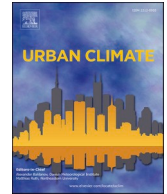




ELSEVIER

Contents lists available at ScienceDirect

## Urban Climate

journal homepage: [www.elsevier.com/locate/uclim](http://www.elsevier.com/locate/uclim)

# A reliable and easy-to-implement approach to estimate daily urban benzene levels

David Galán-Madruga<sup>a,\*</sup>, Parya Broomandi<sup>b</sup>, Jafet Cárdenas-Escudero<sup>c,d,1</sup>,  
J.L. Urraca<sup>d</sup>, Jorge O. Cáceres<sup>d</sup>

<sup>a</sup> National Reference Laboratory of Air Quality, National Centre for Environmental Health (CNSA), Carlos III Health Institute (ISCIII), Ctra, Majadahonda a Pozuelo, Madrid 28222, Spain

<sup>b</sup> Department of Civil and Environmental Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Kabanbay Batyr Ave. 53, Astana 010000, Kazakhstan

<sup>c</sup> Analytical Chemistry Department, FCNET, University of Panama, University City, University Mail, 3366, Panama, 4, Panama City, Panama

<sup>d</sup> Laser Chemistry Research Group, Department of Analytical Chemistry, Faculty of Chemistry, Complutense University of Madrid, Plaza de Ciencias 1, 28040 Madrid, Spain

## ARTICLE INFO

## Keywords:

Air quality  
Urban environment  
Urban pollution  
Estimating benzene  
Fixed station representativeness  
Legislative framework

## ABSTRACT

Assessing air benzene is mandatory in the European Union due to potentially harmful human effects. Current exposure data are lacking since there is a limited number of fixed stations. This research provides an easy approach to estimating daily urban benzene levels potentially applicable in other geographical areas. Madrid City was regarded as a case study. They were calculated mathematical models explaining the relationship between dependent and predictor inputs using an original 2019–2020 dataset ( $r$  from 0.881 to 0.903) and were tested using a 2021 dataset ( $r$  from 0.682 to 0.914). Similarly, the suggested mathematical expressions were satisfactorily implemented in another geographic region ( $r = 0.936$  and  $0.764$  for urban traffic and rural background sites, respectively). Furthermore, the recommended approach helps identify the most representative fixed benzene measuring locations without lost spatial information, given that a similar spatial distribution gradient was sustained between estimated and the most representative benzene levels ( $r = 0.959$ ), which is translated into acceptable quantitative differences ( $0.07 \mu\text{g}/\text{m}^3$ ,  $0.01 \mu\text{g}/\text{m}^3$  and  $5.01 \%$  for RMSE, MAE, and MAPE, respectively). Applying the proposed approach furnishes enriched benzene exposure data, thus decreasing uncertainty generated by the lack of actual information and complements European Legislation directives concerning air pollutants monitoring using air quality networks.

## 1. Introduction

Numerous investigation studies link harmful effects in human beings to air pollutants exposure (Madruga et al., 2019; Ghaffar-pasand et al., 2020; Broomandi et al., 2023). To minimize the potential adverse effects of air pollutants on human beings' health and the environment, the current European legislation sets air quality objectives, which should not be exceeded (Directive 2004/107/EC,

\* Corresponding author at: Department of Atmospheric Pollution, National Center for Environment Health, Health Institute Carlos III, Ctra, Majadahonda a Pozuelo km 2,2, 28222, Majadahonda, Madrid, Spain.

E-mail address: [david.galan@isciii.es](mailto:david.galan@isciii.es) (D. Galán-Madruga).

<sup>1</sup> Doctoral Fellow, FCNET Universidad de Panamá.

<https://doi.org/10.1016/j.uclim.2024.102234>

Received 1 August 2024; Received in revised form 22 November 2024; Accepted 28 November 2024

Available online 5 December 2024

2212-0955/© 2024 Published by Elsevier B.V.

2024; Directive 2008/50/EC, 2024). To comply with these objectives, the European Union urges Member States to monitor ambient air pollutant levels (Galán-Madruga et al., 2023). In this regard, air quality monitoring networks are implemented within each Member States' territory. These networks consist of fixed stations measuring legislated air pollutants (Galán-Madruga, 2023), such as nitrogen oxides, carbon monoxide, sulfur dioxide, ozone, airborne atmospheric particles (PM<sub>10</sub> and PM<sub>2.5</sub>, with an aerodynamic diameter lower than 10 and 2.5 μm, respectively), and benzene. The current legislation provides requirements to lay down the minimum number of fixed stations as a function of the number of inhabitants and pollutants (Annex V from Directive 2008/50/EC, 2024). Based on the previous argument and given that atmospheric pollution is the leading environmental risk for human beings (Galán-Madruga, 2021), exposure control is translated into a primordial activity within the Public Health frame (Qi et al., 2023; Galán-Madruga et al., 2025). The legislation allows Member States to evaluate air quality using diverse methodologies, such as fixed and indicated measurements and modeling studies (more information in Annex I from Directive 2008/50/EC).

Research groups worldwide have been utilizing modeling approaches to estimate polluting levels in ambient air (Mao et al., 2021; Huang et al., 2024; Pyae and Kallawicha, 2024; Zhai et al., 2024; Zhang et al., 2024). In this context, our study aims to provide a reliable and easy-to-implement tool that has the potential to significantly advance the estimation of daily urban benzene levels in ambient air. This study represents a promising step forward in the field of air quality monitoring.

Benzene (C<sub>6</sub>H<sub>6</sub>) is a compound included in the volatile organic compounds family (VOCs) (Sowmya et al., 2024), and it deserves awareness because it is the VOCs pollutant most troublesome in urban areas (Zhao et al., 2004). C<sub>6</sub>H<sub>6</sub> plays a fundamental role in public health due to its potentially harmful effect on human health (Sassano et al., 2024). Over the last several decades, there has been rising proof sustaining that via inhalation exposure to this contaminant relates to acute myeloid leukemia and, presumably, lymphoma and childhood leukemia (Smith, 2010). It is part of group 1 of the IARC carcinogenic classification system (IARC, 2012; IARC, 2018; Andersson et al., 2024).

Estimating a target variable using predictor variables implies a previous relationship between the variable set. In this sense, C<sub>6</sub>H<sub>6</sub> participates in yielding secondary pollutants, being the NO<sub>x</sub>/VOCs ratio a key factor in the atmospheric ozone (O<sub>3</sub>) chemistry (Costagliola et al., 2014), and dependent on the presence of sunlight and other meteorological features, mainly temperature since an increase in temperature is related to the rise in photochemical reactivity (Deng et al., 2019). Other research groups have already demonstrated the strong influence of temperature on ambient air benzene concentrations (Król et al., 2012). Relative to the NO<sub>x</sub>/VOCs ratio, it is important to highlight that NO<sub>x</sub> includes nitrogen monoxide and dioxide (NO and NO<sub>2</sub>, respectively) (Jacob, 2000). Nevertheless, NO reacts within minutes with O<sub>3</sub> to yield additional NO<sub>2</sub> (Lin et al., 2016). In terms of VOCs, toluene, ethylbenzene, xylene, and C<sub>6</sub>H<sub>6</sub> explain about 95 % of the total VOCs (Zhao et al., 2011). The leading VOCs role in the tropospheric O<sub>3</sub> formation is the oxidation of NO to NO<sub>2</sub> (Finlayson-Pitts and Pitts, 1993). Within the COVs family, C<sub>6</sub>H<sub>6</sub> is the major contributor to O<sub>3</sub> formation (Xu et al., 2018). Gómez-Moreno et al., 2022 investigated VOCs levels in Madrid City (CIEMAT facilities), evidencing the highest concentrations in the case of C<sub>6</sub>H<sub>6</sub>. In addition, both compounds (C<sub>6</sub>H<sub>6</sub> and NO<sub>2</sub>) share local emission sources, mainly road traffic (Carr et al., 2002; Schnitzhofer et al., 2008), evidencing a direct chemistry relationship between them. To potential readers, atmospheric chemistry involving NO<sub>x</sub>, COVs, and O<sub>3</sub> has already been described by Atkinson, 2000; Monks et al., 2015; and Lu et al., 2019. In the specific case of Madrid City, Reche et al., 2018 and De la Paz et al., 2024 studied the process influencing the formation of tropospheric O<sub>3</sub>.

