variability coefficient (VC), and skewness coefficient (SC) were also depicted in Table 2. VC allows comparison of the standard deviations of different variables that are in different units of measure. This is useful when comparing measurements across multiple heterogeneous data sets or across multiple measurements taken on the same data set. It is calculated according to the formula $VC = \frac{\text{S.D.}}{\bar{x}}$, while skewness coefficient is a measure which describes asymmetry from the normal distribution in a set of statistical data: negative skewness indicates left-skewed data distribution; while positive, right-skewed. The skewness coefficient values

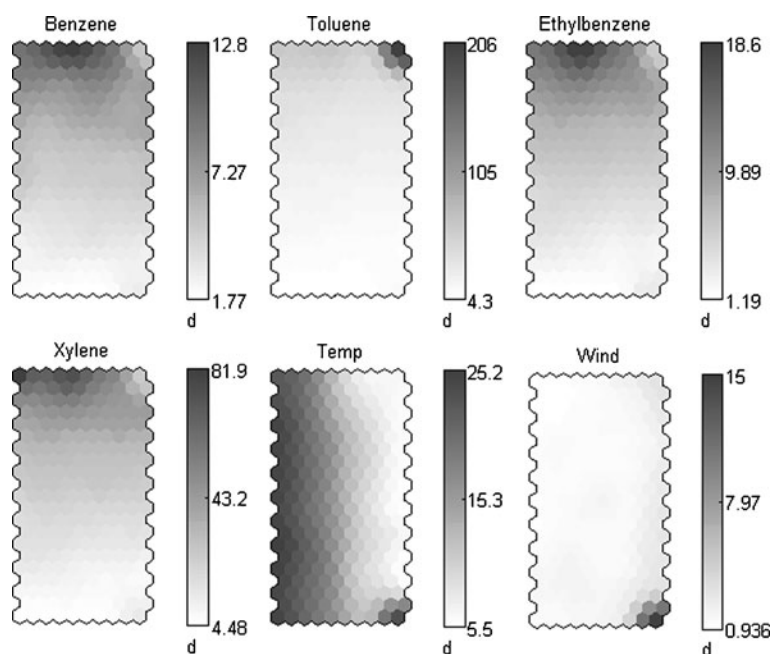presented in this study were calculated according to the formula presented by Doane and Seward (2011):

$SC = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\text{S.D.}} \right)^3$, where n is the sample size.

Considering benzene and toluene mean concentrations, Garibaldi, Canova, Battisti, F. Severo, Corso Italia, d'Alviano streets and Dalmazia square are qualified as the most polluted sites, with a mean higher than 5 and 28 μgm^{-3} of benzene and toluene, respectively. All highly polluted sites are located in the city center or widespread downtown (Figs. 1 and 2). As a result from SC values in majority of the locations, strong (toluene and wind) or moderate (benzene, ethylbenzene, and xylene) data skewness is observed and the application of the most common linear methods for statistical analysis could not be sufficient. In order to analyze the spatial and temporal variations of BTEX concentration in air in the city of Trieste, the data (consisting of 10,368 analytical and meteorological results) were arranged in a two-way array of constant dimensionality of six variables (benzene, toluene, ethylbenzene, and xylene (μgm^{-3}), temperature (°C), and wind speed (ms^{-1})) and 1,728 air samples collected at 21 sampling sites. When the concentration of a given analyte was below the detection limit, the replacement code "NaN" (not a number) was used instead of the missing value in the data set due to software requirements, as the self-organizing map algorithm implemented in the Matlab environment as a SOM Toolbox has the ability to deal with missing data. If an observation is missing a value on a specific variable, and hence coded as NaN, that variable is omitted from the distance calculation for that observation (Sarkisian 2010). As a result, NaNs are neither treated as zero nor lack of sample. Replacement code NaN was put instead of the missing value for benzene (0.06 % of total samples), toluene (0.29 %), ethylbenzene (1.1 %), and xylene (0.23 %). The total number of Kohonen's map neurons was estimated as $n = 5 \cdot \sqrt{1,728} \approx 208$. Such estimated number of map neurons could be obtained by more than one network dimension and this is why several of them were considered at this stage. In order to efficiently use the full space of the SOM, considered dimensions were limited and varied between 15 and 29 for vertical and between 7 and 14 for horizontal. In light of a number of possible combinations of final dimensions which gives the estimated number of neurons (i.e., $29 \times 7 = 203$, $26 \times 8 = 208$, $23 \times 9 = 207$, $21 \times 10 = 210$, $19 \times 11 = 209$, $17 \times 12 = 204$, $16 \times 13 = 208$, and $15 \times 14 = 210$), QE and TE were calculated in all cases. Finally, the

chosen dimensionality of 19×11 was characterized by the lowest values of quantization and topographic errors, 0.520 and 0.043, respectively. Moreover, a hexagonal lattice was preferred because it does not favor either horizontal or vertical direction (Kohonen 2001). Once the SOM's grid had been optimized, the individual variable planes were visualized (Fig. 3).

