Environmental Modeling & Assessment (2024) 29:915–932
https://doi.org/10.1007/s10666-024-09969-7

RESEARCH

Characterization of Non-Conventional Airborne Pollutants (BTEX) by means of Chemometric Techniques

Rubén Albeiro Sánchez-Andica[1] · Wilson Rafael Salas-Chávez[1] · Martha Isabel Páez-Melo[1]

Received: 23 February 2023 / Accepted: 6 March 2024 / Published online: 21 March 2024

Abstract

In this study, chemometric and spatial interpolation methods were employed to characterize non-conventional pollutants in the atmosphere of Santiago de Cali, Colombia. The pollutants were monitored using passive diffusion samplers during two distinct periods (January to February and March to April) in the two years after the pandemic (2021 and 2022). None of the monitored cases exceeded the concentration limits established by the National Air Quality Standard. Cluster analysis revealed distinct groups, categorizing sites into low, medium, and high pollutant influence. Principal Components Analysis (PCA) was employed to condense all variables into two primary constituents. The first component (PC1) serves as an indicator of mobile pollutant sources due to the consistent contribution of pollutants. Conversely, the second component (PC2) indicates punctual emissions of toluene, which made the most significant contribution. Spatial analysis demonstrated that downtown and the northern region of the city were highly influenced by PC1, with a substantial decrease in its effects towards the periphery, particularly the south. Utilizing Inverse Distance Weighting (IDW), we identified hotspots for both PCs, notably in areas undergoing real estate construction and the downtown industrial sector. Finally, our analysis revealed a cancer risk in the downtown and northeast areas of the city, associated with exposure to benzene and ethylbenzene. This observation aligns with the region of incidence indicated by PC1.

Keywords Atmospheric pollution · Volatile organic compounds · Chemometrics

1 Introduction

Air pollution in cities and towns is an escalating concern due to the continuous emission of a diverse range of pollutants. According to the World Health Organization (WHO) in 2012, air quality issues were linked to an estimated 6.5 million deaths, out of which 3.5 million were premature [1]. Urban centers primarily face atmospheric contamination from vehicle emissions, combustion processes, and industrial activities. Priority pollutants in atmospheric contamination studies include particulate matter, nitrogen and sulfur oxides, heavy metals, and various volatile organic compounds (VOCs), notably BTEX (benzene, toluene, ethylbenzene, and xylenes) [2, 3]. These pollutants have diverse impacts on human health. For instance, benzene affects the nervous system and is carcinogenic, linked to leukemia [4]. Toluene primarily impacts the nervous system, and prolonged exposure can lead to vision loss and permanent brain damage [5]. Ethylbenzene, when present in high levels, can cause vertigo, eye and throat irritation, and is considered a potential carcinogen [6]. Xylenes affect the nervous system (e.g., short-term memory) and the airways, potentially causing kidney and liver problems at high concentrations [7]. Various methodologies have been employed to determine contamination sources, enabling the establishment of preventive plans. Among these, chemometrics, a set of statistical methods, stands out [8]. Cluster Analysis (CA) is a prominent chemometric method enabling the classification of elements or variables [9–12]. Principal Component Analysis (PCA) is another vital method, summarizing variables

✉ Rubén Albeiro Sánchez-Andica
ruben.sanchez@correounivalle.edu.co

Wilson Rafael Salas-Chávez
wilson.salas@correounivalle.edu.co

Martha Isabel Páez-Melo
martha.paez@correounivalle.edu.co

[1] Grupo de Investigación en Contaminación Ambiental por Metales Pesados y Plaguicidas, Departamento de Química, GICAMP, Universidad del Valle, Calle 13 No 100-00, Cali 760032, Colombia

to a smaller set and facilitating the identification of pollution sources [13–19]. To correlate geographic information with pollutant concentration in hard-to-reach areas, spatial representation of pollution phenomena has been essential. Spatial interpolation methods like Inverse Distance Weighting (IDW) interpolation [20, 21] and kriging [22–27] have been utilized for this purpose. Risk assessment, a methodology estimating risk levels for living beings based on possible exposure scenarios, is also crucial. Due to the persistence of VOCs in urban areas, several authors suggest that the characterization of pollutants and spatial representation should be accompanied by risk assessment studies, considering the toxicological characteristics of polluting chemical compounds [28–32].

The primary objective of this study was to apply chemometric techniques to characterize and spatially represent VOCs, as well as to assess the carcinogenic risk based on data from passive pollutant sampling in Santiago de Cali, Colombia.

2 Methodology

2.1 Study Area

The pollutants were monitored within the city of Santiago de Cali, Colombia. Geographically, the city is positioned at 3°27'26" North latitude and 76°31'42" West longitude, situated at an elevation of 1,070 m above sea level (masl). Santiago de Cali is nestled between the Western Cordillera and the Central Cordillera. Serving as the capital of the Valle del Cauca department, it ranks as the third most populous city in Colombia, housing approximately 2,394,870 inhabitants (refer to Fig. 1).

The city spans an approximate area of 561.6 km², with the urban area constituting 120.9 km². Administratively, it is organized into 22 communes and 248 neighborhoods. The city shares its boundaries; to the north with the municipalities of Yumbo, La Cumbre, and Palmira, to the east with the municipality of Candelaria, to the west with the municipalities of Buenaventura and Dagua, and to the south with the municipality of Jamundí. The administrative division comprises 22 communes and 248 neighborhoods [33].

2.2 Sampling and Analysis

The sampling of contaminants was conducted by the passive sampling technique using Carbopack B as an adsorbent material (procured from the Swedish Environmental Research Institute).

A total of four (4) monitoring campaigns for two months (according to the Swedish Environmental Research Institute IVL recommendations) were conducted as follows: s1 (February-March 2021), s2 (March-April 2021), s3 (February-March 2022), and s4 (March-April 2022) in Santiago de Cali city. Across the city, 41 monitoring points were strategically located along streets and road corridors, covering all 22 communes (refer to Table 1).

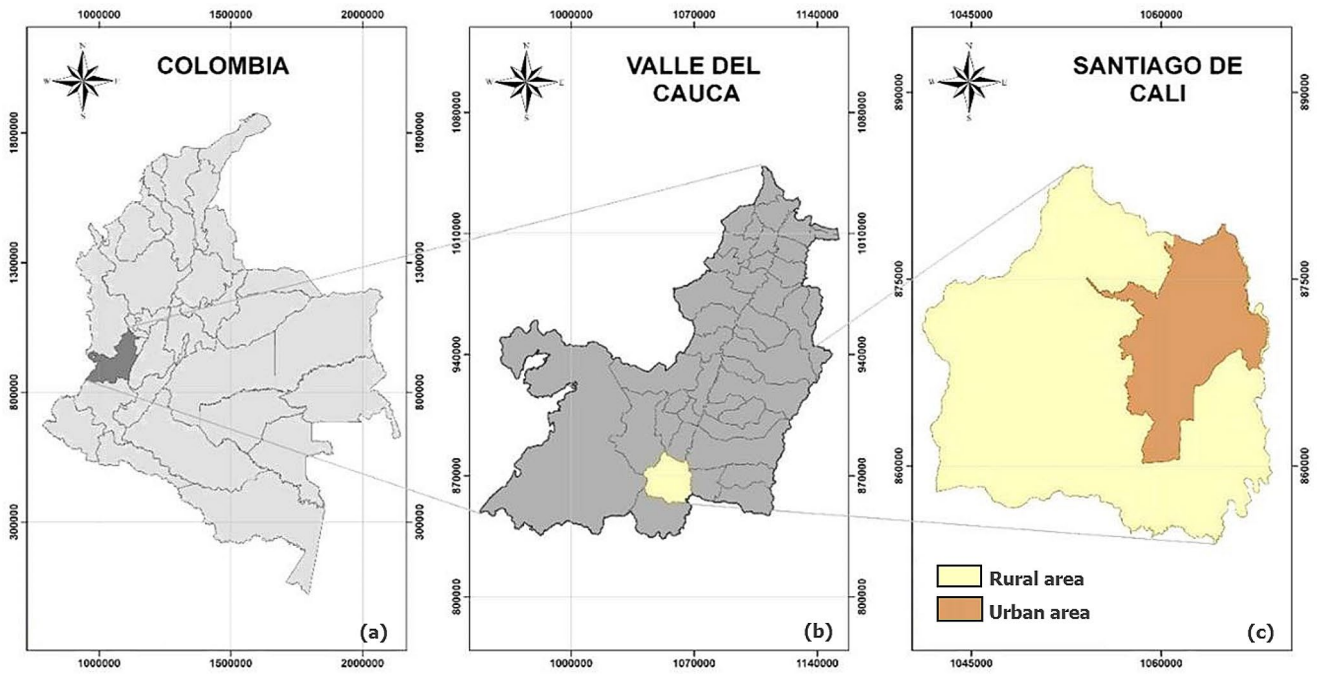

Fig. 1 Spatial geographical location in (a) Colombia of (b) Departamento del Valle and (c) the municipality of Santiago de Cali and the city of Santiago de Cali (urban area)

Notably, some of these sampling stations are operated by the *Departamento Administrativo de Gestión del Medio Ambiente* (DAGMA) as part of the Santiago Cali Air Quality Surveillance System. The spatial distribution of these stations within the city is illustrated in Fig. 2.

During the sampling campaigns, the average weather conditions were recorded as follows:

Period s1 (February-March 2021): Minimum temperature (Tmin) of 19.2 °C, maximum temperature (Tmax) of 28.6 °C, and rainfall of 5.7 mm [34].

Period s2 (March-April 2021): Tmin of 19.7 °C, Tmax of 28.4 °C, and rainfall of 2.7 mm [34].

Period s3 (February-March 2022): Tmin of 18.7 °C, Tmax of 29.2 °C, and rainfall of 3.4 mm [34].

Period s4 (March-April 2022): Tmin of 19.4 °C, Tmax of 29.6 °C, and rainfall of 6.1 mm [34].

According to data from the *Instituto de Hidrología, Meteorología y Estudios Ambientales* (IDEAM), the foremost hydroclimatology authority in Colombia, no significant deviations were observed in temperature fluctuations between the sampling periods and the respective years in 2021 and 2022.

The samplers were anchored to posts alongside the road at an average height of 6 m and left for a period of two months. After this time, the samplers were removed and taken to the laboratory for the respective quantification.

Analytical determination of volatile organic compounds (VOCs) was carried out at IVL laboratories following the standardized method EN 13528-2 2003 [35]. In short, this method involves thermal desorption of the compounds at 250 °C for 5 min, followed by sample concentration in a cold trap maintained at -30 °C and quantification by Gas Chromatography with flame ionization detector (GC-FID).