The assessment of ambient air benzene is paramount and mandatory in European Member States. However, not all fixed stations within an air quality monitoring network measure this atmospheric pollutant. For this reason, modeling studies are highly needed, providing complementary polluting data in previously non-monitored points. In the case study, our proposed approach estimates daily ambient air benzene levels using variables easily monitored by air quality networks. This tool will enrich the available information within any network on the benzene levels in the ambient atmosphere, decreasing uncertainty derived from the lack of actual benzene exposure information when supplying target pollutant data reliably in non-measured zones.

## 2. Materials and methods

### 2.1. Study area and reference pollution dataset

Madrid City was regarded as a case study for developing the proposed objective. It is Spain's capital and is located in the Community of Madrid (central region of the Iberian Peninsula). The city of Madrid has an estimated population of over 3.3 million people and encompasses a land area of approximately 600 km<sup>2</sup> (National Statistical Institute 2023, <http://www.ine.es>). It is divided into 21 districts and 128 neighborhoods. In a European context, Madrid City is one of the European capitals with greater population density (see <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20200430-1>, accessed May 29, 2024).

Given that each geographic area is characterized by a particular polluting level, the proposed model will be implemented in another air quality network to guarantee its independence from the potential effect of meteorological conditions and emission sources specific to each zone (Dayan et al., 2023).

The covered period to develop the suggested objective involved 2019 and 2020. Daily C<sub>6</sub>H<sub>6</sub>, NO<sub>2</sub>, O<sub>3</sub>, and temperature (T) data were acquired from Madrid City Council's open data portal (<https://www.madrid.es/portal/site/munimadrid>). C<sub>6</sub>H<sub>6</sub> corresponds to the dependent variable, whereas the rest of the input are predictor variables. Madrid City's network uses reference methods for measuring the involved gases (gas chromatography for C<sub>6</sub>H<sub>6</sub>, chemiluminescence for NO<sub>2</sub>, and ultraviolet absorption for O<sub>3</sub>, according to the current European Directive). The previously mentioned reference methods are applicable up to 50 μg C<sub>6</sub>H<sub>6</sub>/m<sup>3</sup>, 500 μg NO<sub>2</sub>/m<sup>3</sup>, and 500 μg O<sub>3</sub>/m<sup>3</sup> (more information on method foundation, maintenance, verifications, calibrations, among other factors related to the measurement method, see EN 14662–1, 2005, EN 14211, 2013, and EN 14625, 2013 for C<sub>6</sub>H<sub>6</sub>, NO<sub>2</sub> and O<sub>3</sub>, respectively).

It's important to note that the original dataset used to develop the mathematical models (urban traffic, suburban background, and rural background) involved data from 2019 to 2020. To ensure the relevance and applicability of our research, 2021 data was selected to test the proposed model's performance.

As additional information, 2019–2020 polluting concentrations recorded by the selected fixed stations reached average levels of  $0.43 \pm 0.35$ ,  $52.29 \pm 24.72$ , and  $30.42 \pm 20.28$   $\mu\text{g}/\text{m}^3$  for  $\text{C}_6\text{H}_6$ ,  $\text{O}_3$ , and  $\text{NO}_2$ , respectively. Regarding meteorology, Madrid City is characterized by temperate and dry winters and warm summers (Kottek et al., 2006). Over the study time, an average temperature value of  $16.29 \pm 7.58$  °C was reached.

## 2.2. Modeling daily urban benzene levels from other inputs measured in air quality networks

The leading objective proposed in this study is to formulate a mathematical model that can effectively explain the relationship between the dependent variable and those predictors. Over the covered time, the target network consisted of 24 fixed stations (see Fig. S1). The station organization within Madrid City's air quality network is based on two criteria: (i) location (urban or suburban) and (ii) main pollution source (traffic or background) (more information in Table S1). The current legislation reports the definition of the different types of fixed stations (Directive 2008/50/EC, 2024), as well as other International Organisms (such as European Environment Agency, <https://www.eea.europa.eu/en/topics/in-depth/air-pollution/monitoring-station-classifications-and-criteria>, accessed September 10, 2024).

A comprehensive multiple linear regression analysis (MLR) was conducted to include all variables involved in the study, allowing the construction of a robust mathematical model to estimate ambient air urban benzene concentrations from the limited predictor variables, which are commonly monitored by air quality networks.

The MLR analysis deemed the organization criteria of fixed stations, both the location and main combustion source (urban traffic, urban background, and suburban background). The mathematical formula explaining the relationship between urban daily benzene levels in ambient air and the predictor inputs is reported in eq. (1):

$$\text{C}_6\text{H}_6 = \text{Cte} + ax_1 + bx_2 + cx_3 + \varepsilon \quad (1)$$

where, *Cte* equals to a constant value (intercept); *a*, *b*, and *c* correspond to linear regression coefficients (slope);  $x_1$ ,  $x_2$  and  $x_3$  are the predictor inputs ( $\text{NO}_2$ ,  $\text{O}_3$  and *T*, respectively), and  $\varepsilon$  is the bias associated with the proposed model development. Table S2 reports the outcomes reached when implementing MLR analysis regarding regression coefficients, intercept values, and Pearson coefficients.

## 2.3. Validating the proposed model to estimate ambient air daily urban benzene levels

The validation of suggested methodologies to enhance or complement any subject at the science level is a highly relevant instrument that allows for quantifying the degree of certainty regarding the functional implementation of the proposed method. A linear regression analysis was performed between the estimated and current urban daily benzene concentrations (candidate and reference method, respectively) to assess the confidence of the proposed method. The criteria used to value the reached outcomes when applying linear regression were the following: (i) measuring the correlation between estimated and current urban daily benzene levels, (ii) statistical significance of the simple linear regression equation with one-way ANOVA (Analysis of Variance), and (iii) statistical significance of the regression coefficient of the independent variable.

Within the same context, the association degree between a candidate method (in this case, the proposed method to estimate urban daily benzene levels) and a reference method (real benzene levels measured by reference methods in the fixed stations within an air quality network) is evaluated from bias determination (Eurachem, 2012). In the function of the type of station, the proposed model bias was quantified, calculating the difference between the daily estimated and actual benzene concentrations during 2019–2020. Subsequently, every model equation (urban traffic, suburban background, and rural background) was adjusted with the average value of the bias, thereby allowing for improvement in the proposed model performance (Bravo et al., 2012).

## 2.4. Applicability of the suggested model to complement the information provided by air quality networks

During 2021, the target air quality network measures urban benzene levels in six sampling points, although it comprises 24 stations. Implementing the proposed model in those stations measuring predictor inputs ( $\text{NO}_2$ ,  $\text{O}_3$ , and *T*) will provide an estimate of urban daily benzene concentrations, thereby supplementing the information supplied by the network. In this regard, estimating benzene may be addressed in additional 10 fixed stations, substantially improving the available information on daily exposure to urban benzene levels (see Fig. S2).