Each variable has its own component plane. This representation usually is a slice through the weights vector, scaling the highest (or most positive) weight to 1 and the lowest (or least positive) to 0 for each of the variables. In our case, component planes were scaled to represent the range of changeability of a given parameter. Analyzing those planes, we concluded that high concentration values of the majority of BTEXs are located in the middle and left parts of the upper area, while high contents of toluene and low content of benzene, ethylbenzene, and xylene are located in the right part of the upper area. Simultaneously, relatively well-organized distribution of temperature's plane suggested seasonality. Taking into consideration the important information from the SOM theory that each node of the SOM map could be consecutively referred to one or more air samples, it could be summarized that the differentiated structure of BTEX abundance (reflected in different color scales in the planes) suggested the existence of several similarity clusters in the set of samples.

As a consequence mentioned above, weight vectors of the converged map were clustered based on a K-means clustering mode. Different predefined numbers of clusters (K) were tried, and the sum of squares for each run was calculated. The best partition was obtained for a seven-cluster configuration having the lowest DB index value (Fig. 4). The numbers plotted within the map cells show how many times the weights of a given map neuron resembled the most of the input vector. The node with a weight vector closest to the input vector is tagged as the best matching unit, and the number of tagging is summarized. Finally, the distribution of the sample vectors along a Kohonen map can be studied by decoding the best matching unit selection events. Setting up the cluster composition with codes of the streets and names of the months (Fig. 5) allowed assessment of the spatial and temporal variations of BTEXs in the city of Trieste. The differences in BTEX concentration in air samples clustered as CI–CVII, presented in Fig. 6, were examined by Dunn's multiple comparison test (Table 4).

Its results show that in the case of more than 70 % comparisons per variable, the difference in rank sum was statistically important with the domination of p level value lower than 0.001. Most often, the differences were discovered between CI, CIII, and CIV and other clusters. Almost total lack of differences between CII

Fig. 3 SOM planes for meteorological variables as well as BTEXs determined in air samples

Fig. 4 Clustering pattern according to the Davies–Bouldin index minimum value
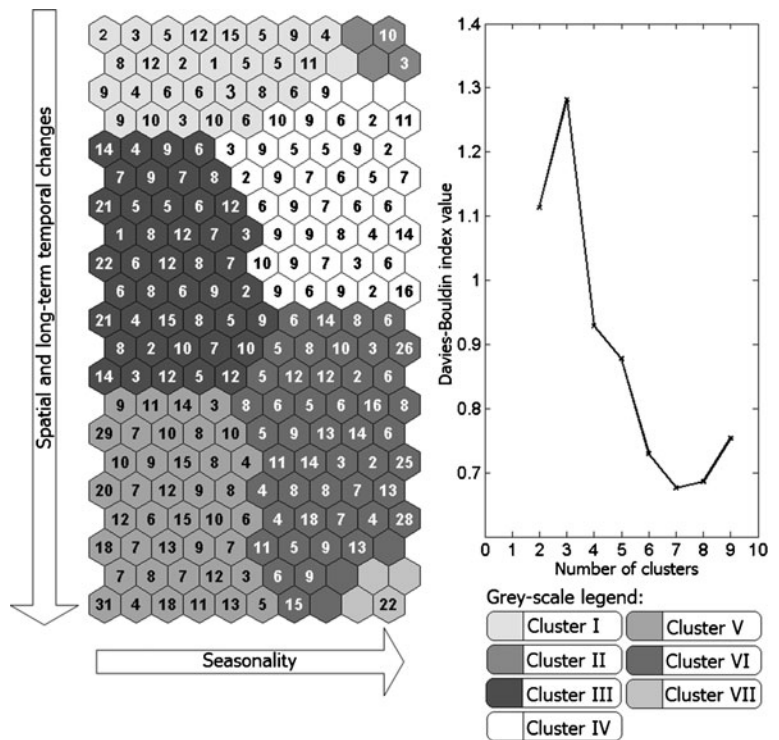
and other clusters were discovered for benzene, ethylbenzene, and xylene. CV, CVI, and CVII did not differ one another, while BTEXs were considered. It might be concluded that the clusters are not homogeneous. Among the seven clusters, cluster I includes 179 samples (10.36 %) with the highest benzene, ethylbenzene, and xylene concentrations. Most of the samples included in CI were collected at Garibaldi Sqr. (42), Battisti (32) and Canova (26) Streets and Corso Italia (22), which create a compact city center. These data show how the downtown area of Trieste looks the most polluted by BTEXs part. There is a lack of the seasonal pattern in this case since the contribution of samples