2.3 Analytical Performance

Stock solution of COVs were prepared from benzene, toluene, ethylbenzene, m,p-xylene, o-xylene, n-octane, and nonane (Sigma-Aldrich, USA) analytical standards. Chromatographic quantification of VOCs was conducted in a GC-FID Hewlett Packard 6890 series. A CP-Sil 5 CB GC column (0.32 mm id., 1.2 µm x 50 m) was used in the chromatographic separation. The oven temperature was raised from 35 to 300 °C with a rate of 20 °C/s. The measurements were acquired by external calibration in triplicate and reported as mean values. Precision was calculated as standard deviation (SD). The accuracy of the analytical method was determined by the recovery method by spiking a known concentration of VOCs at three levels (20, 50 y 90 µg/L) to the extracts and the recovery percentage (% R) for each VOC was calculated. The average of recovery percentages of all compounds was 90%. The Lineal range of the calibration plots were between 5 µg/L and 120 µg/L on average. The linearity of each plot was evaluated by the correlation coefficient (r) giving 0,995 (p-value < 0,05) on average. The instrumental limit of detection (LOD) and limit of quantification (LOQ) were calculated for each compound based on 3 and 10 times the Signal Noise ratio (S/N) respectively. The LOD and LOQ were 3.0 µg/L and 5.0 µg/L on average, respectively. Excel® was used for mathematical and statistical calculations when necessary.

2.4 Chemometric Characterization

2.4.1 Data Exploratory Analysis

For the exploratory statistical analysis, R® software was employed, utilizing the graphical interface of R Studio. Box plots were generated for each pollutant in every sampling period to identify potential trends. A statistical significance level of $p < 0.05$ was established. Additionally, each variable was normalized to a statistical value of Z_i calculated by

$$Z_i = \frac{x_i - \mu}{\sigma}, \tag{1}$$

Table 1 Codes for the places where the Volatile Organic Compounds (VOCs) were monitored

Code	Name	Code	Name	Code	Name	Code	Name
1	Icesi†	12	Nueva Floresta	23	Av 3ª C44†	34	Ermita Dagma* †
2	Pance Dagma*	13	Chapinero†	24	Chipichape	35	Luna*†
3	Calicanto	14	B Herrera	25	Primavera	36	Univalle Dagma*
4	MI Urrutia†	15	Comfandi Delicias	26	Guabal	37	Madrigal
5	Vallado	16	Maizena	27	Autopista K44†	38	Premier†
6	ET Dagma*	17	Flora Dagma*†	28	Club Noel†	39	Siloé
7	Calipso†	18	Floralia	29	Aguacatal†	40	Cañaveralejo Dagma*†
8	Mojica	19	Calima†	30	Terrón†	41	Parque Banderas†
9	Compartir Dagma*	20	Ciudad de Cali †	31	San Cayetano		
10	Puente Mil Días †	21	Cl 44 K1 †	32	CC Caleño †		
11	Base Area †	22	Bolivariano	33	ERA Dagma *†		

* = SVCASC's stations; † = Near to main streets

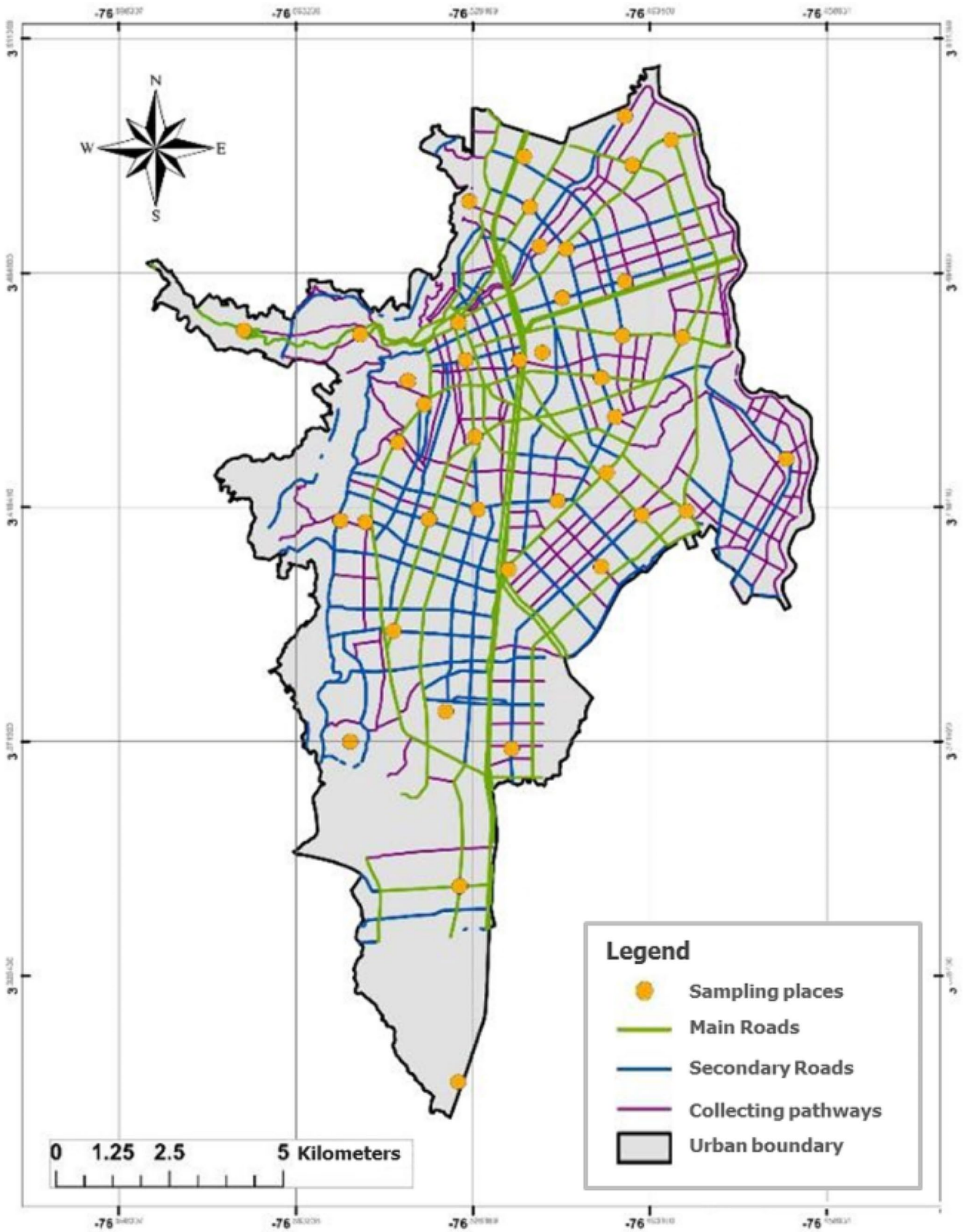

Fig. 2 Spatial location of the samplers in the VOCs monitoring in the urban area of the city of Santiago de Cali

where x_i represents the concentration, μ stands for the mean, and σ represents the standard deviation of the compound [36].

Furthermore, the correlation between different compounds was assessed using the Pearson correlation test. The correlation charts were created using the *cor* function and the *corrplot* package, enabling the examination of potential linear relationships among the VOCs, thus serving as an indicator of common emission sources [36].

2.4.2 Cluster Analysis

The cluster analysis is an unsupervised grouping method in such a way that objects are classified according to their similarities and differences [36]. Euclidean distance and Ward's agglomeration method have been reported as those that present the best structure for environmental data, for this reason, they were the methods that were used in the cluster analysis, to this end, it was set as criteria of similarity between two points *i* and *j*, Euclidian distance d_{ij}, was calculated using the Eq. 2.

Cluster analysis is an unsupervised grouping method used to classify objects based on their similarities and differences [36]. In this study, the Euclidean distance and Ward's agglomeration method were selected as they are known to provide the most effective structure for environmental data. For this purpose, the Euclidean distance, d_{ij}, between two points *i* and *j* was calculated by the following expression:

$$d_{ij} = \sqrt{\sum_{n=1}^{7} (x_{i,n} - x_{j,n})^2}, \tag{2}$$

where x_i and x_j represent the *n* characteristics of those points, in this case the 7 VOC concentrations. Initially, each object was considered as its own cluster, and iteratively, the most similar objects were linked. Using Ward's method, clusters were formed until a single total cluster composition was achieved [36]. The cluster analysis was conducted using the *hclust* function within the *stats* package.

2.4.3 Principal Components Analysis

Principal Components Analysis (PCA) is a multivariate analysis method commonly employed to reduce the dimensionality of a data array comprising interrelated variables. It effectively preserves most of the original information while simplifying the interpretation of variation sources [36, 37]. The method involves decomposing the original data matrix (*X*) as expressed by

$$\mathrm{X} = TL^T, \tag{3}$$

where *X* is a matrix of dimensions (n,p), *T* is the matrix of *scores* (n,d), *L* is the matrix of *loadings* (p,d) and the superscript *T* indicates the transpose of the matrix.

The resulting new variables, known as principal components (PC), are derived from the projection of *X* onto the matrix of scores (*L*), yielding *T* (Eq. 4), where columns represent the principal components [12, 36, 37]

$$\mathrm{T} = \mathrm{XL} \tag{4}$$

.The PCA was conducted using the PCA function from the *FactoMineR* package.

2.4.4 Spatial Analysis

Two interpolation methods were employed for spatial analysis: Inverse Distance Weighting (IDW) [19, 38] and kriging [38, 39].

Inverse Distance Weighting (IDW) This method assumes that each sampled point has a local influence, which decreases with the distance, giving more weight to closer points. IDW estimates a property x̂ at a point in space s_0 through the following expression.

$$\widehat{x}(s_0) = \frac{\sum_{i=1}^{N} (x(s_i)/d_i^p)}{\sum_{i=1}^{N} (1/d_i^p)}, \tag{5}$$

where *N* is the number of neighbors, $x(s_i)$ is the property at point s_i, d_i is the distance from s_i to s_0, and *p* is the power to which the distance raised. Here, $N = 12$ neighbors and $p = 2$, were used.

Kriging Kriging is based on the statistical principle of spatial autocorrelation, suggesting greater similarity between nearby points. Similar to IDW, it estimates the value of a property, x̂, at s_0 using the following expression.

$$\widehat{x}(s_0) = \sum_{i=1}^{N} \lambda_i x(s_i). \tag{6}$$

Here, N is the number of measured values, $x(s_i)$ is the property at a point, and λ_i, represents the weighting of the measured value at the location *i*. The weighting factor, λ, is known as the semi-variance, and was calculated by

$$\lambda(d_{ij}) = \frac{1}{2} \sum_{i=1}^{N} (x(s_i) - x(s_j))^2, \tag{7}$$

where d_{ij} is the distance between 2 points s_i and s_j and $x(s_i)$ and $x(s_j)$ are the values at those points. This yields a semi-variogram representing the degree of spatial correlation. In this study, ordinary kriging with a spherical semi-variogram model was employed. The calculations and cartographic representations were conducted using ArcGIS® 10.1.