At the environmental level, a study to identify the most representative fixed stations within the target network monitoring benzene is performed since the total number of fixed stations that provide benzene information (current and estimated measurements in six and ten stations, respectively) is higher than the requirement set in the current European Directive. Note that the spatial benzene distribution gradient in the target area should not vary once removing those non-representative fixed stations. A hierarchical method of partitive clustering (k-means clustering with 10 maximum iterations) was applied to achieve this objective (Salem et al., 2018). Clustering analysis is usually employed in diverse scientific subjects due to its efficiency and simplicity (Jain, 2010). To not repeat information already published, those readers needing more information on clustering analysis, see Govender and Sivakumar, 2020. In the case study, applying the selected clustering test to the sixteen fixed stations providing benzene data (estimated and current)

identifies different clusters of fixed stations with similar benzene levels. Each cluster is represented by the average benzene concentration calculated from every fixed station included in this cluster. In line with other studies, Euclidean distance and cluster standard deviation were used to value each cluster as the spatial indicator and cluster membership identifier, respectively (Galán-Madruga et al., 2023). Note that the Euclidean distance is the difference between the average benzene concentration of a target cluster (representative concentration) and the benzene concentration for each fixed station included in this cluster. In contrast, the standard deviation of the benzene concentrations of this cluster corresponds to the cluster standard deviation term (Galán Madruga et al., 2018).

Identifying the most representative fixed stations is to adapt the existing resources within the network without losing information regarding the spatial benzene distribution gradient. To test this hypothesis, monthly average isoconcentration maps of estimated and current benzene were built using a geographical information system (Surfer for Windows, Win32: Surface Mapping System, v.6.04. Golden Software, Inc., Golden, CO, USA). The Kriging method was used for estimating benzene levels in non-measured points since it is appointed as the best geospatial interpolation method (Cintrón et al., 2024). In order to quantify the similarity percentage of the interpolated spatial information at sixteen fixed monitoring stations versus those most representative stations, a linear regression-based analysis was used. On the other hand, to assess the performance of the previous hypothesis, root mean square error (RMSE), mean prediction error (MAE), and mean absolute percentage error (MAPE) were employed as metric measures, which is in line with other studies within the environmental sciences field (Karunasingha, 2022). The following equations were used to calculate previously cited indicators (Dai et al., 2022).

$$RMSE = \sqrt{\frac{\sum (C_6H_6_E - C_6H_6_C)^2}{N}} \quad (2)$$

$$MAE = \frac{\sum (C_6H_6_E - C_6H_6_C)}{N} \quad (3)$$

$$MAPE = \frac{\sum |C_6H_6_E - C_6H_6_C|}{N} \times \frac{100\%}{C_6H_6_C} \quad (4)$$

where,  $C_6H_6_E$  ( $\mu\text{g}/\text{m}^3$ ) and  $C_6H_6_C$  ( $\mu\text{g}/\text{m}^3$ ) correspond to the estimated and current benzene concentrations, respectively. N equals the number of samples.

### 3. Results and discussion

#### 3.1. Modeling daily urban benzene levels from other inputs measured in air quality networks

In order to offer an overview of air quality status over the study time, a statistical summary is presented. In the function of the category of the fixed station, and during the 2019–2020 period, averaged  $C_6H_6$  levels of  $0.62 \pm 0.39$ ,  $0.35 \pm 0.30$  and  $0.32 \pm 0.27$   $\mu\text{g}/\text{m}^3$  in the target urban traffic and background, and suburban background fixed station (ESA, FAR, and CDC, respectively) provide crucial insights. Similarly, average  $O_3$  values of  $46.35 \pm 21.74$ ,  $55.00 \pm 26.48$  and  $55.42 \pm 24.67$   $\mu\text{g}/\text{m}^3$  were observed in the target urban traffic and background and suburban background location, while  $NO_2$  levels of  $43.33 \pm 20.01$  (urban traffic),  $29.88 \pm 17.91$  (urban background) and  $18.36 \pm 14.28$   $\mu\text{g}/\text{m}^3$  (suburban background) were averaged over the studied period. Regarding temperature, similar levels were recorded ( $16.27 \pm 7.26$  in the urban traffic site,  $17.13 \pm 7.65$  in the urban background site, and  $15.01 \pm 7.74$  °C in the suburban background).

According to the information reported in the 2.2 section, mathematical models were developed in the function of the type of station. Irrespective of the station type, the mathematical equations demonstrated an outstanding predictive daily benzene capacity, relying on the independent variables (with a coefficient of correlation ranging from 80 % to 90 %). They sustained the highest predictive ability in suburban and urban background categories, respectively, and in the urban traffic site as the minor.  $C_6H_6$  emission into the atmosphere is higher and more fluctuant in traffic sites than in background sites, which generates the greatest predictive difficulty. It needs to be highlighted that all 2019–2020 data monitored by Madrid City's air quality network were included in the suggested daily benzene model study since there were no run outlier tests.

Once daily urban benzene levels were modeled for each type of fixed station (2019–2020), the bias associated with the proposed approach was calculated according to the 2.2 section, and then, the mean bias value was included in each model (urban traffic, urban background, and suburban background). In this sense, average bias values of  $-0.002$   $\mu\text{g}/\text{m}^3$  (0.25 % expressed as a relative value),

**Table 1**

Equations of generic multiple linear regression to estimate the daily urban benzene concentrations in ambient air as the function of the type of station.

Category of station	Mathematical equation (proposed model for estimating daily urban benzene levels)
Urban traffic	$C_6H_6 = 0.3210 - (0.005 \times O_3^a) + (0.013 \times NO_2^b) - (0.001 \times T^c)$
Urban background	$C_6H_6 = 0.2173 - (0.002 \times O_3^a) + (0.011 \times NO_2^b) - (0.005 \times T^c)$
Suburban background	$C_6H_6 = 0.2900 - (0.001 \times O_3^a) + (0.013 \times NO_2^b) - (0.009 \times T^c)$

Key: <sup>a</sup> Ozone ( $O_3$ ) in  $\mu\text{g}/\text{m}^3$ , <sup>b</sup> nitrogen dioxide ( $NO_2$ ) in  $\mu\text{g}/\text{m}^3$ , and <sup>c</sup> Temperature (T) in °C.

$-0.0003 \mu\text{g}/\text{m}^3$  (0.09 %), and  $1.7 \cdot 10^{-15} \mu\text{g}/\text{m}^3$  ( $5.3 \cdot 10^{-15}\%$ ) were observed for urban traffic and background, and suburban background, respectively (see Fig. S3). Table 1 shows equations of generic multiple linear regression to estimate the daily urban benzene concentrations in ambient air as the function of the type of station. When considering all predictor variables together, the individual influence that each one exerts on the estimated daily urban benzene concentrations evidenced that  $\text{NO}_2$  was the dominant independent input in all stations regardless of its category. Nevertheless, the least influencing predictor variable was  $\text{O}_3$ , except for the urban traffic category, for which T was the least dominant predictor variable. As a potential explanation and in line with the atmospheric chemistry of nitrogen oxides-ozone, lower  $\text{O}_3$  levels in the ambient air of traffic locations occur than in urban and suburban background environments. Similar findings were observed regarding individual predictor variables ( $\text{NO}_2$  as the most influencing predictor variable, followed by  $\text{O}_3$  and T, see Fig. S4). As can be seen,  $\text{NO}_2$  levels were positively correlated with  $\text{C}_6\text{H}_6$  concentrations (they are primary pollutants and share common emission sources, Hien et al., 2014) since ambient air  $\text{NO}_2$  formation occurs within minutes (see introduction section). Nevertheless, the relationship between the rest of the independent variables and benzene is negatively correlated, which was expected since both predictor inputs favor the photochemical reactions of this polluting compound in ambient air, assisting its reduction (Liu et al., 2019). These observations align with other outcomes reported by scientific literature in urban environments (Marc et al., 2014).