collected in successive months is comparable (median contribution to the cluster's population is 13.5 samples per month). A temporal assessment on the whole span of years proves consecutive improvement of the air quality in the city center: 86 % of the samples of CI were uniformly collected in the period between January 2001 and December 2004, while only 14 % in the period between January 2005 and December 2008. This phenomenon, being in agreement with Fig. 2, proves that the strongest pollution events were common before 2004 and that the air quality in the city of Trieste significantly increases. As a result from the comparison of variable planes (Fig. 3), high abundance of BTEXs
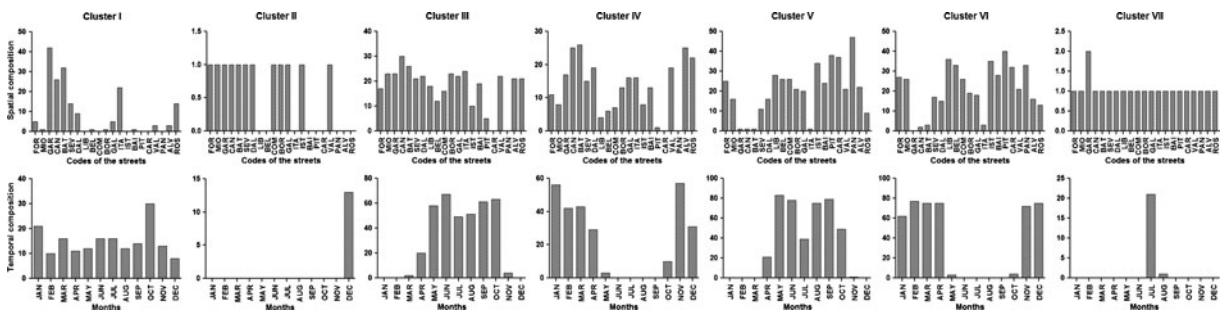

Fig. 5 Cluster composition according to the quantity of samples collected in a given location and in a given month
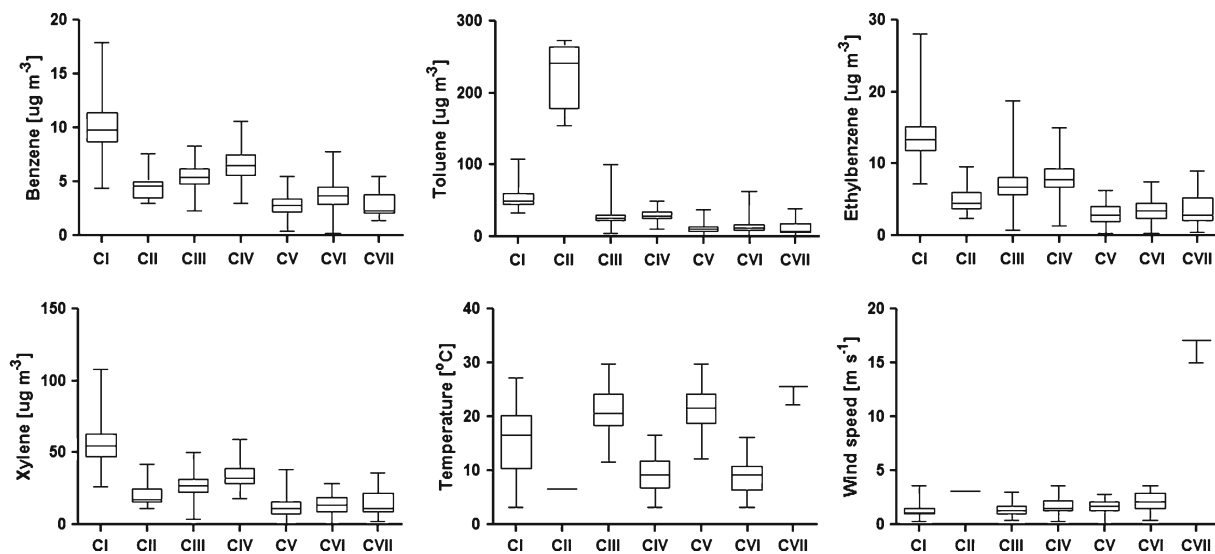
Fig. 6 Benzene, toluene, ethylbenzene, and xylene levels as well as temperature and wind speed values according to clustering pattern (*central horizontal line* median, *box* 25–75 % percentile, *whiskers* minimum–maximum)

Table 4 Statistical analysis (Dunn's test) in the combination of clusters obtained using SOM approach