2.4.5 Carcinogenic Risk Assessment

Carcinogenic risk (R) assessment was conducted for benzene and ethylbenzene using the EPA model [14, 28], as depicted as

$$R = SF\frac{CxIRxEFxETxED}{BWxAT}, \quad (8)$$

where SF represents the slope factor (2,73×10-2 y 3,85×10-3 mg/kg/day for benzene and ethylbenzene, respectively), C(si) is the pollutant concentration at the point si (mg/m3), IR is the inhalation rate (0,875 m3/h for adults), ET is the exposure time (4 h/day), EF is the exposure frequency (350 days/year), ED is the exposure duration (30 years), BW is the body weight (70 kg) and AT is the averaging Time (70 years). The study framework considered an adult spending a minimum of 4 h per day outdoors, for 350 days each year, over a minimum of 30 years [30]. Risk values were calculated for each sampling period *i* and both pollutants as the total carcinogenic risk TR_i (Eq. 9):

$$TR_i = BR_i + ER_i, \quad (9)$$

where BR_i and ER_i represent the benzene and ethylbenzene risk for the sampling period i. This model assumes non-interaction among pollutants and was represented cartographically using the kriging method to generate carcinogenic risk maps for the city of Santiago de Cali.

3 Results and Discussion

3.1 Exploratory Analysis

The box plots, depicted in Fig. 3, illustrate the concentrations of each analyte across the four sampling campaigns. Notably, there is a noticeable congruency between the s1-s3 and s2-s4 samplings, indicating a certain temporal stability in the monitored contaminants. The average VOC concentrations obtained from the four campaigns underwent One-way ANOVA, yielding a significance level (p-value) of 0.246, exceeding 0.05. Therefore, no significant difference was observed between each group.

The specific concentrations of each analyte during each sampling period are summarized in Table 2. Notably, the first campaign exhibited the highest average concentrations for all the VOCs analyzed and displayed the greatest dispersion of data.

In comparison with data from various cities worldwide (see Table 3) [40], the average benzene concentrations in this study are comparable to most cities, except for Rome and Bombay, where they are an order of magnitude lower. For toluene, the concentration in Cali is higher than that of benzene. However, the concentration ranges align with those of the cities mentioned in Table 3. The xylene isomers' concentrations in Cali were an order of magnitude higher than benzene, akin to the data from the Sao Paulo study.

Several studies suggest that the m,p-xylene/ethylbenzene (mpX/E) ratio can serve as an emissions indicator, primarily from mobile sources. In this study, the mpX/E ratio averages 4.8, closely resembling ratios reported by other authors

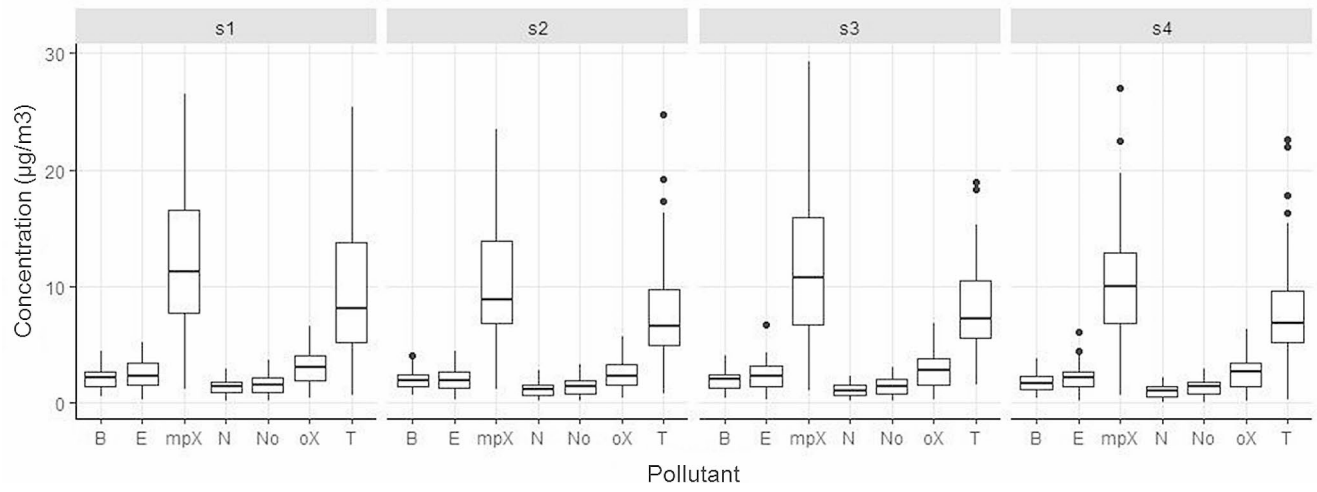

Fig. 3 Box plot of VOC concentrations for the comparison of pollutants in the 4 monitoring (s1, s2, s3 and s4)

Table 2 Summary of VOC concentrations (μgm^{-3}) for each sampling period (S_1, S_2, S_3 and S_4)

Analite (μg/m^3)	S_1			S_2			S_3			S_4		
	m ± s	min	max	m ± s	min	max	m ± s	min	max	m ± s	min	max
Benzene	2.3 ± 1.0	0.51	4.40	2.0 ± 0.9	0.68	4.08	2.1 ± 0.9	0.41	4.05	1.8 ± 0.8	0.39	3.81
Toluene	9.5 ± 5.8	0.63	25.4	8.0 ± 5.0	0.85	24.8	8.3 ± 4.1	1.59	18.9	8.4 ± 5.0	0.28	22.5
Ethylbenzene	2.6 ± 1.2	0.29	5.15	2.1 ± 1.0	0.30	4.46	2.4 ± 1.2	0.29	6.72	2.2 ± 1.2	0.19	6.10
m,p-Xylene	12.6 ± 6.3	1.13	26.6	10.5 ± 5.2	1.14	23.5	11.5 ± 6.0	1.10	29.3	10.8 ± 5.6	0.70	27.0
o-Xylene	3.2 ± 1.5	0.40	6.59	2.6 ± 1.2	0.38	5.68	2.9 ± 1.5	0.3	6.8	2.7 ± 1.3	0.23	6.29
n-Octane	1.6 ± 0.9	0.15	3.72	1.5 ± 0.8	0.14	3.31	1.5 ± 0.8	0.17	3.05	1.5 ± 0.7	0.08	2.99
Nonane	1.4 ± 0.7	0.15	2.97	1.3 ± 0.6	0.13	2.81	1.1 ± 0.6	0.12	2.37	1.1 ± 0.5	0.08	2.16

m = mean; s = standard deviation; min = minimum concentration; max = maximum concentration; s1 = sampling on February-March 2021, s2 = sampling on March-April 2021, s3 = sampling on February-March 2022 and s4 = sampling on March-April 2022

Table 3 BTEX concentrations in different cities worldwide (μg.m^{-3}) compared from those obtained in this study. Values adapted from Kerchich & Kerbachi [41]*

City	Benzene	Toluene	Ethylbenzene	m,p-Xylene	o-Xylene
Argiel	1.94	4.57	1.20	1.07	0.55
Rome	33.5	99.7	17.6	54.1	25.1
Yokohama	1.7–3.7	4.7–34.3	0.5–3.8	1.0–2.0	0.1–0.8
London	2.7	7.2	1.4	3.7	2.9
Mumbai	13.7	11.1	0.4	1.3	2.2
São Paulo	4.6	44.8	13.3	26.1	6.9
La Coruña	3.4	23.6	3.3	5.1	2.7
Toulouse	2	6.6	NA	1.2	3.7
Pamplona	2.8	13.3	2.15	6.01**	NA
Seoul	3.2	24.5	3.0	10.0	3.5
Hong Kong	30.51	200.82	15.07	15.67**	N.A.
Cali (This study)	**2.1**	**8.6**	**2.3**	**11.4**	**2.8**

*Data collected between 1997 to 2012. ** (o + m + p)-xylenes

(ranging between 2 and 5) [41–43 Given that the air quality standard in Colombia only establishes a threshold limit for benzene (5 μg/m3 annual average) [3] and considering the global standard values ranging from 1.3 to 10 μg/m3 (annual averages) [44], the measured values fall within compliance with Colombian legislation and are within the concentration range of the majority of legislations worldwide.

The correlation graph (Fig. 4) reveals that toluene has the lowest correlations with all VOCs, particularly benzene, while displaying a high correlation with the rest of the VOCs. Previous research investigating the correlation between these compounds has also noted that high correlations between VOCs suggest similar emission sources, particularly mobile sources [41–43], as observed for benzene, xylenes, ethylbenzene, n-octane, and nonane. The low correlation between benzene and toluene (BT) can be explained by the (B/T) ratio, which can be considered an indicator of vehicle emissions. In this study, B/T ratio values between 4 and 5 were obtained, closely resembling ratios reported in other works [41–43]. However, upon detailed analysis of the B/T ratio, it was observed that for the highest concentrations of both pollutants, the ratio increases. Kerchich and Kerbachi (2012) suggested that this phenomenon might be attributed to the presence of point emission sources of toluene, since the (BT) correlation is high in road corridors [40].

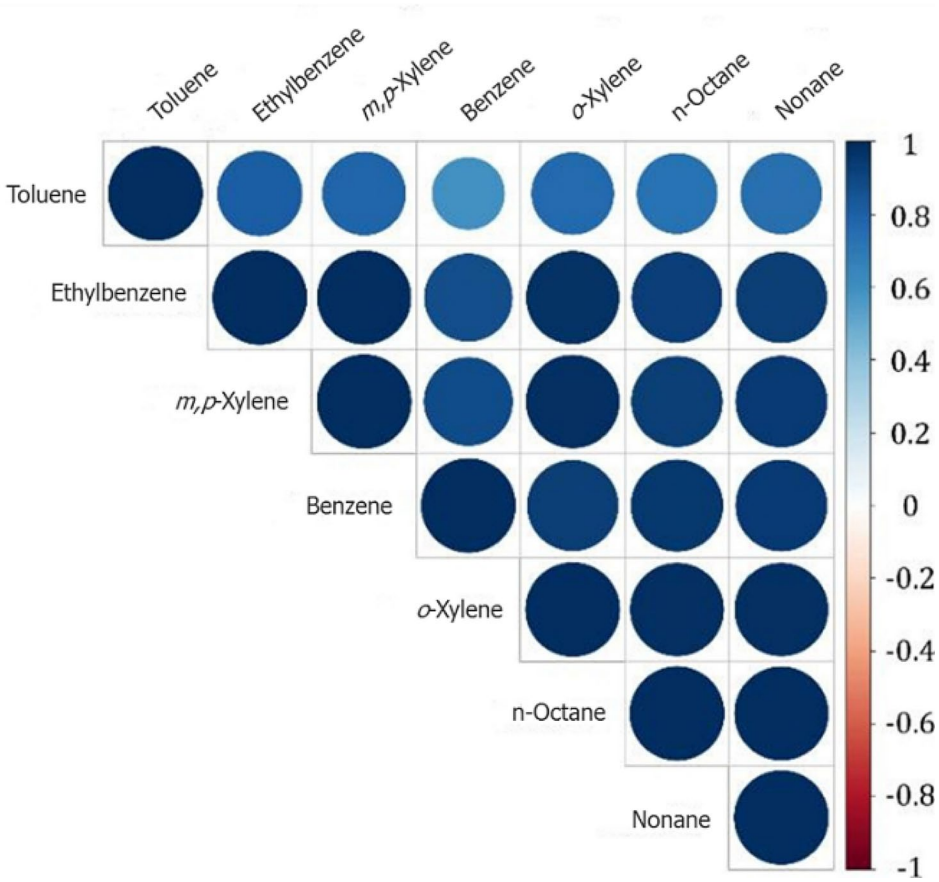

Fig. 4 Pair-wise correlation analysis for the 7 VOCs monitored using 'corrplot' package. In the right side of the correlogram, the legend color shows the correlation coefficients and the corresponding colors, blue for positive correlations and red for negative correlations. All samples showed positive correlation with each other, and the size of the circle were proportional to correlation coefficients