Given the need-to-know exposure benzene levels, paramount research groups have developed different approaches to estimate ambient air benzene concentrations in unmeasured points. So, the passive methodology has been employed in this sense, given that it allows monitoring of an extensive study surface (Bozkurt et al., 2018; Khuriganova et al., 2019). The measured benzene data employing this methodology have a higher uncertainty level than those monitored by air quality networks (reference methods) since the passive methodology doesn't comply with the reference method set in the current European legislation (Directive 2008/50/EC, 2024), sustaining the main difference concerning the current research study (data of the dependent and independent variables were collected using reference methods).

Other authors used historical data series to conduct descriptive and estimate analysis. So, an investigation group belonging to the Department of Computer Science of the Bhagwan Parshuram Inst. of Technology, Rohini, in New Delhi, forecasted benzene trends up to 2022, employing data from 2009 to 2017 (Sharma et al., 2018). They did not validate the developed approach, so forecasted benzene concentrations were not tested (either in the same geographical area or another one); therefore, the approach's performance is not known. In the present study, a validation process of mathematical expressions suggested for estimating ambient air benzene levels was addressed. Other groups have conducted their investigations within the same objective based on diverse approaches. Chang et al., 2015 regarded a novel modeling approach combining dispersion models, traffic emissions, and transportation networks to estimate exposure levels to traffic-related air pollutants in Portland, Oregon. They implemented the proposed model in two distinct geographical areas but did not set an uncertainty value associated with this model, constituting the leading difference concerning our work. Using regression models, Su et al., 2010 assessed the impact of physical geography, traffic density, population density, and land use on benzene levels, among other volatile organic compounds. They developed the modeling approach based on a short-term time (from July 25 to August 9, 2006, 16 days), unlike our work (from January 1, 2019, to December 31, 2020, 731 days). This fact is highly relevant because modeling that covers a lot of time carries greater model reliability. The approach proposed by Su et al., 2010 presents a more complex implementation than our approach, given that it needs a more significant number of study variables that are difficult to obtain and set. In addition, they did not test the suggested approach in other geographical areas, unlike our work. Like the present work, other authors implemented an MLR model to investigate the influence of meteorological conditions and emission sources on benzene and other volatile organics in outdoor residential locations of diverse cities in the USA (Kwon et al., 2016). They did not validate their approach using data non-included in the development of the approach. In the same context, more complex techniques were used to estimate spatiotemporal variations in benzene concentrations (Hsu et al., 2022). They employed various land-use Regression-based multiple machine learning algorithms, such as random forest, deep neural network, gradient boosting, light gradient boosting, CatBoost, extreme gradient boosting, and ensemble algorithms. A validation process was not addressed. In addition, this approach is complex to implement due to the application of machine learning techniques.

In summary, our research work offers scientific advances concerning the notorious previously reported investigations, as listed following: (i) a model estimating ambient air benzene levels has been proposed only using three predictor variables (commonly measured in air quality networks), (ii)  $\text{C}_6\text{H}_6$  and  $\text{NO}_2$  data used in this investigation, to develop the suggested model, were measured by reference methods according to the current European Legislation, (iii) modeling outcomes complies with legislative air quality modeling criteria, (iv) the suggested model has been validated using data from a different period to that one employed to develop the model, (v) the generated mathematical expressions have been satisfactorily tested in another geographical area, and (vi) the proposed model has allowed identifying the most representative fixed stations for measuring benzene in the studied domain.

Although the previously indicated notable studies contributed to scientific advancement, the performance of the suggested models was not validated. The proposed mathematical model in this work was tested in a period other than that employed to engineer the estimating model and applied to a different geographic area as leading contributions.

### 3.2. Validating the proposed model to estimate ambient air urban daily benzene levels

Given that 2019–2020 daily benzene concentrations were used to develop the mathematical models, 2021 daily benzene data were employed to test the performance of the suggested approach. To achieve that objective, the same fixed stations in both sceneries (2019–2020 and 2021) were considered, encompassing an urban traffic station (identified as ESA), urban background station (FAR), and suburban background station (CAC).

A simple linear regression analysis was conducted between 2021 daily estimated vs current benzene concentrations to evaluate the

reliability of the suggested model. Pearson’s coefficient was utilized to value the previously cited study. So, coefficients of correlation of 0.682 (urban traffic location), 0.642 (urban background), and 0.914 (suburban background) were observed. In this context, the daily current benzene level (independent input) explains 46.5 %, 41.2 %, and 83.6 % (urban traffic and background and suburban background, respectively) of the daily estimated benzene concentration’s total variance (dependent variable). According to the approach proposed by [Dancey and Reidy \(2007\)](#), the relationship between 2021 daily estimated vs current benzene levels is moderate (in the case of the urban background site) and strong (for the urban traffic and suburban background locations). The equation of the simple linear regression between estimated and current benzene concentrations resulted in values close to 1 and 0 for the slope (0.74, 0.80, and 0.85 for the urban traffic, urban background, and suburban background stations, respectively) and intercept (0.19, 0.13, and 0.07 for urban traffic, urban background, and suburban background, respectively). Considering together, the linear reliance is significant in terms of statistics, promoting the implementation of the suggested mathematical approach to estimate daily urban ambient air benzene levels. According to results reached in the statistical analysis (ANOVA) of the simple linear regression equation, the prediction model significantly improves the estimate of daily benzene levels with an F value of 265.430 ( $p < 0.001$ ), 214.434 ( $p >$

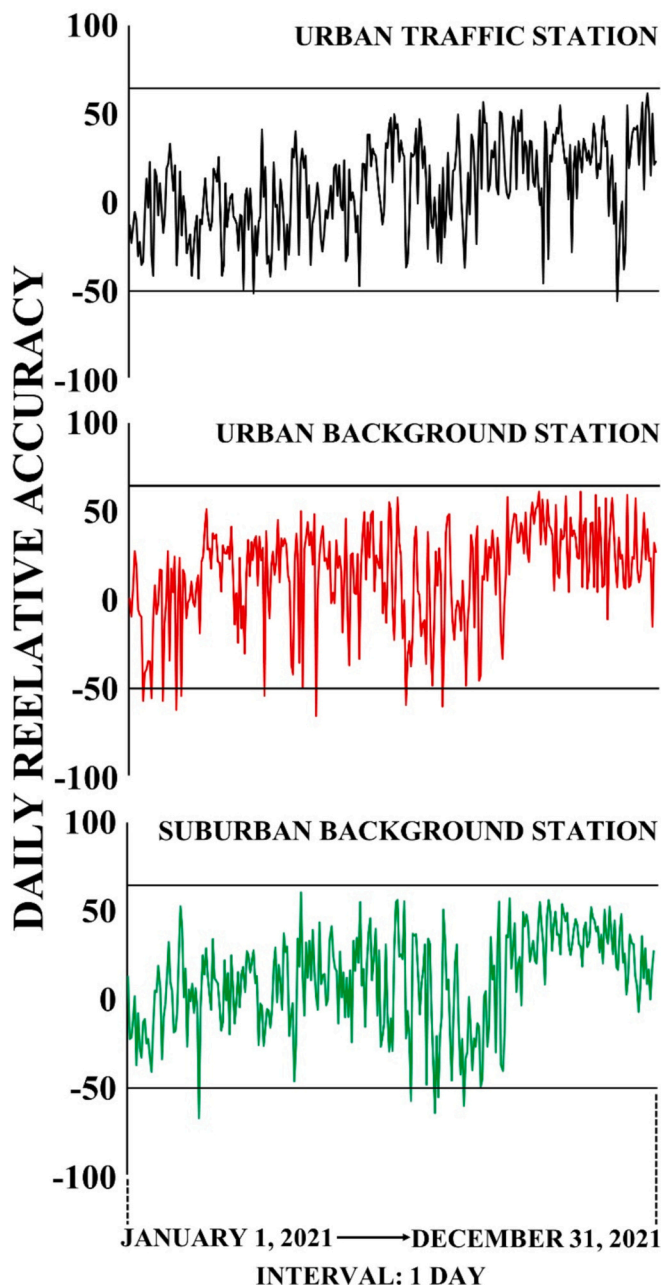


Fig. 1. Daily accuracy (%) resulting when comparing daily estimated vs current 2021 benzene data.