	Benzene		Toluene		Ethylbenzene		Xylene	
	Differences in rank sum	*p* level	Differences in rank sum	*p* level	Differences in rank sum	*p* level	Differences in rank sum	*p* level
CI vs. CII	746	<0.001	−114.4	n.s.	821.8	<0.001	789.3	<0.001
CI vs. CIII	504.9	<0.001	479.8	<0.001	489.1	<0.001	507	<0.001
CI vs. CIV	315.6	<0.001	346.6	<0.001	336.3	<0.001	303.6	<0.001
CI vs. CV	1,252	<0.001	1,190	<0.001	1,202	<0.001	1,200	<0.001
CI vs. CVI	1,001	<0.001	1,083	<0.001	1,098	<0.001	1,092	<0.001
CI vs. CVII	1,227	<0.001	1,124	<0.001	1,091	<0.001	1,071	<0.001
CII vs. CIII	−241.1	n.s.	594.1	<0.001	−332.7	n.s.	−282.3	n.s.
CII vs. CIV	−430.4	n.s.	461	<0.05	−485.6	<0.05	−485.7	<0.05
CII vs. CV	506.1	<0.01	1,304	<0.001	380.2	n.s.	410.8	n.s.
CII vs. CVI	255	n.s.	1,197	<0.001	276.5	n.s.	302.7	n.s.
CII vs. CVII	480.6	n.s.	1239	<0.001	269.3	n.s.	281.9	n.s.
CIII vs. CIV	−189.3	<0.001	−133.1	<0.05	−152.8	<0.01	−203.4	<0.001
CIII vs. CV	747.1	<0.001	709.9	<0.001	712.9	<0.001	693.1	<0.001
CIII vs. CVI	496.1	<0.001	603	<0.001	609.2	<0.001	585.1	<0.001
CIII vs. CVII	721.6	<0.001	644.4	<0.001	602.1	<0.001	564.2	<0.001
CIV vs. CV	936.4	<0.001	843.1	<0.001	865.7	<0.001	896.5	<0.001
CIV vs. CVI	685.4	<0.001	736.1	<0.001	762.1	<0.001	788.5	<0.001
CIV vs. CVII	910.9	<0.001	777.5	<0.001	754.9	<0.001	767.6	<0.001
CV vs. CVI	−251	<0.001	−107	<0.05	−103.7	<0.05	−108	<0.05
CV vs. CVII	−25.49	n.s.	−65.54	n.s.	−110.9	n.s.	−128.9	n.s.
CVI vs. CVII	225.5	n.s.	41.42	n.s.	−7.184	n.s.	−20.85	n.s.

n.s. not significant

was discovered in the downtown area during the meteorological condition characterized by low wind speed (<0.93 ms^{-1}) both at low and high temperatures.

Cluster II consists of only 13 (0.75 %) samples collected in December 2002 with a concentration of toluene exceeding 150 μgm^{-3}. The small size of the CII and the fact that all other pollutants were determined on relatively low levels indicate that the samples belonging to CII represent accidental pollution by toluene, mainly limited to the sampling sites belonging to compact downtown area (Battisti Str. (BAT), Borsa square (BOR), Gallina Str. (GAL), Corso Italia (ITA), Canova Str. (CAN), and Garibaldi Str. (GAR)) or widespread downtown (F. Severo Str. (SEV), Dalmazia square (DAL), Mioni Ave. (MIO), Foraggi square (FOR), and dell'Istria Str. (IST)). The reason of this accidental increase in toluene concentration could be related to street floor painting for road indications.

Clusters III (375 samples, 21.70 %) and IV (271 samples, 15.68 %) consist of samples of moderate BTEX concentrations. Spatial analysis indicates that most of the samples in both cases were collected in the downtown (GAR (23), CAN (30), BAT (26), BOR (23), GAL (22), and ITA (24)) or in the widespread downtown (SEV (21), DAL (22), ALV (21), and ROS (21)) of the city, while temporal assessment proves an existence of a clear seasonal pattern. Samples clustered in CIII were mainly collected in summer; while in CIV, in winter time. Sampling stations located in the southern suburban areas (Monte S. Pantaleone (PAN), Carpineto Str. (CAR), and Pitacco Str. (PIT)) were not included to CIII and CIV or appeared occasionally. The only exception from the rule described above is VAL, which even if being located southerly from downtown still appears in CIII and CIV. The reason of this phenomenon might be related to the traffic caused by the proximity of the stadium. Higher median (Table 2) and mean (Fig. 6) concentrations of BTEXs determined in winter than in summer time can be explained with the concurrence of several factors as seasonality in emissions and in atmospheric oxidants that lower BTEX as well as with lower height of mixed layer provoking higher pollutant concentrations during cold months. For the pluriennal assessment, 68 and 64 % of the samples derived respectively from CIII and CIV were collected in the period between January 2001 and December 2004, while the rest (32 and 36 %) were collected after January 2005. The relative prevalence of data collected

before 2005 also in the moderately polluted clusters CIII and CIV confirms the improvement of the air quality in the city. Average monthly concentrations in the whole city of Trieste for the different BTEXs within the 8 years are reported in Fig. 7 in order to support this observation.