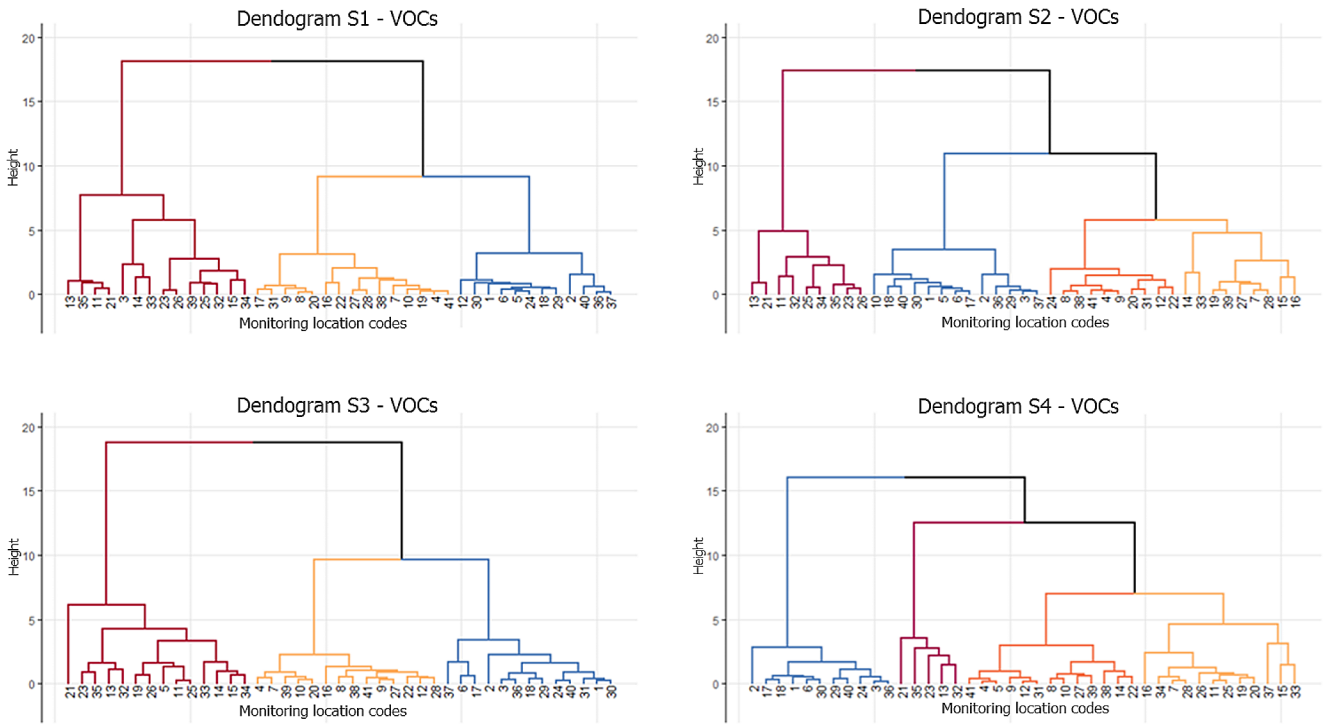

Fig. 5 Mean similarity dendrograms for the pollutants assemblages (VOCs) of the four sampling periods (s1, s2, s3 and s4). Each group formed cluster is denoted by the red, blue, orange and yellow colors, which agglomerate pollutants according to the characteristics of the sampling place

3.2 Cluster Analysis

The cluster analysis was employed to characterize monitoring locations based on the concentrations of VOCs found. This method involves a global-type multivariate classification considering all measured variables [45].

Figure 5 displays the dendrograms resulting from the cluster analysis for each sampling (s1, s2, s3, and s4). The graphic illustrates distinct clustering patterns of contaminants for s1-s3 (3 clusters) and s2-s4 (4 clusters), which aligns with the expectation as s1-s3 and s2-s4 represent samplings from the same period in different years. The resulting clusters in the four periods are as follows, categorized by sampling sites:

a) *Cluster H (Red)*: (13) Chapinero, (21) C44K1, (23) Av3 C44, (32) CC Caleño, (35) Luna. These locations are characterized by high vehicular traffic, particularly emissions from two-stroke motorcycle engines [44].
b) *Cluster L (Blue)*: (1) Icesi, (2) Pance Dagma, (6) ET Dagma, (18) Floralia, (29) Aguacatal, (36) Univalle Dagma, (40) Cañaveralejo Dagma. These places correspond to residential areas close to low-traffic roads or spacious areas with better airflow.
c) *Clusters M and ML (Orange and Yellow respectively)*: (4) MI Urrutia, (7) Calipso, (8) Mojica, (9) Compartir Dagma, (16) Maizena, (20) Ciudad Cali, (22) Bolivariano, (27) Autop 44, (28) Club Noel, (38) Premier, (41) P Banderas. These clusters comprise locations near high-traffic areas with significant mobility and high motorcycle circulation. Although individual motorcycles have low VOC emissions, their sheer number results in high VOC concentrations.

Table 4 presents the average VOC concentrations for each cluster group during the four campaigns. ANOVA analysis indicates no statistical difference among all the cluster groups (p-value (0.58) > 0.05). The model's sensitivity to variations in pollutant concentrations can cause sampling points to shift from one cluster to another. Certain cases are highlighted, such as calicanto, which responds to an increase in toluene concentrations (likely due to residential buildings), and terron, where low VOCs are in similar proportions to those with high pollutant levels. For the latter, a noticeable transition occurs from the marked cluster L to cluster H, representing high-traffic areas. For the remaining locations, variations are attributed to proximity to road corridors, resulting in smoother transitions.

Table 4 Average VOC concentrations ($\mu g/m^3$) for each cluster group during the four campaigns

Sampling campaign	Period time	Clusters			
		H	L	M	ML
S1	Feb – Mar (2021)	2,23	5,27	8,29	N.A.
S2	Mar – Abr (2021)	1,95	6,86	3,53	5,49
S3	Feb – Mar (2022)	6,60	3,87	1,81	N.A.
S4	Mar – Abr (2022)	1,81	3,14	7,70	4,69

3.3 Principal Components Analysis (PCA)

The aim of this technique was to condense the dimensionality of the original variables into a smaller number of elements, known as principal components (PCs), to facilitate data interpretation while preserving the contained information. A scree plot was initially constructed, a tool to determine the number of principal components based on the percentage of variance explained and the abrupt change in the slope of the curve [12, 37].

This change becomes apparent from the second PC, as depicted in Fig. 6a. The total variance explained by the first two PCs amounts to over 95% of the total variance across all samples. For these reasons, it was concluded that these two PCs sufficiently explain the original set of variables [36, 37].

Figure 6b displays correlation circles, representing each variable's contribution to each PC [12, 36, 37]. These graphs illustrate the qualitative and semi-quantitative contribution of each variable to PCs. Notably, all compounds, except toluene, primarily contribute to PC1, aligning with its horizontal axis. Moreover, the small angle between these compounds indicates their high correlation [12], consistent with the results of the correlation analysis. Concerning PC2, toluene emerges as the major contributor. It's worth noting the behavior of benzene, showing a negative association with PC2 and being nearly orthogonal to toluene, signifying the low correlation between these contaminants [12, 37]. Despite toluene's positive contribution to PC1, its greatest impact is on PC2. The strong correlation and contribution of all VOCs to PC1 suggest a common emission source, aligning with previous reports by Laowagul et al. (2008) and Bruno et al. (2001) [43, 45]. In contrast, toluene, being the primary contributor to PC2, is linked to specific emission sources of this pollutant [47], such as paints and dyes.

Similarly, monitoring points can be depicted as a function of the variables in a biplot [12, 36], shown in Fig. 7. To enhance comprehension, colors corresponding to the clusters obtained from CA have been utilized. In general, the distribution of points with respect to PC1 for each sampling period suggests mobile sources, with limited instances related to point sources (PC2). Notably, the previously obtained clusters are mainly conserved concerning PC1. The cluster L (blue) exhibits low PC1 influence, transitioning to

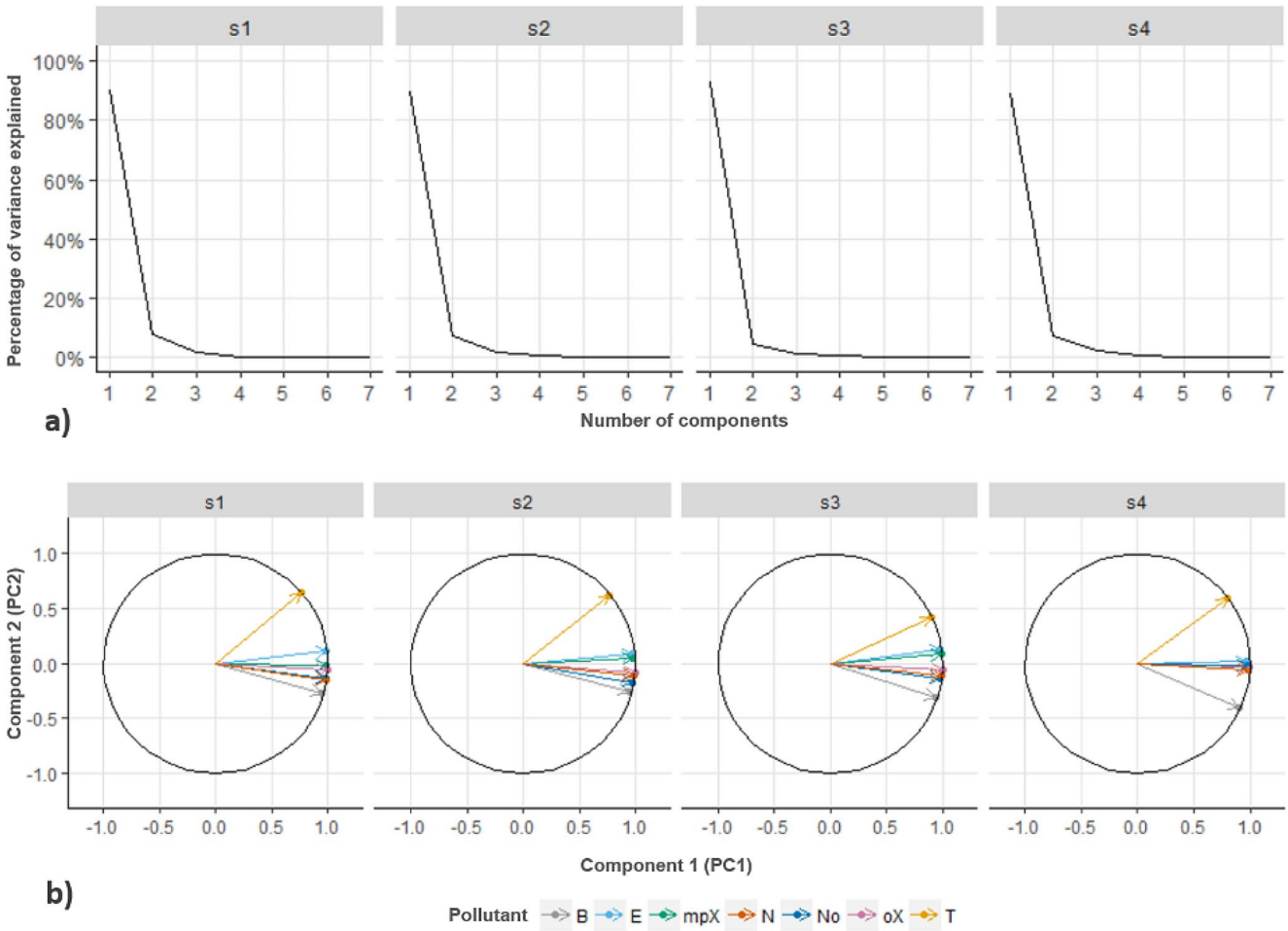

Fig. 6 (a) Scree plots of the variance explained and the corresponding principal components (PC) number for the four sampling periods of VOCs, (b) Correlation circles plots of the two Principal Components (PC1 and PC2) where the qualitative and semi-quantitative contribution of each pollutant (denoted by the color in the legend) is observed