0.001) and 1501.598 ( $p < 0.001$ ) for urban traffic and background, and suburban background. On the other hand, ANOVA of the regression coefficient of the predictor inputs suggested their insertion in the linear regression equation ( $p < 0.001$ ).

As a second validation test, the current normative framework was regarded. In this sense, the European legislation lays down data quality objectives for ambient air quality assessment. In this context, it sets a modeling uncertainty value to estimate annual ambient air benzene levels of 50 % (Directive 2008/50/CE, see Annex I). During 2021, it was observed an annual average current  $C_6H_6$  level of  $0.32 \pm 0.23 \mu\text{g}/\text{m}^3$ . The 2021 annual average  $C_6H_6$  level estimated by the suggested mathematical model was  $0.38 \pm 0.25 \mu\text{g}/\text{m}^3$ . For comparing the estimated vs current value, the relative accuracy was calculated according to the Guide to the Expression of Uncertainty in Measurement published by the International Bureau of Weights and Measures (JCGM 100, 2008; JCGM GUM-6, 2020). In this sense, a 19.75 % relative accuracy was observed. This value is notably lower than that reported by current legislation. In the function of the type of station, annual current vs estimated benzene levels of  $0.43 \pm 0.25$  and  $0.51 \pm 0.23 \mu\text{g}/\text{m}^3$  for the urban traffic station,  $0.23 \pm 0.17$  and  $0.51 \pm 0.23 \mu\text{g}/\text{m}^3$  for the urban background, and  $0.29 \pm 0.26$  and  $0.30 \pm 0.24 \mu\text{g}/\text{m}^3$  for the suburban background were averaged. Note that the corresponding relative accuracies comply with the legislative requirements (18, 38, and 4 % for urban traffic, urban background, and suburban background, respectively). Therefore, the proposed model to forecast daily average urban benzene levels properly and reliably estimates annual average urban benzene levels for each type of fixed station and may be potentially applied in other geographical areas. This reflection is relevant since the limit value established by current legislation to protect human beings' health and the environment is  $5 \mu\text{g}/\text{m}^3$ , expressed as an annual average value. In the case study, yearly average benzene concentrations (2019–2021) are below the annual air quality standard.

The current European legislation does not set a modeling daily uncertainty value when focusing on daily estimated benzene data. Fig. 1 represents the daily relative accuracy obtained when comparing daily estimated vs current 2021 benzene concentrations. As can be seen, most outcomes fall into the  $\pm 50$  % interval laid down by the current legislation (regardless of the station category), and a data minority drops outside this interval (daily maximum relative accuracy value equals  $-56$ ,  $67$ , and  $68$  % for urban traffic, urban background, and suburban background, respectively). To value this scenery, the difference between daily and annual modeling uncertainties for other air pollutants included in the legislation (such as sulfur dioxide, nitrogen dioxide, oxides of nitrogen, and carbon monoxide) was considered. So, daily and annual modeling uncertainties of 50 and 30 % are established for other air pollutants distinct from benzene. Note that uncertainty associated with daily polluting data is significantly higher than annual, supporting an increase of 63 % (Directive 2008/50/EC, 2024). Whether a similar variation is applied to this case study, modeling daily benzene uncertainty of 83 % is obtained. In this context, daily urban benzene concentrations estimated using the suggested approach are notably lower than that previously reported uncertainty value. Therefore, the recommended approach helps estimate daily urban benzene levels in no-measure points from selected predictor variables, complying with legislative uncertainty criteria sets for other pollutants.

As a third validation test, the proposed model was implemented in another geographic area, given that the occurrence and permanence of air pollutants depend on emission air pollutants sources and meteorological conditions, respectively, specific to each geographic region. In this regard, fixed stations measuring both dependent and predictor variables and located in other geographic zones were selected to elucidate whether the recommended mathematical model depends on topography, human activities, and existing transport networks, among other factors. Within the Community of Madrid's air quality network, they were regarded the stations placed in Alcobendas (urban traffic station, national code: 28134002, European code: ES1564A. Distance to the Madrid City: 26 km according to Google Maps, <https://www.google.es/maps>), and El Atazar (rural background station, national code: 28016001, European code: ES1802A. Distance to the Madrid City: 99 km according to Google Maps). Fig. 2 shows the obtained outcomes when

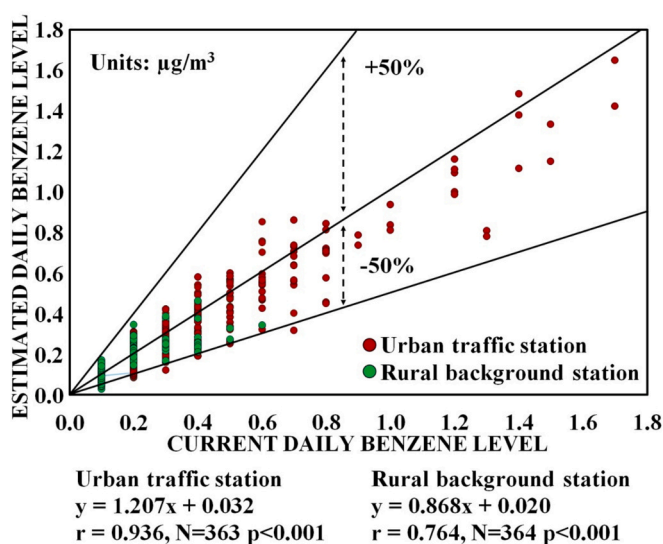


Fig. 2. Plot of estimated vs current daily benzene levels (2021). Obtained outcomes when applying the suggested approach to a different geographical area concerning the target studied area.

applying the proposed approach to another air quality network. The mathematical formula implanted in El Atazar corresponds to the proposed suburban background model. Note that the performance of the recommended model to forecast daily benzene concentrations in ambient air, in a geographic zone distinct to that conducted to develop the approach, complies with the annual benzene modeling uncertainty value set by the current legislative frame. In line with other legislated air pollutants, to highlight that the daily modeling uncertainty is more permissive than that annual.

### 3.3. Applicability of the suggested mathematical model

Developing this research has allowed us to elucidate the relationship between daily urban benzene levels (dependent variable) and three predictor variables commonly measurable in air quality networks, which has been translated into a validated mathematical equation letting estimate the dependent variable in unmeasured points. This approach may serve as a valuable tool within the environmental sciences, complementing benzene exposure information provided by air quality networks and identifying the most representative fixed benzene stations within the study area.

#### 3.3.1. Complementing the polluting information provided by air quality networks

Given the notable increase in the number of fixed stations providing daily benzene information (current and estimated values), a comparison of monthly 2021 concentrations is conducted. Table 2 shows the monthly average benzene levels for two sceneries (A: stations measuring current benzene,  $n = 6$ , and B: stations providing current and estimated benzene,  $n = 16$ ). When running a paired samples test, significant differences are observed ( $p > 0.001$ ). In this regard, annual levels of  $0.31 \pm 0.17 \mu\text{g}/\text{m}^3$  (scenery A) and  $0.51 \pm 0.23 \mu\text{g}/\text{m}^3$  (scenery B) were averaged in 2021, evidencing different polluting sceneries. Note that when covering more study area (new fixed stations estimating benzene,  $n = 10$ ), the pollution information may vary due to considering new emission sources and meteorological features.