The slope coefficient of the time series is negative for each pollutant confirming definitively the trend of air quality improvement. Similarly to cluster composition described above, CV (425 samples, 24.60 %) and CVI (443 samples, 25.63 %) reflect to clear seasonal pattern, and in both cases, they consist of samples showing the lowest concentrations of BTEXs. In agreement with the expectations, most of the samples were generally collected in the northern (COM, BEL, and LIB) and southern (PAN, CAR, and PIT) Trieste. The relatively high contribution of samples collected after 2005 in CV and CVI (68 and 67 %, respectively), a significant part of which are collected also in downtown, again proves continues improvement of air quality in the city.

The last cluster CVII consists of 22 samples (1.27 %), and almost all of them (20) were collected in July 2007 when very strong wind occurred occasionally (17.1 ms^{-1}). An isolation of the CVII is mainly due to extraordinary meteorological condition.

4 Conclusions

A multi-year study on spatial and temporal variation of BTEX concentrations in an urban environment is reported. More than 1,700 samples, consisting of passive samples each exposed for 3 weeks in 21 sites in 8 years, were collected in the city of Trieste (Italy) and analyzed for BTEX concentrations together with the meteorological data providing a relevant base of 10,368 data useful for the assessment of urban VOC evolution. The method used in this research allowed, by application of one chemometric technique, for a logical spatiotemporal grouping of air quality samples collected in the city of Trieste. The SOM modeling approach demonstrated to be adequate for managing such an extended data collection, characterized by relative numerosity and some missing data. Interpretable clusterization of data and meaningful data representation are provided.

A partition of urban areas/zonation based on air quality parameters is provided, and a change in time
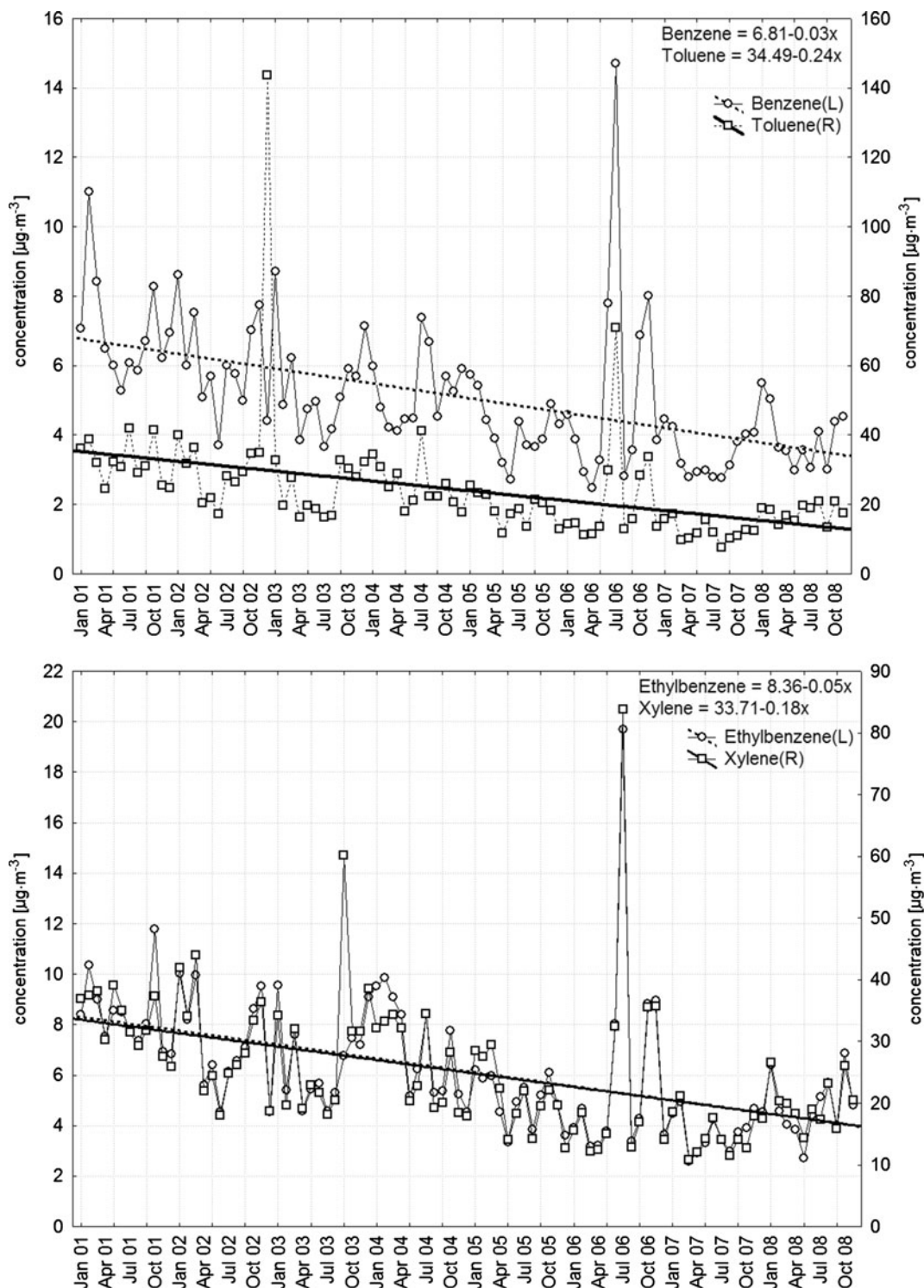
Fig. 7 Long-term trend of BTEX concentration in air of the city of Trieste