clusters ML (yellow), M (orange), and finally H (red) as the PC1 effect intensifies. This indicates that the mobile sources indicator (PC1) is primarily responsible for organizing the clusters.

Additionally, sampling sites 3 and 37 (Calicanto and Madrigal respectively) demonstrate a low PC1 contribution (indicating a minor influence of mobile sources in minor road areas and the residential sector) but a high PC2 contribution, likely due to point sources of toluene stemming from construction activities (paints, solvents, etc.) in both zones. Likewise, places like Benjamín Herrera and ERA Dagma (sampling sites 14 and 33) situated in industrial sectors and near automotive workshops exhibit a significant PC2 effect despite their proximity to high-traffic roads. Finally, it is evident that the concentrations of VOCs in cluster H places predominantly stem from mobile sources.

3.4 Spatial Characterization

Figures 8 and 9 depict spatial interpolation maps for PC1 (a mobile sources indicator) during the s1 sampling period (February – March 2021), generated using Inverse Distance Weighting (IDW) and ordinary Kriging methodologies, respectively. The spatial interpolation of PC1 of the remaining periods (s2 period (March – April 2021), s3 period (February – March 2022), s4 period (March – April 2022)), generated by IDW and Kriging methods can be found in figures S1 to S6, respectively, available in online resources. Both IDW and Kriging are spatial interpolation methods used to estimate values between data points in unsampled geographic spaces. While IDW assigns weights to data points based on their distances to the point of estimation, Kriging employs a statistical model that considers spatial correlation structures among the points. It encompasses both overall trend and spatial correlation, making both models complementary [48]. For instance, in an epidemiological

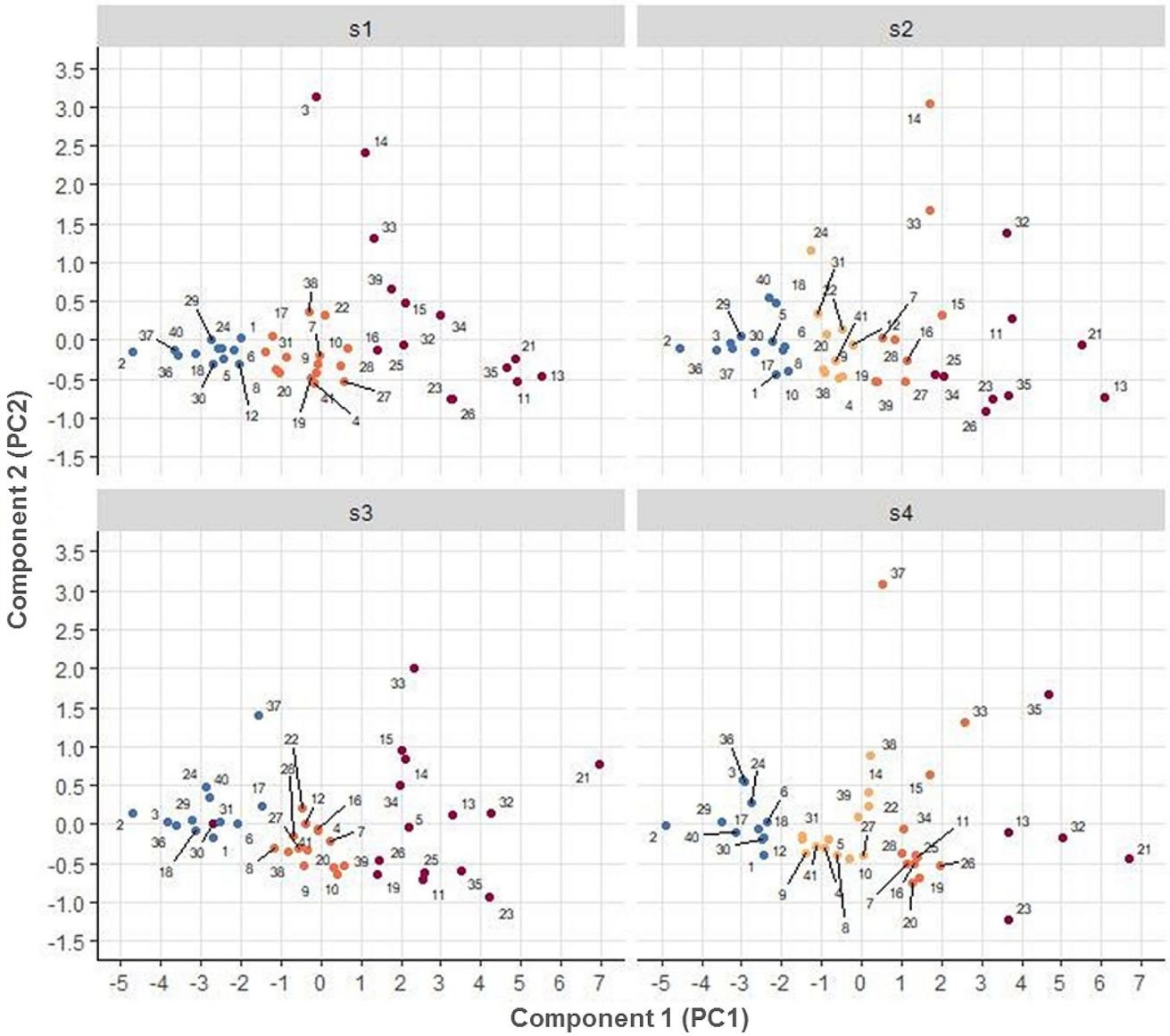

Fig. 7 Principal component biplot (PC1 and PC2) of the sampling sites and pollutants in the air of Santiago de Cali. The numbers correspond to the sampling site codes listed in Table 1

malaria study, IDW highlighted areas with the highest incidence of cases, while Kriging provided an estimate of the overall impact in a city region [20].

In both cases, the spatial behavior of PC1 is predominantly centered in the downtown area where streets are narrow, and traffic is slow. This situation hampers air mass recirculation, consequently delaying the dilution of VOCs in the atmosphere. The phenomenon also extends to the north and east of the city, corresponding to areas of high population density. These population groups primarily belong to a low socio-economic background [33], often relying on motorcycles as their mode of transportation (a significant source of VOCs) [46, 48, 50].

The effects of PC1 diminish in the western (hillside sector) and southern directions, characterized by less dense road corridors and greater amplitude, allowing for better circulation of VOCs. Temporally, few differences are observed between each of the sampling periods. In the case of PC1, both methods identify specific events with a high vehicular impact (as observed in the Siloé area), as well as the broader affected region (center-northern-eastern area). It's important to note that due to the low density of sampling points, the resolution of the Kriging method is affected, as seen in the northern sector (Flora, Chipichape, and Floralia), where PC1 levels are low but the method overestimates its effect.

Exploiting the properties of IDW and the characteristics of PC2, spatial interpolation of this component was

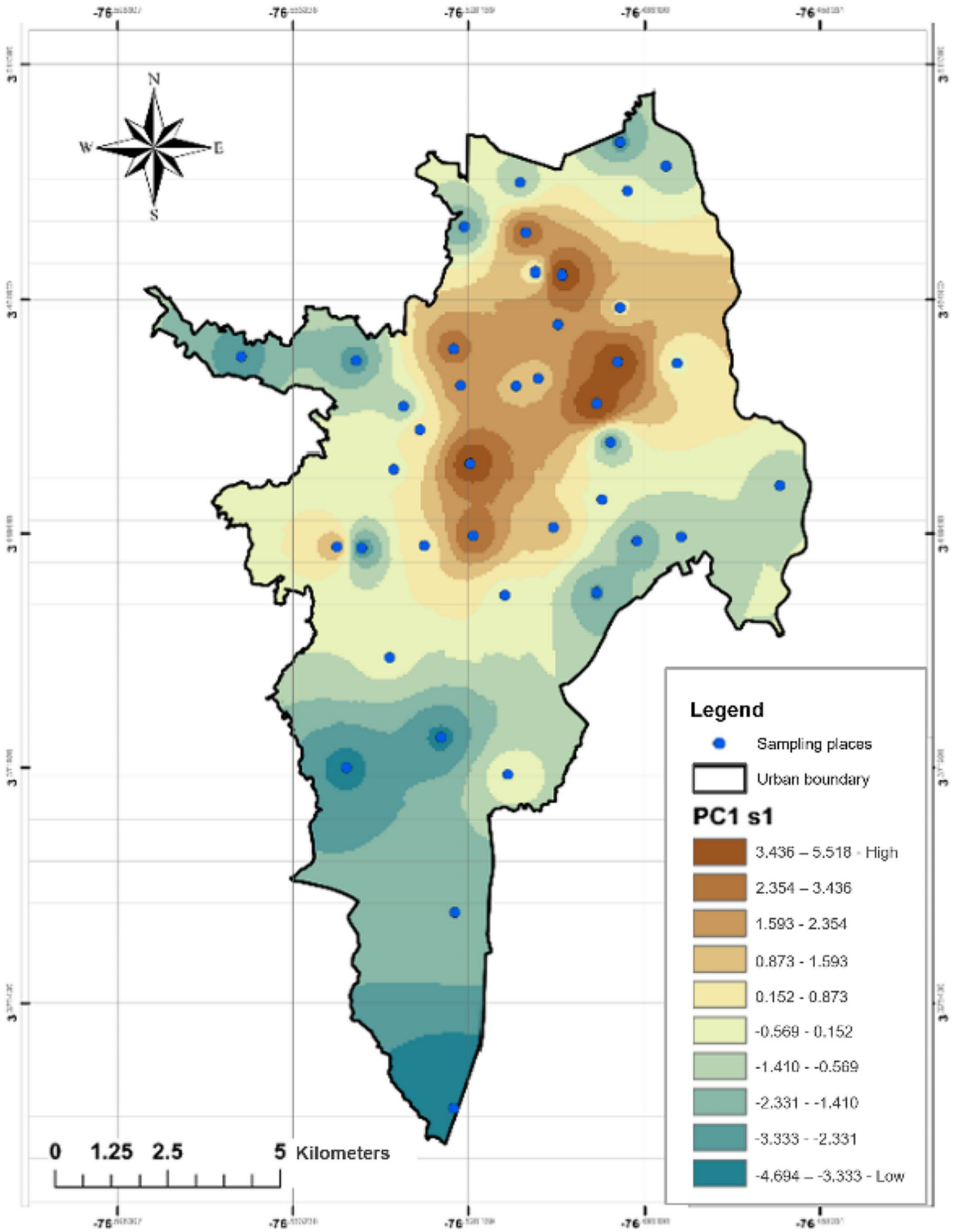

Fig. 8 Inverse distance weighting (IDW) map for the spatial distribution of PC1 in the city of Santiago de Cali for s1 sampling campaign (February to March 2021)