According to the reported information in the 2.4 section, Fig. 3 represents the 2021 monthly average isoconcentration maps of benzene within the covered territory. Broadly, a spatial pattern is observed throughout the year, with the highest monthly average levels located in the east-north zone, due mainly to the presence of Adolfo Suárez-Madrid Barajas airport (Latitude:  $40^\circ 29' 43''\text{N}$ , Longitude:  $3^\circ 33' 36''\text{W}$ ). It is well known that polluting emissions from airport areas have a direct impact on the surrounding local atmosphere (Stettler et al., 2011; Christodoulakis et al., 2022; Trebs et al., 2023). Airport areas have been identified as remarkable emission sources of volatile organic compounds in urban environments (Zhu et al., 2023), such as benzene (Gaeta et al., 2016). In this context, Yang et al., 2018 measured volatile organic compounds at Beijing Capital International Airport, identifying benzene as one of the most abundant species. Note that this environmental scenery is not detected only using the network's current benzene data. As an example, average levels of benzene in July 2021 were depicted (see Fig. S5). As can be seen, the influence of airport emissions on the surrounding atmosphere is not covered only considering the fixed stations monitoring current benzene ( $n = 6$ ). Therefore, the easy application of the suggested mathematical model sustains a notable input regarding benzene levels information in no-monitored areas, which is essential in the public health frame.

Despite this, Madrid City's air quality network widely complies with criteria laid down in Directive 2008/50/EC regarding a micro- and macro-implantation requirement and a minimum number of sampling benzene points. Whereas the legislation lays down 3 sampling points for a population from 2,750,000 to 3,749,000 inhabitants (see Annex V of the European Directive), Madrid City's network monitoring benzene in 6 sampling locations.

The reported outcomes sustain that the proposed approach in this research work allows easy and reliable short-time urban benzene estimating, which is relevant to extending urban pollution information provided by any air quality networks at the global level and expanding the scope of potential epidemiological studies.

The reported outcomes sustain that the proposed approach in this research work allows easy and reliable short-time urban benzene estimating. This is relevant to extending urban benzene pollution information from three predictor inputs and expanding the scope of

**Table 2**  
Monthly 2021 benzene concentrations. Units:  $\mu\text{g}/\text{m}^3$ .

Monthly	Scenery A <sup>a</sup>	Scenery B <sup>b</sup>
January	$0.59 \pm 0.20$	$0.55 \pm 0.13$
February	$0.38 \pm 0.15$	$0.48 \pm 0.13$
March	$0.34 \pm 0.13$	$0.51 \pm 0.17$
April	$0.25 \pm 0.09$	$0.51 \pm 0.24$
May	$0.22 \pm 0.10$	$0.53 \pm 0.30$
June	$0.18 \pm 0.08$	$0.55 \pm 0.35$
July	$0.17 \pm 0.07$	$0.54 \pm 0.35$
August	$0.17 \pm 0.06$	$0.52 \pm 0.33$
September	$0.18 \pm 0.04$	$0.42 \pm 0.22$
October	$0.37 \pm 0.16$	$0.49 \pm 0.15$
November	$0.41 \pm 0.12$	$0.49 \pm 0.11$
December	$0.48 \pm 0.11$	$0.51 \pm 0.09$

Key: <sup>a</sup> Monthly average level obtained from six fixed stations measuring current daily benzene, and <sup>b</sup> Monthly average level obtained from 16 fixed stations providing daily benzene data (6 and 10 stations measuring current and estimate, respectively).

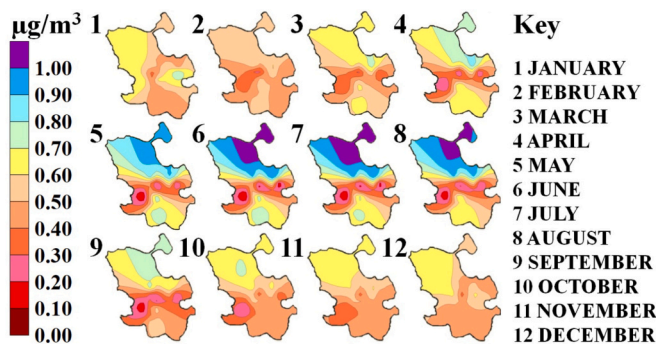


Fig. 3. Iso-concentration maps of monthly 2021 benzene levels applying the suggested model ( $N =$  sixteen fixed stations, of which six and ten measure current and estimate benzene, respectively).

potential epidemiological studies, which could potentially be applicable in other geographical zones.

### 3.3.2. Identifying the most representative fixed benzene stations

2021 monthly studies were conducted to detect those most representative stations within the fixed station set, consisting of stations measuring current and estimating daily benzene levels ( $n = 16$ ). To run the k-means clustering analysis, the final number of clusters ( $k$  value) should previously be defined. In this sense, a  $k = 3$  value was considered, coinciding with the legislative requirement about the minimal number of sampling benzene points in the target domain. Table S3 reports monthly clustering outcomes in the function of the locations. Broadly, good relationships between monthly average benzene concentrations obtained when applying the suggested approach ( $N = 16$  fixed stations) and k-means analysis ( $k = 3$ ) were observed, exhibiting coefficients of determination between 0.816 and 0.971 (average equals  $0.918 \pm 0.054$ ) (see Table S4). This finding is relevant since it sustains similar datasets. In the same context, maximum values of 0.160 (Euclidean distance) and 0.051 (standard deviation) were observed. To easily understand the results reported in Table S3, Fig. S6 pictures the monthly representativity for 2021, appointing March and April as the months needing more fixed stations for keeping adequate similitude between the suggested approach and k-means analysis outcomes. Regarding location, CUC and VIL have the highest number of representations (6), and PLC has the lowest (0). In the function of the fixed station category, urban background was the station type with the greatest number of times appointing as representative (21 times), followed by urban traffic (18 times) and suburban background (9). In terms of season, the Spring period (from March to May), Winter (from December to February), Autumn (from September to November), and finally, Summer (from June to August) with 16, 13, 10, and 9 locations, respectively. Nevertheless, the cold (Winter and Autumn) and warm (Spring and Summer) periods maintain reasonably similar statistics (23 and 25 locations). Based on the previous evidence, applying the proposed model in this study to estimate daily urban benzene levels is also adequate to identify the fixed benzene stations with the highest monthly representativity.

Given that current European legislation sets an annual average benzene concentration to protect human health as a unique air quality objective, the previously developed exercise is addressed in terms of a yearly concentration of 2021. In this sense, Table S5 reports the reached outcomes. It was observed a maximum Euclidean distance and standard deviation value of  $0.131$  and  $0.039 \mu\text{g}/\text{m}^3$ , respectively. The most representative fixed annual benzene stations were VIL (urban background), EPA (suburban background), and CDC (suburban background). In order to quantify potential changes in spatial distribution 2021 information, a spatial gradient valuation is addressed to discern whether spatial information is lost or not when removing those non-representative fixed stations. In this sense, Fig. 4 represents maps of annual average benzene isoconcentrations (2021). Note that the spatial distribution gradient is highly similar between both sceneries, exhibiting a coefficient of determination of 0.919, which is translated into acceptable quantitative differences ( $0.07 \mu\text{g}/\text{m}^3$  of RMSE,  $0.01 \mu\text{g}/\text{m}^3$  of MAE, and 5.01 % of MAPE). Therefore, the easy application of the proposed approach allows for the reliable estimation of annual benzene concentrations and the identification of those more representative fixed

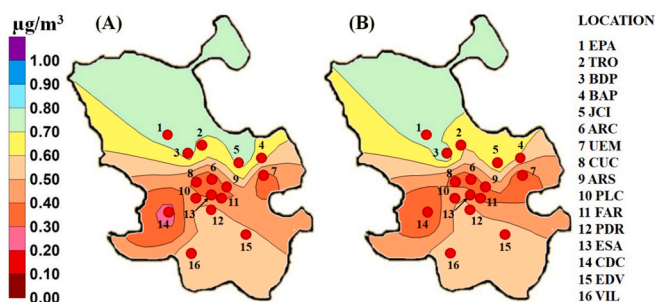


Fig. 4. Iso-concentration maps of annual 2021 benzene levels. (A) Applying the suggested model ( $N =$  sixteen fixed stations, of which six and ten measure current and estimate benzene, respectively), and (B) Implementing k-means analysis ( $k = 3$ ).

stations.