is shown, making the most polluted clusters/situations to disappear in recent years. This evidence has

probably some co-determining factors (Nuvolone et al. 2009); the changes in private and public vehicular

fleets, urban management measures as the ban of most of fueling stations from city center, and the limitation for the access to the city center imposed to vehicles with the highest emission had reasonably played a role. The zonation based on similarities revealed by passive samplers positioned at different locations can provide hints for optimization of the air quality monitoring network equipped with more expensive wired automatic instruments.

Results indicate that during the considered timeframe accordingly to the sampling scheme, downtown Trieste results the most polluted area among those considered. The level of pollution—and hence hazard for humans—decreases together with an increase of the distance between particular location and city center. Clustering based on self-organizing allowed to discover both long- and short-term (seasonality) changes in the air quality in the city of Trieste. An existence of seasonality in BTEX atmospheric concentrations was confirmed by higher abundance of benzene, toluene, ethylbenzene, and xylene in winter than in summer time, pointing at the importance of dispersion associated to the expansion of the mixing layer during the warm season.

The considered monitoring and data analysis scheme allows an effective assessment of the evolution of pollution mitigation associated with traffic impacts. The negative slope coefficient of the BTEX concentration time series confirmed definitively the tendency of air quality improvement in the area of interest, showing the experimental effects of EU Directives aimed at the reduction of BTEX concentrations in air.

References

Agency for Toxic Substances and Disease Registry (ATSDR). (2012). *Interaction profile for benzene, toluene, ethylbenzene, and xylenes (BTEX)*. Atlanta: Department of Health and Human Services, Public Health Service.

Arsuaga, U. A., & Díaz, M. F. (2005). Topology preservation in SOM. *International Journal of Mathematical and Computer Sciences, 1*(1), 19–22.

Astel, A., Tsakovski, S., Barbieri, P., & Simeonov, V. (2007). Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research, 41*, 4566–4578.

Astel, A., Walna, B., Simeonov, V., & Kurzyca, I. (2008). Multivariate statistics as means of tracking atmospheric pollution trends in western Poland. *Journal of Environmental Science and Health, Part A, 43*, 313–328.

Astel, A. (2010). Air contaminants modeling by use of several receptor oriented models. *International Journal of Environment and Pollution, 42*, 32–57.

Astel, A., Simeonov, V., Bauer, H., & Puxbaum, H. (2010). Multidimensional modeling of atmospheric monitoring data. *Environmental Pollution, 158*, 3201–3208.

Astel, A., Cozzutto, S., Cozzi, F., Adami, G., Barbieri, P., Tsakovski, S., et al. (2010). Seasonal apportionment of the sources of ambient air particulates in the city of Trieste. *International Journal of Environment and Pollution, 41*, 70–89.

Brocco, D., Fratarcangeli, R., Lepore, L., Petricca, M., & Ventrone, I. (1997). Determination of aromatic hydrocarbons in urban air of Rome. *Atmospheric Environment, 31*, 557–566.

Bruno, P., Caselli, M., de Gennaro, G., Scolletta, L., Trizio, L., & Tutino, M. (2008). Assessment of the impact produced by the traffic source on VOC level in the urban area of Canosa di Puglia (Italy). *Water, Air and Soil Pollution, 193*, 37–50.