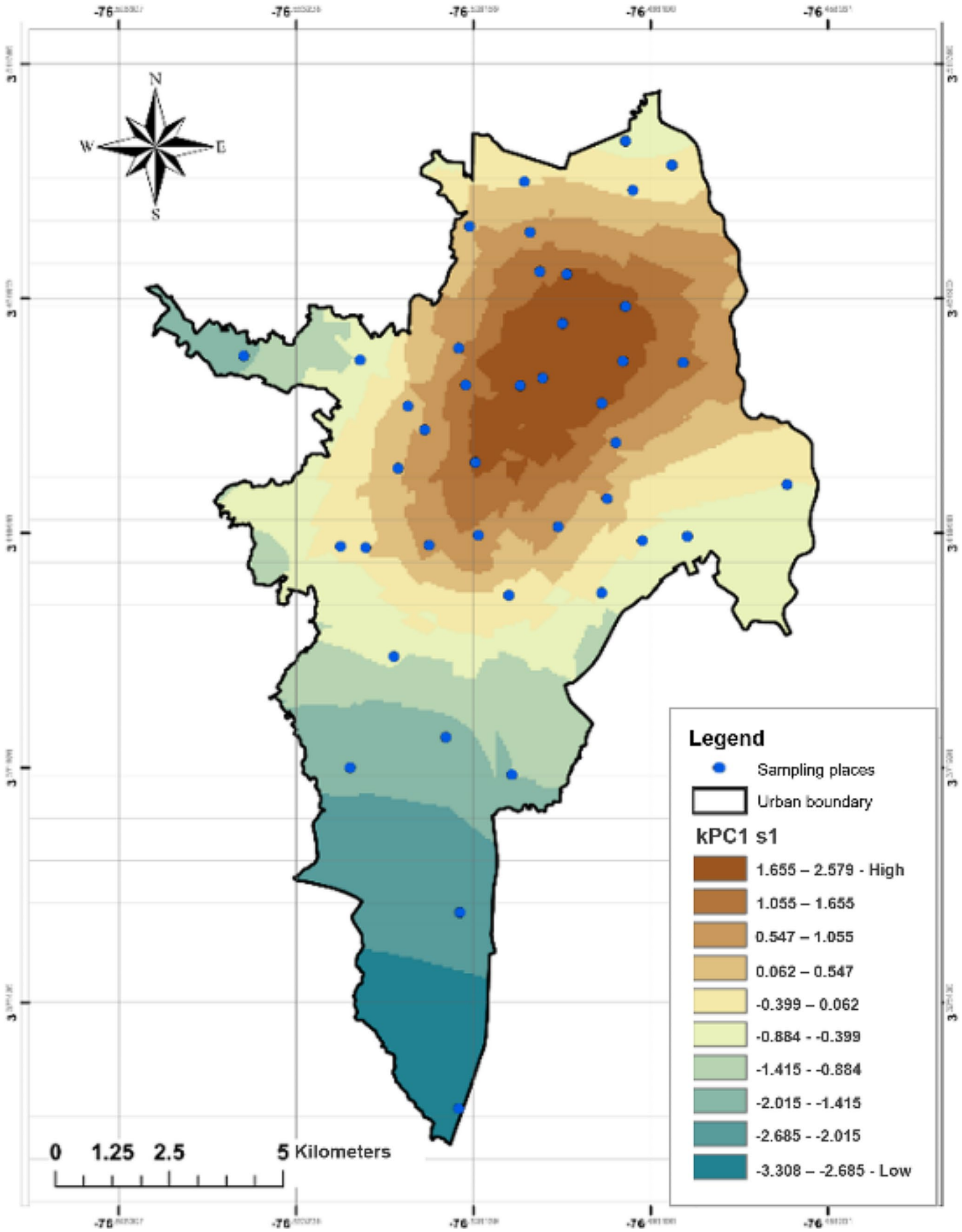

Fig. 9 Spatial interpolation map prepared by kriging method of PC1 in the city of Santiago de Cali for s1 sampling campaign (February to March 2021)

conducted (Fig. 10). As depicted in the Principal Component Biplot (Fig. 7), PC2's incidence on most points throughout the city is very low, thus resulting in low spatial predominance values. However, specific sources of toluene emission are emphasized. Notably, the Calicanto site stands out in samplings s1, s3, and s4, as does the Chipichape area in samplings s3 and s4, both of which are associated with ongoing construction of residential complexes. Conversely, in all samplings, downtown and its surrounding areas predominate as sources. These areas correspond to industrial zones with automotive activities, tanneries, paper industries, textiles, and common usage of toluene [33]. IDW maps for the spatial interpolation of PC2 of the remaining periods (s2 period (March – April 2021), s3 period (February – March 2022), s4 period (March – April 2022)) can be found in figures S7 to S9, respectively, available in online resources. Both for PC1 and PC2, the employed spatial interpolation methods effectively elucidate the phenomena, enabling an estimate of the impact of pollutants in non-monitored areas.

3.5 Carcinogenic Risk Assessment

Given the detrimental impact of VOCs identified in the city's air and acknowledging benzene as a known carcinogen [4, 49] and ethylbenzene as a potential carcinogen, it was imperative to evaluate the inhalation-based carcinogenic risk in the population, aligning with the guidelines set forth by the Environmental Protection Agency of the United States (EPA) [31].

The scenario considered for this assessment was geared towards adults. Deliberations on exposure time (ET) led to an estimation of 4 h, encompassing the representative duration an individual typically spends outdoors, either commuting to work or engaging in leisure activities. A standard average time (AT) of 70 years was established, with an exposure period of 35 years. Other exposure factors were adopted in accordance with EPA recommendations [30, 51]. The carcinogenic risk results pertaining to benzene and ethylbenzene, calculated utilizing Eq. 9, are succinctly presented in Table 5.

The EPA has set carcinogenic risk reference values greater than 1×10^{-6} as the acceptable threshold. Based on this criterion, it is evident that none of the ethylbenzene concentrations pose a carcinogenic risk under the stipulated exposure conditions. Conversely, for benzene, across samplings s1 to s3, at least half of the measured concentrations surpass the 1×10^{-6} median, indicating a potential cancer risk.

Moreover, in sampling s4, the average concentration exceeds the permissible risk level. The calculation of the total carcinogenic risk (Eq. 9) was performed, assuming a lack of additive effects between the health impacts of both pollutants, indicating no synergy or antagonism [52]. Consequently, cancer risk maps were generated for all four samplings, as illustrated in Fig. 11.

The spatial distribution of the risk aligns predominantly with the behavior of PC1, attributable to the substantial contribution of benzene and ethylbenzene concentrations in this principal component. The prevailing trend underscores a heightened carcinogenic risk in the downtown area and, to a lesser extent, in the northern and eastern regions of the city during all sampling periods. In contrast, both the hillside areas and the southern region of the city exhibit negligible risks to human health.

4 Conclusions

This study successfully determined the concentrations of volatile organic compounds (VOCs) in the air of Santiago de Cali using passive diffusion samplers. Notably, none of the monitored compounds exceeded the national air quality standards. The observed pollutant concentrations were comparable to those found in other major cities. The correlation analysis provided valuable insights, revealing an unusual behavior in toluene concentrations compared to the other VOCs, which exhibited strong correlations. Utilizing Cluster Analysis (CA), monitoring sites were effectively grouped based on similarities, demonstrating relative homogeneity among different sampling periods. Principal Component Analysis (PCA) further summarized the original 7 variables into 2 principal components. The first component suggested a potential indicator of mobile emissions, while the second highlighted specific toluene emissions. The synergy between CA and PCA was evident, emphasizing the dependency of conglomerates on the behavior of PC1.

Spatial interpolation through Inverse Distance Weighting (IDW) and kriging facilitated the representation and prediction of VOC concentrations in non-monitored areas. PC1 analysis identified critical areas, including the downtown, east, and north of the city, as regions of heightened concern. Additionally, PC2 representation by IDW confirmed punctual emissions primarily associated with toluene. Furthermore, the carcinogenic risk assessment for benzene and ethylbenzene shed light on inhalation-induced carcinogenic risk within the proposed scenario. The generated risk maps proved to be invaluable tools for identifying critical areas concerning human health, notably the downtown and northeast sectors of the city.

In conclusion, this comprehensive study employing chemometric techniques and spatial analysis provided a holistic understanding of VOC concentrations, their sources, and associated health risks. The findings contribute significantly to environmental monitoring efforts and lay a foundation for