Since the managers of air quality networks cannot change the location of a specific fixed station monthly, identifying the most representative fixed stations in air quality networks for an area of study allows for maximizing the polluting information in the function of the existing resources, which is essential within a limited resources framework (Guigoz et al., 2017), without losing spatial information due to the lack of measurements at the rest of the stations (Christakos et al., 2017), provided that this data is generated by spatial extrapolation algorithms and/or statistical procedures, being in line with other authors (Aguilera et al., 2016).

Given that limited data on air pollutants creates uncertainties about current people's exposure, the proposed approach's easy implementation may address paramount issues in environmental sciences, thereby helping to improve air quality network management.

The proposed model's capability to reliably estimate ambient air daily urban benzene levels in zones not currently measured may furnish relevant added data on benzene exposure and its spatial-temporal behavior, thus enhancing adequate compliance with the air quality networks' functions.

#### 4. Conclusions

This study presents a straightforward, fast, and reliable approach to estimating daily urban ambient air benzene levels. It uses predictor variables commonly measured in air quality networks (NO<sub>2</sub>, O<sub>3</sub>, and T) and sustains its validity from exhibited outcomes.

Predicting approaches such as this are particularly relevant for urban air quality networks comprising a few fixed stations that really measure ambient air benzene, thus evading the additional expenses of increasing the number of measuring benzene sites. Within the same context, air quality networks with an elevated number of fixed benzene stations might be adjusted to the suggested estimating model, thereby directing surplus economic resources toward other network interests (such as measuring emergent pollutants and chemical composition of particulate matter, among others).

As a derived application, the proposed model is also valid for evaluating the representativeness of fixed measuring benzene stations regarding the minimum number of sampling points for monitoring this pollutant in ambient air, which aligns with the requirements of European legislation.

This model exhibits a robust scientific base that completes the directrices set by European benzene monitoring through air quality networks legislation, thereby assisting national, regional, and municipal decision-making about future representative fixed stations. To highlight that the current legislation does not establish a harmonized methodology to assess the representativeness of fixed air pollutant stations.

In a broader context, the recommended model is not limited to a specific region. It may be potentially implemented in other geographical areas to estimate daily urban benzene levels in non-measured points using the target predictor variables. This possible global implementation will help decrease uncertainty and provide reliable data on the target pollutant in those areas, fostering a sense of inclusivity in the scientific community.

#### Funding

This research does not have commercial purposes, only scientific ones. This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

#### Author statement

The Authors certify that have seen and approved the final version of the manuscript being submitted. They warrant that the article is an authors' original work, hasn't received prior publication and isn't under consideration for publication elsewhere.

#### CRedit authorship contribution statement

**David Galán-Madruga:** Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization, Writing – review & editing, Writing – original draft. **Parya Broomandi:** Formal analysis, Writing – review & editing. **Jafet Cárdenas-Escudero:** Formal analysis, Writing – review & editing. **J.L. Urraca:** Investigation, Writing – review & editing. **Jorge O. Cáceres:** Supervision, Investigation, Writing – review & editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

## Acknowledgement

The authors thank the Madrid City's government for the data use and spread transparency policy, as well as Carlos III Health Institute and Complutense University of Madrid for providing adequate resources to the development of this research work.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.uclim.2024.102234>.

## References

- Aguilera, R., Sabater, S., Marcé, R., 2016. A methodological framework for characterizing the spatiotemporal variability of river water-quality patterns using dynamic factor analysis. *J. Environ. Inf.* 31 (2), 97–110.
- Andersson, E.M., Barregard, L., Akerstrom, M., Sallsten, G., Jarvholm, B., Nilsson, R.L., 2024. Cancer incidence in Swedish oil refinery workers exposed to benzene. *Int. J. Hyg. Environ. Health* 261, 114420.
- Atkinson, R., 2000. Atmospheric chemistry of VOCs and NOx. *Atmos. Environ.* 34, 2063–2101.
- Bozkurt, Z., Üzmez, Ö.Ö., Dögeroglu, T., Artun, G., Gaga, E.O., 2018. Atmospheric concentrations of SO<sub>2</sub>, NO<sub>2</sub>, ozone and VOCs in Düzce, Turkey using passive air samplers: sources, spatial and seasonal variations and health risk estimation. *Atmos. Pollut. Res.* 9, 1146–1156.
- Bravo, M.A., Fuentes, M., Zhang, Y., Burr, M.J., Bell, M.L., 2012. Comparison of exposure estimation methods for air pollutants: ambient monitoring data and regional air quality simulation. *Environ. Res.* 116, 1–10.
- Broomandi, P., Rodríguez-Sejjo, A., Janatian, N., Fathian, A., Tleuken, A., Mohammadpour, K., et al., 2023. Health risk assessment of the European inhabitants exposed to contaminated ambient particulate matter by potentially toxic elements. *Environ. Pollut.* 323, 121232.
- Carr, D., von Ehrenstein, O., Weiland, S., Wagner, C., Wellie, O., et al., 2002. Modeling annual benzene, toluene, NO<sub>2</sub>, and soot concentrations on the basis of road traffic characteristics. *Environ. Res.* 90, 111–118.
- Chang, S.Y., Vizuete, W., Valencia, A., Naess, B., Isakov, V., Palma, T., et al., 2015. A modeling framework for characterizing near-road air pollutant concentration at community scales. *Sci. Total Environ.* 538, 905–921.
- Christakos, G., Angulo, J.M., Yu, H.L., Wu, J., 2017. Space-time metric determination in environmental modeling. *J. Environ. Inf.* 30, 29–40.
- Christodoulakis, J., Karinou, F., Kelemen, M., Kouremadas, G., Fotaki, E.F., Varotsos, C.A., 2022. Assessment of air pollution from Athens international airport and suggestions for adaptation to new aviation emissions restrictions. *Atmos. Pollut. Res.* 13, 101441.
- Cintrón, M.L.B., Broomandi, P., Cárdenas-Escudero, J., Cáceres, J.O., Galán-Madruga, D., 2024. Elucidating best geospatial estimation method applied to environmental sciences. *Bull. Environ. Contam. Toxicol.* 112 (6).
- Costagliola, M.A., Murena, F., Prati, M.V., 2014. Exhaust emissions of volatile organic compounds of powered two-wheelers: effect of cold start and vehicle speed. Contribution to greenhouse effect and tropospheric ozone formation. *Sci. Total Environ.* 468–469, 1043–1049.
- Dai, H., Huang, G., Wang, J., Zeng, H., Zhou, F., 2022. Spatio-temporal characteristics of PM<sub>2.5</sub> concentrations in China based on multiple sources of data and LUR-GBM during 2016–2021. *IJERPH* 19, 6292.
- Dancey, C.P., Reidy, J., 2007. Statistics without Maths for Psychology.
- Dayan, U., Koch, J., Agami, S., 2023. Atmospheric conditions leading to buildup of benzene concentrations in urban areas in Israel. *Atmos. Environ.* 300, 119678.
- De la Paz, D., Borge, R., de Andrés, J.M., Tovar, L., Sarwar, G., et al., 2024. Summertime tropospheric ozone source apportionment study in the Madrid region (Spain). *Atmos. Chem. Phys.* 24, 4949–4972.
- Deng, Y., Li, J., Li, Y., Wu, R., Xie, S., 2019. Characteristics of volatile organic compounds, NO<sub>2</sub>, and effects on ozone formation at a site with high ozone level in Chengdu. *J. Environ. Sci.* 75, 334–345.
- Directive 2004/107/EC, 2024. Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 Relating to Arsenic, Cadmium, Mercury, Nickel and Polycyclic Aromatic Hydrocarbons in Ambient Air. available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32004L0107>. accessed on June 3.
- Directive 2008/50/EC, 2024. Directive 2008/50/EC of the European Parliament and of the Council on 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe. available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008L0050>. accessed on June 3.
- EN 14211, 2013. Ambient air - Standard Method for the Measurement of the Concentration of Nitrogen Dioxide and Nitrogen Monoxide by Chemiluminescence.
- EN 14625, 2013. Ambient Air - Standard Method for the Measurement of the Concentration of Ozone by Ultraviolet Photometry.
- EN 14662–1, 2005. Ambient Air Quality - Standard Method for Measurement of Benzene Concentrations - Part 1: Pumped Sampling Followed by Thermal Desorption and Gas Chromatography.
- Eurachem, 2012. Quantifying Uncertainty in Analytical Measurement. EURACHEM /CITAC Guide CG 4 3rd ed.
- Finlayson-Pitts, B.J., Pitts, J.N., 1993. Atmospheric chemistry of tropospheric ozone formation: scientific and regulatory implications. *Air Waste* 43, 1091–1100.
- Gaeta, A., Cattani, G., Menno, Di, di Bucchianico, A., De Santis, A., Cesaroni, G., et al., 2016. Development of nitrogen dioxide and volatile organic compounds land use regression models to estimate air pollution exposure near an Italian airport. *Atmos. Environ.* 131, 254–262.
- Galán-Madruga, D., Fernández Patier, R., Sintés Puertas, M.A., Romero García, M.D., Cristóbal-López, A., 2018. Characterization and local emission sources for Ammonia in an urban environment. *Bull. Environ. Contam. Toxicol.* 100, 593–599.
- Galán-Madruga, D., 2021. A methodological framework for improving air quality monitoring network layout. Applications to environment management. *J. Environ. Sci.* 102, 138–147.
- Galán-Madruga, D., 2023. Urban air quality changes resulting from the lockdown period due to the COVID-19 pandemic. *Int. J. Environ. Sci. Technol.* 20, 7083–7098.
- Galán-Madruga, D., Cárdenas-Escudero, J., Broomandi, P., Olenlacz, R., Cáceres, J.O., 2023. Performance assessment of air quality monitoring networks. A specific case study and methodological approach. *Air Qual. Atmos. Health* 16, 113–126.
- Galán-Madruga, D., Broomandi, P., Satyanaga, A., Jahanbakhshi, A., Bagheri, M., et al., 2025. A methodological framework for estimating ambient PM<sub>2.5</sub> particulate matter concentrations in the UK. *J. Environ. Sci.* 150, 676–691.
- Ghaffarpasand, O., Nadi, S., Shalamzari, Z.D., 2020. Short-term effects of anthropogenic/natural activities on the Tehran criteria air pollutants: source apportionment and spatiotemporal variation. *Build. Environ.* 186, 107298.
- Gómez-Moreno, F.J., Alonso-Blanco, E., Díez, E., Coz, E., Molero, F., Núñez, L., et al., 2022. On the influence of VOCs on new particle growth in a continental-Mediterranean region. *Environ. Res. Commun.* 4, 125010.
- Govender, P., Sivakumar, V., 2020. Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019). *Atmos. Pollut. Res.* 11, 40–56.
- Guigoz, Y., Giuliani, G., Nonguierma, A., Lehmann, A., Mlisa, A., Ray, N., 2017. Spatial data infrastructures in Africa: a gap analysis. *J. Environ. Inf.* 30 (1), 53–62.
- Hien, P.D., Hangartner, M., Fabian, S., Tan, P.M., 2014. Concentrations of NO<sub>2</sub>, SO<sub>2</sub>, and benzene across Hanoi measured by passive diffusion samplers. *Atmos. Environ.* 88, 66–73.