Caselli, M., de Gennaro, G., Marzocca, A., Trizio, L., & Tutino, M. (2010). Assessment of the impact of the vehicular traffic on BTEX concentration in ring roads in urban areas of Bari (Italy). *Chemosphere, 81*, 306–311.

Chen, X., Zhang, G., Zhang, Q., & Chen, H. (2011). Mass concentrations of BTEX inside air environment of buses in Changsha, China. *Building and Environment, 46*, 421–427.

Cocheo, C., Boaretto, C., Pagani, D., Quaglio, F., Sacco, P., Zaratin, L., et al. (2009). Field evaluation of thermal and chemical desorption BTEX radial diffusive sampler radiello® compared with active (pumped) samplers for ambient air measurements. *Journal of Environmental Monitoring, 11*, 297–306.

Cogliano, V. J., Baan, R., & Straif, K. (2011). Updating IARC's carcinogenicity assessment of benzene. *American Journal of Industrial Medicine, 54*, 165–167.

Cottrell, M., Fort, J. C., & Pagès, G. (1998). Theoretical aspects of the SOM algorithm. *Neurocomputing, 21*, 119–138.

Davies, D.L., Bouldin, D.W. (1979). A cluster separation measure. In: *Proceeding of the IEEE Transactions on pattern recognition and machine intelligence* 1(2), 224–227.

Doane, D. P., & Seward, L. E. (2011). Measuring skewness: a forgotten statistic? *Journal of Statistics Education, 19*(2), 1–18.

Edwards, R. D., Jurvelin, J., Saarela, K., & Jantunen, M. (2001). VOC concentrations measured in personal samples and residential indoor, outdoor and workplace microenvironments in EXPOLIS-Helsinki, Finland. *Atmospheric Environment, 35*(27), 4531–4543.

Esplugues, A., Ballester, F., Estarlich, M., Llop, S., Fuentes-Leonarte, V., Mantilla, E., et al. (2010). Indoor and outdoor air concentrations of BTEX and determinants in a cohort of one-year old children in Valencia, Spain. *Science of the Total Environment, 409*(1), 63–69.

Fort, J. C. (2006). SOM's mathematics. *Neural Networks, 19*, 812–816.

Giraudel, J. L., & Lek, S. (2001). A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecological Modelling, 146*, 329–339.

Hoque, R. R., Khillare, P. S., Agarwal, T., Shridhar, V., & Balachandran, S. (2008). Spatial and temporal variation of BTEX in the urban atmosphere of Delhi, India. *Science of the Total Environment, 392*, 30–40.

Hun, D. E., Corsi, R. L., Morandi, M. T., & Siegel, J. A. (2011). Automobile proximity and indoor residential concentrations of BTEX and MTBE. *Building and Environment, 46*(1), 45–53.

Infante, P. F. (2011). The IARC October 2009 evaluation of benzene carcinogenicity was incomplete and needs to be reconsidered. *American Journal of Industrial Medicine, 54*, 157–164.

Johnson, M. M., Williams, R., Fan, Z., Lin, L., Hudgens, E., Gallagher, J., et al. (2010). Participant-based monitoring of indoor and outdoor nitrogen dioxide, volatile organic compounds, and polycyclic aromatic hydrocarbons among MICA-Air households. *Atmospheric Environment, 44*(38), 4927–4936.

Kim, Y. P., Kim, J. Y., Ghim, Y. S., & Na, K. (1997). Ambient concentrations of volatile organic compounds in Seoul. *Environmental Research Forum, 7–8*, 265–269.

Keymeulen, R., Görgényi, M., Héberger, K., Priksane, A., & Van Langenhove, H. (2001). Benzene, toluene, ethyl benzene and xylenes in ambient air and *Pinus sylvestris* L. needles: a comparative study between Belgium, Hungary and Latvia. *Atmospheric Environment, 35*, 6327–6335.

Khoder, M. I. (2007). Ambient levels of volatile organic compounds in the atmosphere of Greater Cairo. *Atmospheric Environment, 41*, 554–566.

Kohonen, T. (1984). *Self-organization and associative memory.* Berlin: Springer Verlag.

Kohonen, T. (1995). *Self-organizing maps.* New York: Springer Inc.

Kohonen, T. (2001). Self-organizing maps, third ed. Berlin: Springer.

Kohonen, T., Oja, E., Simula, O., Visa, A., Kangas, J. (1996). Engineering applications of the self-organizing map. In: *Proceeding of the IEEE* 84, 1358–1384.

Monteiro Martins, E., Arbilla, G., Favilla Bauerfeldt, G., & de Paula, M. (2007). Atmospheric levels of aldehydes and BTEX and their relationship with vehicular fleet changes in Rio de Janeiro urban area. *Chemosphere, 67*, 2096–2103.