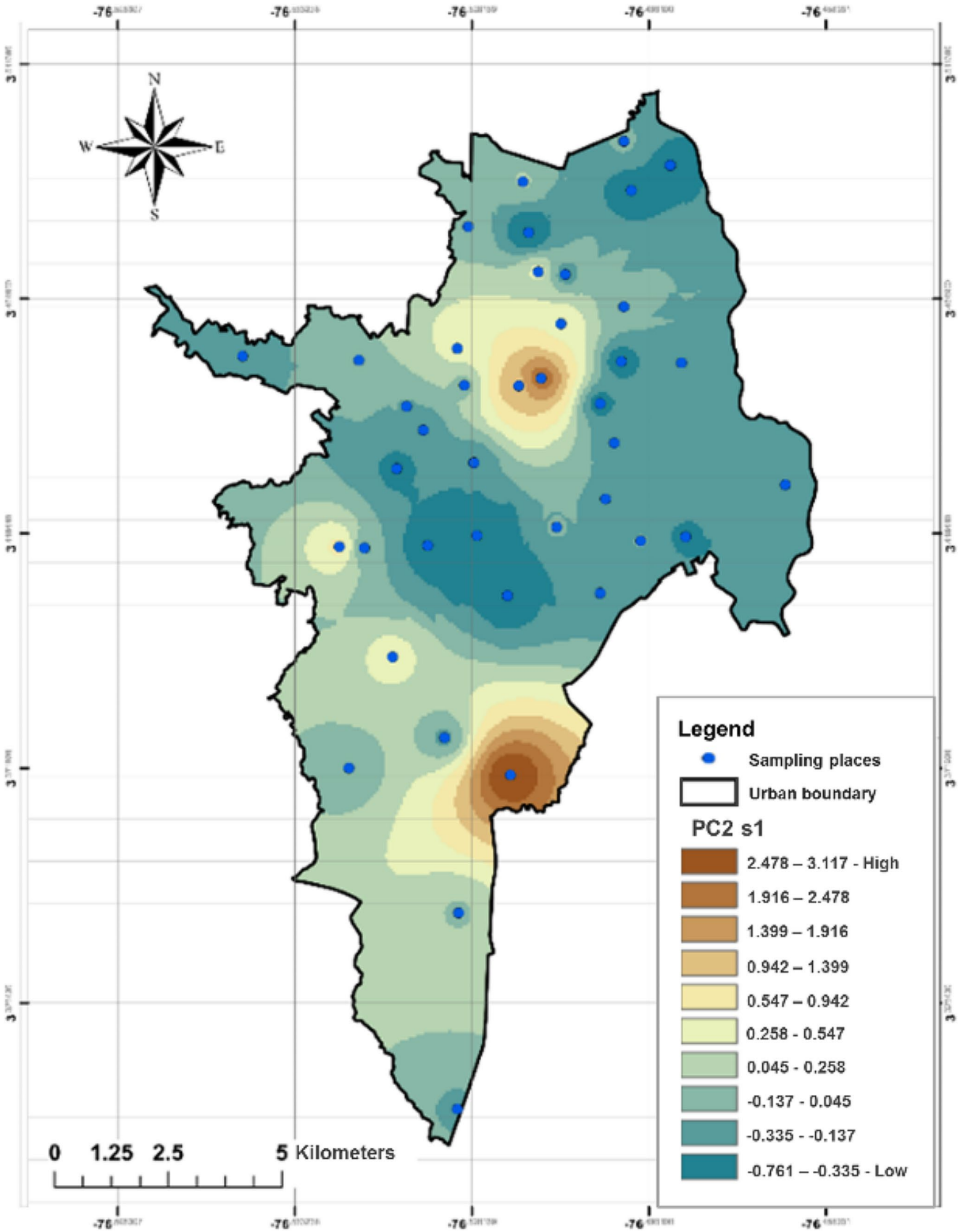

Fig. 10 Inverse distance weighting (IDW) map for the spatial distribution of PC2 in the city of Santiago de Cali for s1 sampling campaign (February to March 2021)

Table 5 Summary of cancer risk for adults from inhalation benzene and ethylbenzene

	Cancer risk (x 10^{-6})							
	S1		S2		S3		S^4	
	Benzene	Ethylbenzene	Benzene	Ethylbenzene	Benzene	Ethylbenzene	Benzene	Ethylbenzene
Minimum	0.29	0.02	0.38	0.02	0.23	0.02	0.22	0.02
Maximum	2.47*	0.41	2.29*	0.35	2.27*	0.53	2.14*	0.48
Average	1.29*	0.20	1.14*	0.17	1.16*	0.19	1.01*	0.18
Median	1.20*	0.19	1.10*	0.15	1.18*	0.18	0.94	0.17

*Values above 1×10^{-6} are considering with cancer risk

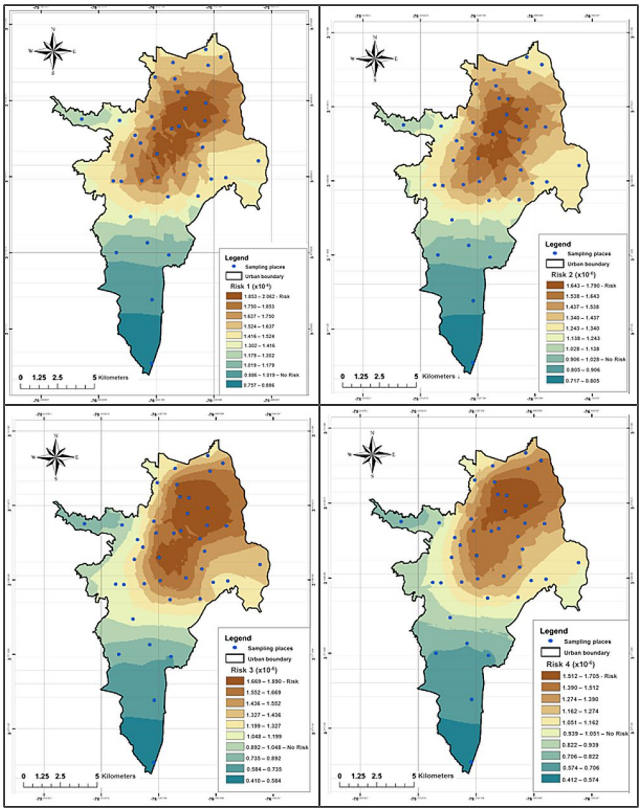

Fig. 11 Spatial interpolation map prepared by kriging method for cancer risk caused by benzene and ethylbenzene in the city of Santiago de Cali in each of the four sampled periods: Risk 1 = s1, Risk 2 = s2, Risk 3 = s3 and Risk 4 = s4

potential targeted interventions to mitigate air pollution in Santiago de Cali.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10666-024-09969-7.

Acknowledgements The authors thank Engineer Gisela Arizabaleta and Mr. Jairo Copete from the Air Quality group of the *Departmento Administrativo de Gestión Ambiental* (DAGMA), for allowing the use of the Santiago de Cali Air Quality Surveillance System stations.

Author Contributions Rubén Albeiro Sánchez-Andica: Designed and Supervised analytical measurements also wrote and revised the manuscript. Wilson Rafael Salas-Chávez: Monitoring coordination, performed analytical measurements and wrote and reviewed manuscript. Martha Isabel Páez-Melo: Searched for resources, supervised the experiments and reviewed and approved the manuscript.

Funding This work was funded by *Corporación Autónoma del Valle* (CVC) and *Universidad del Valle* with 059 and 077 Grant numbers.

Data Availability The authors confirm that the data supporting the findings of this study are available within the article.

Declarations

Ethical Approval Not aplicable.

WRS-C Monitoring coordination, performed analytical measurements and wrote and reviewed manuscript.

MIP-M Searched for resources, supervised the experiments and reviewed and approved the manuscript.

Competing Interests The authors declare no competing interests.

References

1. Organización Mundial de la Salud. OMS (2022). diciembre 19), Contaminación del aire ambiente (exterior) y salud http://www.who.int/mediacentre/factsheets/fs313/es/ (accessed Dec 26, 2022).
2. Colman Lerner, J. E., Elordi, M. L., Orte, M. A., Giuliani, D., de los Angeles Gutierrez, M., Sanchez, E., & Porta, A. A. (2018). Exposure and risk analysis to particulate matter, metals, and polycyclic aromatic hydrocarbon at different workplaces in Argentina. *Environmental Science and Pollution Research*, *25*, 8487–8496. https://doi.org/10.1007/s11356-017-1101-0.
3. Ministerio de Ambiente, M. (2010). *Vivienda Y Desarrollo Territorial Resolución Número* 610 8.
4. Agency for Toxic Substances and Disease Registry (ATSDR). (2007). *Toxicological profile for Benzene; Atlanta, GA, Agency for toxic substances and Disease Registry*. Division of Toxicology.
5. Agency for Toxic Substances and Disease; Registry (ATSDR). (2015). *Toxicological Profile for Toluene (draft for public comment); Atlanta, GA, Agency for toxic substances and Disease Registry*. Division of Toxicology.
6. Agency for Toxic Substances and Disease; Registry (ATSDR). (2010). *Toxicological Profile for Ethylbenzene; Atlanta, GA, Agency for toxic substances and Disease Registry*. Division of Toxicology.
7. Agency for Toxic Substances and Disease; Registry (ATSDR). (2007). *Toxicological Profile for Xylenes, Atlanta, GA, Agency for toxic substances and Disease Registry*. Division of Toxicology.