Muezzinoglu, A., Odabasi, M., & Onat, L. (2001). Volatile organic compounds in the air of Izmir, Turkey. *Atmospheric Environment, 35*, 753–760.

Mukherjee, A. (1997). Self-organizing neural network for identification of natural modes. *Journal of Computing in Civil Engineering, 11*, 74–77.

Nuvolone, D., Barchielli, A., & Forastiere, F. (2009). Assessing efficacy of actions on urban mobility for the improvement of air quality and citizen health: a review (in Italian). *Epidemiological Preview, 33*(3), 79–78.

Peeters, L., Bação, F., Lobo, V., & Dassargues, A. (2006). Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM, a variant of Kohonen's self-organizing map. *Hydrology and Earth System Sciences Discussions, 3*, 1487–1516.

Pekey, B., & Yilmaz, H. (2011). The use of passive sampling to monitor spatial trends of volatile organic compounds (VOCs) at an industrial city of Turkey. *Microchemical Journal, 97*, 213–219.

Ras-Mallorquí, M. R., Marcé-Recasens, R. M., & Borrull-Ballarín, F. (2007). Determination of volatile organic compounds in urban and industrial air from Tarragona by thermal desorption and gas chromatography–mass spectrometry. *Talanta, 72*, 941–950.

Ripley, B. D. (1996). *Pattern recognition and neural networks* (pp. 154–155). UK: Cambridge University Press.

Roukos, J., Riffault, V., Locoge, N., & Plaisance, H. (2009). VOC in an urban and industrial harbor on the French North Sea coast during two contrasted meteorological situations. *Environmental Pollution, 157*, 3001–3009.

Roukos, J., Locoge, N., Sacco, P., & Plaisance, H. (2011). Radial diffusive samplers for determination of 8-h concentration of BTEX, acetone, ethanol and ozone in ambient air during a sea breeze event. *Atmospheric Environment, 45*(3), 755–763.

Sarkisian, N. (2010). Neural networks as an emergent method in quantitative research, an example of self-organizig maps. In S. Nagy Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 630–634). New York: Guilford Press.

Simon, V., Baer, M., Torres, L., Olivier, S., Meybeck, M., & Della Massa, J. P. (2004). The impact of reduction in the benzene limit value in gasoline on airborne benzene, toluene and xylenes levels. *Science of the Total Environment, 334–335*, 177–183.

Stanimirova, I., Daszykowski, M., & Walczak, B. (2007). Dealing with missing values and outliers in principal component analysis. *Talanta, 72*, 172–178.

Sturaro, A., Rella, R., Parvoli, G., & Ferrara, D. (2010). Long-term phenol, cresols and BTEX monitoring in urban air. *Environmental Monitoring and Assessment, 164*, 93–100.

Ultsch, A., Siemon, H.P. (1990). Kohonen's self organizing feature maps for exploratory data analysis. In: Proceedings of International Neural Network Conference (INNC'90). Kluwer Netherlands, Dordrecht.

Vesanto, J. (2000). *Neural network tool data mining: SOM Toolbox. In: Proceedings of Symposium on Tool Environments and Development Methods for Intelligent Systems (TOOL-MET2000)* (pp. 184–196). Oulu: Oulun yliopistopaino.

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankagas, J. (2000). SOM Toolbox for Matlab 5, Report A57, (Accessed 15 February 2007, http://www.cis.hut.fi/projects/somtoolbox/).

Villalta, R., Beziza, S., Bortolussi, A., Cecchin, W., Pellegrini, I., Giorgini, L., Salvagni, E., Moimas, F., Marchiol, A., Polesello, S. (2004). Determinazione di composti aromatici BTEX in aria ambiente per assorbimento con campionatori passivi ed analisi gascromatografica con rilevatore a ionizzazione di fiamma. Esperienza di intercalibrazione a livello regionale tra i laboratori dell'ARPA FVG, Bolletino dei Chimici Igienisti, 55 (in Italian).

Zhang, J., Dong, Y., & Xi, Y. (2008). A comparison of SOFM ordination with DCA and PCA in gradient analysis of plant communities in the midst of Taihang Mountains, China. *Ecological Informatics, 3*(6), 367–374.

Zielinska, B., Fujita, E., Ollison, W., Campbell, D., Sagebiel, J., Merritt, P., et al. (2012). Relationships of attached garage and home exposures to fuel type and emission levels of garage sources. *Air Quality, Atmosphere and Health, 5*, 89–100.

Radiello. (2003). Composti organici volatili (COV) desorbiti con CS2, Edizione 02/2003, http://www.radiello.com/italiano/cov_chim_it.htm. Accessed 10 Feb 2003.