8. Inobeme, A., Nayak, V., Mathew, T. J., Okonkwo, S., Ekwoba, L., Ajai, A. I., … Singh, K.R. (2022). Chemometric approach in environmental pollution analysis: A critical review. *Journal of Environmental Management*, *309*, 114653. https://doi.org/10.1016/j.jenvman.2022.114653.
9. Liu, X., & Wania, F. (2014). Cluster analysis of passive air sampling data based on the relative composition of persistent organic pollutants. *Environmental Science: Processes & Impacts*, *16*(3), 453–463. https://doi.org/10.1039/C3EM00605K.
10. Ossai, C. J., Iwegbue, C. M., Tesi, G. O., Olisah, C., Egobueze, F. E., Nwajei, G. E., & Martincigh, B. S. (2021). Distribution, sources and exposure risk of polycyclic aromatic hydrocarbons in soils, and indoor and outdoor dust from Port Harcourt City, Nigeria. *Environmental Science: Processes & Impacts*, *23*(9), 1328–1350. https://doi.org/10.1039/D1EM00094B.
11. Kovač-Andrić, E., & Arh, G. (2016). Atmospheric VOCs measurement in nature reserve Kopački Rit, Croatia. *Air Quality Atmosphere & Health*, *9*, 681–686. https://doi.org/10.1007/s11869-015-0374-z.
12. Zhang, J., Zhang, L. Y., Du, M., Zhang, W., Huang, X., Zhang, Y. Q., … Xiao, H. (2016). Indentifying the major air pollutants base on factor and cluster analysis, a case study in 74 Chinese cities. *Atmospheric Environment*, *144*, 37–46. https://doi.org/10.1016/j.atmosenv.2016.08.066.
13. Pérez-Arribas, L. V., León-González, M. E., & Rosales-Conrado, N. (2017). Learning principal component analysis by using data from air quality networks. *Journal of Chemical Education*, *94*(4), 458–464. https://doi.org/10.1021/acs.jchemed.6b00550.
14. Arhami, M., Hosseini, V., Shahne, M. Z., Bigdeli, M., Lai, A., & Schauer, J. J. (2017). Seasonal trends, chemical speciation and source apportionment of fine PM in Tehran. *Atmospheric Environment*, *153*, 70–82. https://doi.org/10.1016/j.atmosenv.2016.12.046.
15. Singh, D., Kumar, A., Kumar, K., Singh, B., Mina, U., Singh, B. B., & Jain, V. K. (2016). Statistical modeling of O3, NOx, CO, PM2. 5, VOCs and noise levels in commercial complex and associated health risk assessment in an academic institution. *Science of the Total Environment*, *572*, 586–594. https://doi.org/10.1016/j.scitotenv.2016.08.086.
16. Rueda-Holgado, F., Calvo-Blázquez, L., Cereceda-Balic, F., & Pinilla-Gil, E. (2016). Temporal and spatial variation of trace elements in atmospheric deposition around the industrial area of Puchuncaví-Ventanas (Chile) and its influence on exceedances of lead and cadmium critical loads in soils. *Chemosphere*, *144*, 1788–1796. https://doi.org/10.1016/j.chemosphere.2015.10.079.
17. Riley, E. A., Schaal, L., Sasakura, M., Crampton, R., Gould, T. R., Hartin, K., …Yost, M. G. (2016). Correlations between short-term mobile monitoring and long-term passive sampler measurements of traffic-related air pollution. *Atmospheric Environment*, *132*, 229–239. https://doi.org/10.1016/j.atmosenv.2016.03.001.
18. Sakai, N., Yamamoto, S., Matsui, Y., Khan, M. F., Latif, M. T., Mohd, M. A., & Yoneda, M. (2017). Characterization and source profiling of volatile organic compounds in indoor air of private residences in Selangor State, Malaysia. *Science of the Total Environment*, *586*, 1279–1286. https://doi.org/10.1016/j.scitotenv.2017.02.139.
19. Gibergans Bàguena, J., Sala, H., C., & Jarauta Bragulat, E. (2020). The quality of urban air in Barcelona: A new approach applying compositional data analysis methods. *Emerging Science Journal*, *4*(2), 113–121. https://doi.org/10.28991/esj-2020-01215.
20. Habibi, R., Alesheikh, A. A., Mohammadinia, A., & Sharif, M. (2017). An assessment of spatial pattern characterization of air pollution: A case study of CO and PM2. 5 in Tehran, Iran. *ISPRS International Journal of Geo-information*, *6*(9), 270. https://doi.org/10.3390/ijgi6090270.
21. Cleary, E., Hetzel, M. W., Siba, P. M., Lau, C. L., & Clements, A. C. (2021). Spatial prediction of malaria prevalence in Papua New Guinea: A comparison of bayesian decision network and multivariate regression modelling approaches for improved accuracy in prevalence prediction. *Malaria Journal*, *20*, 1–16. https://doi.org/10.1186/s12936-021-03804-0.
22. Xu, W., Riley, E. A., Austin, E., Sasakura, M., Schaal, L., Gould, T. R., & Vedal, S. (2017). Use of mobile and passive badge air monitoring data for NOX and ozone air pollution spatial exposure prediction models. *Journal of Exposure Science & Environmental Epidemiology*, *27*(2), 184–192. https://doi.org/10.1038/jes.2016.9.
23. Sau, T. K., Truong, N. X., Hanh, T. T. T., Le Hung, B., Thang, N. D., & Le Anh, L., T (2021). Ambient air monitoring around the dioxin remediation site in Da Nang, Vietnam, using passive air samplers. *Environmental Monitoring and Assessment*, *193*, 1–14. https://doi.org/10.1007/s10661-021-09223-7
24. [24] Lin, H. C., Hung, P. H., Hsieh, Y. Y., Lai, T. J., Hsu, H. T., Chung, M. C., & Chung, C. J. (2022). Long-term exposure to air pollutants and increased risk of chronic kidney disease in a community-based population using a fuzzy logic inference model. *Clinical Kidney Journal*, *15*(10), 1872–1880. https://doi.org/10.1093/ckj/sfac114.
25. Requia, W. J., Koutrakis, P., Roig, H. L., Adams, M. D., & Santos, C. M. (2016). Association between vehicular emissions and cardiorespiratory disease risk in Brazil and its variation by spatial clustering of socio-economic factors. *Environmental Research*, *150*, 452–460. https://doi.org/10.1016/j.envres.2016.06.027.
26. Theophilo, C. Y. S., Ribeiro, A. P., Moreira, E. G., Aranha, S., Bollmann, H. A., Santos, C. J., & Ferreira, M. L. (2021). Biomonitoring as a nature-based solution to assess Atmospheric Pollution and impacts on Public Health. *Bulletin of Environmental Contamination and Toxicology*, *107*, 29–36. https://doi.org/10.1007/s00128-021-03205-8.
27. Ribeiro, M. C., Pinho, P., Branquinho, C., Llop, E., & Pereira, M. J. (2016). Geostatistical uncertainty of assessing air quality using high-spatial-resolution lichen data: A health study in the urban area of Sines, Portugal. *Science of the Total Environment*, *562*, 740–750. https://doi.org/10.1016/j.scitotenv.2016.04.081.
28. Hong, S. H., Shin, D. C., Lee, Y. J., Kim, S. H., & Lim, Y. W. (2017). Health risk assessment of volatile organic compounds in urban areas. *Human and Ecological Risk Assessment: An International Journal*, *23*(6), 1454–1465. https://doi.org/10.1080/10807039.2017.1325714.
29. Khoshakhlagh, A. H., Askari Majdabadi, M., Yazdanirad, S., & Carlsen, L. (2023). Health risk assessment of exposure to benzene, toluene, ethylbenzene, and xylene (BTEX) in a composite manufacturing plant: Monte-Carlo simulations. *Human and Ecological Risk Assessment: An International Journal*, 1–15. https://doi.org/10.1080/10807039.2023.2167193.
30. Abd Hamid, H. H., Jumah, N. S., Latif, M. T., & Kannan, N. (2017). BTEXs in indoor and outdoor air samples: Source apportionment and health risk assessment of benzene. *Journal of Environmental Science and Public Health*, *1*(1), 49–56. https://doi.org/10.26502/JESPH.005.
31. Dehghani, M., Mohammadpour, A., Abbasi, A., Rostami, I., Gharehchahi, E., Derakhshan, Z.,… Conti, G. O. (2022). Health risks of inhalation exposure to BTEX in a municipal wastewater treatment plant in Middle East city: Shiraz, Iran. *Environmental Research*, *204*, 112155. https://doi.org/10.1016/j.envres.2021.112155.
32. Wang, Q., Yan, S., Chang, C., Qu, C., Tian, Y., Song, J., & Guo, J. (2023). Occurrence, potential risk Assessment, and Source Apportionment of Polychlorinated Biphenyls in Water from Beiluo River. *Water*, *15*(3), 459. https://doi.org/10.3390/w15030459.

33. Bravo, L. E., García, L. S., Collazos, P., Carrascal, E., Grillo-Ardila, E. K., Millán, E., & Holguín, J. (2022). Epidemiología Del cáncer en cali, 60 años de experiencia. *Colombia Médica*, *53*(1). https://doi.org/10.25100/cm.v53i1.5050.
34. Instituto de, & Hidrología (2023). Meteorología y Estudios Ambientales (IDEAM), Precipitación y temperatura diaria-Principales ciudades, Retrieved September 21, from http://www.pronostico-syalertas.gov.co/informacion-diaria-de-precipitacion-y-temperatura-de-los-principiales-aeropuertos-del-pais.
35. European Committee for Standardization. Ambient air quality - Diffusive samplers for the determination of concentrations of gases and vapours - Requirements and test methods - Part 3: Guide to selection, use and maintenance (2004). EN 13528-3; CEN: Bruxelles.
36. Otto, M. (2016). *Chemometrics: Statistics and computer application in analytical chemistry*. Wiley.
37. Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101.
38. Webster, R. (2005). Barnett, V. Environmental Statistics: Methods and Applications. John Wiley & Sons, Chichester, 2004. xi + 293 pp.£ 55, hardback. ISBN 0-471-48971-9. https://doi.org/10.1111/j.1365-2389.2004.0694g.x.
39. Chiles, J. P., & Delfiner, P. (2009). *Geostatistics: Modeling spatial uncertainty* (Vol. 497). Wiley.
40. Kerchich, Y., & Kerbachi, R. (2012). Measurement of BTEX (benzene, toluene, ethybenzene, and xylene) levels at urban and semirural areas of Algiers City using passive air samplers. *Journal of the Air & Waste Management Association*, *62*(12), 1370–1379. https://doi.org/10.1080/10962247.2012.712606.
41. Kim, S. J., Kwon, H. O., Lee, M. I., Seo, Y., & Choi, S. D. (2019). Spatial and temporal variations of volatile organic compounds using passive air samplers in the multi-industrial city of Ulsan, Korea. *Environmental Science and Pollution Research*, *26*, 5831–5841. https://doi.org/10.1007/s11356-018-4032-5.
42. Groth, C. P., Huynh, T. B., Banerjee, S., Ramachandran, G., Stewart, P. A., Quick,H., … Stenzel, M. R. (2022). Linear relationships between total hydrocarbons and benzene, toluene, ethylbenzene,xylene, and n-hexane during the Deepwater Horizon response and clean-up. *Annals of work exposures and health*, *66*(Supplement_1), i71-i88. https://doi.org/10.1093/annweh/wxac018.
43. Laowagul, W., Garivait, H., Limpaseni, W., & Yoshizumi, K. (2008). Ambient air concentrations of benzene, toluene, ethylbenzene and xylene in Bangkok, Thailand during April-August in 2007. *Asian Journal of Atmospheric Environment*, *2*(1), 14–25. https://doi.org/10.5572/ajae.2008.2.1.014.
44. Sekar, A., Varghese, G. K., & Varma, M. R. (2019). Analysis of benzene air quality standards, monitoring methods and concentrations in indoor and outdoor environment. *Heliyon*, *5*(11). https://doi.org/10.1016/j.heliyon.2019.e02918.
45. Bruno, P., Caselli, M., De Gennaro, G., & Traini, A. (2001). Source apportionment of gaseous atmospheric pollutants by means of an absolute principal component scores (APCS) receptor model. *Fresenius' Journal of Analytical Chemistry*, *371*, 1119–1123. https://doi.org/10.1007/s002160101084.
46. Gómez, M. V. T., & Rendón, D. M. Q. (2015). Energy demand and vehicle emissions estimate in Aburra Valley from 2000 to 2010 using LEAP model. *Dyna*, *82*(189), 45–51.
47. Feng, Y., An, J., Tang, G., Zhang, Y., Wang, J., & Lv, H. (2022). Characteristics and sources of volatile Organic compounds in the Nanjing Industrial Area. *Atmosphere*, *13*(7), 1136. https://doi.org/10.3390/atmos13071136.
48. Dörter, M., Mağat-Türk, E., Döğeroğlu, T., Özden-Üzmez, Ö., Gaga, E. O., Karakaş, D., & Yenisoy-Karakaş, S. (2022). An assessment of spatial distribution and atmospheric concentrations of ozone, nitrogen dioxide, sulfur dioxide, benzene, toluene, ethylbenzene, and xylenes: Ozone formation potential and health risk estimation in Bolu city of Turkey. *Environmental Science and Pollution Research*, *29*(35), 53569–53583. https://doi.org/10.1007/s11356-022-19608-x.
49. Franco, J. F., Pacheco, J., Behrentz, E., & Belalcázar, L. C. (2015). Characterization and source identification of VOC species in Bogotá. *Colombia Atmósfera*, *28*(1), 1–11. https://doi.org/10.1016/S0187-6236(15)72155-7.
50. Londoño, J., Correa, M. A., & Palacio, C. A. (2011). Estimation of the emissions of Atmospheric pollutants from Mobile sources in the Urban Area of Envigado, Colombia. *Revista EIA*, *16*, 149–162.
51. USEPA - U.S. OSWER Directive 9200.1–120 Standard Default Exposure Factors (2014). Vol. 2004.
52. U.S. Environmental Protection Agency (EPA). EPA Guide (2003). No. May 1–129.