- Hsu, C.Y., Xie, H.X., Wong, P.Y., Chen, Y.C., Chen, P.C., et al., 2022. A mixed spatial prediction model in estimating spatiotemporal variations in benzene concentrations in Taiwan. *Chemosphere* 301, 134758.
- Huang, H., He, H., Fu, Q., Pan, J., Peng, Z., 2024. A deep learning model incorporating frequency domain information for ultra multi-step air pollutant forecasting: a case study of Shanghai. *Atmos. Pollut. Res.* 15, 102247.
- IARC, 2012. A Review of Human Carcinogens: Chemical Agents and Related Occupations, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, 100F. International Agency for Research on Cancer. available at <https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Chemical-Agents-And-Related-Occupations-2012>. accessed June 4, 2024.
- IARC, 2018. Benzene. IARC (Int. Agency Res. Cancer) Monogr. Eval. Carcinog. Risks Hum, 120 available at <https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Benzene-2018>. accessed July 29, 2024.
- Jacob, D.J., 2000. Heterogeneous chemistry and tropospheric ozone. *Atmos. Environ.* 34, 2131–2159.
- Jain, A.K., 2010. Data clustering: 50 years beyond K-means. *Pattern Recogn. Lett.* 31, 651–666.
- JCGM 100, 2008. Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement. BIPM available at [https://www.bipm.org/documents/20126/2071204/JCGM\\_100\\_2008\\_E.pdf](https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf). accessed on June 20, 2024.
- JCGM GUM-6, 2020. Guide to the Expression of Uncertainty in Measurement — Part 6: Developing and using Measurement Models. available at [https://www.bipm.org/documents/20126/2071204/JCGM\\_GUM\\_6\\_2020.pdf](https://www.bipm.org/documents/20126/2071204/JCGM_GUM_6_2020.pdf). accessed on June 20, 2024.
- Karunasingha, D.S.K., 2022. Root mean square error or mean absolute error? Use their ratio as well. *Inf. Sci.* 585, 609–629.
- Khuriganova, O.I., Obolkin, V.A., Golobokova, L.P., Bukin, Y.S., Khodzher, T.V., 2019. Passive sampling as a low-cost method for monitoring air pollutants in the Baikal region (eastern Siberia). *Atmos* 10 (8), 470.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. *Metz* 15, 259–263.
- Król, S., Zabiegała, B., Namieśnik, J., 2012. Measurement of benzene concentration in urban air using passive sampling. *Anal. Bioanal. Chem.* 403, 1067–1082.
- Kwon, J., Weisel, C.P., Morandi, M.T., Stock, T.H., 2016. Source proximity and meteorological effects on residential outdoor VOCs in urban areas: results from the Houston and Los Angeles RIOPA studies. *Sci. Total Environ.* 573, 954–964.
- Lin, C., Feng, X., Heal, M.R., 2016. Temporal persistence of intra-urban spatial contrasts in ambient NO<sub>2</sub>, O<sub>3</sub> and Ox in Edinburgh. *UK Atmos. Pollut. Res.* 7, 734–741.
- Liu, C., Zhang, X., Wang, Q., Shi, K., 2019. Role of PM<sub>2.5</sub> in the photodegradation of the atmospheric benzene. *Environ. Pollut.* 247, 447–456.
- Lu, X., Zhang, L., Shen, L., 2019. Meteorology and climate influences on tropospheric ozone: a review of natural sources, chemistry, and transport patterns. *Curr. Pollut. Rep.* 5, 238–260.
- Madruga, D.G., Ubeda, R.M., Terroba, J.M., dos Santos, S.G., García-Camero, J.P., 2019. Particle-associated polycyclic aromatic hydrocarbons in a representative urban location (indoor-outdoor) from South Europe: assessment of potential sources and cancer risk to humans. *Indoor Air* 29, 817–827.
- Mao, W., Wang, W., Jiao, L., Zhao, S., Liu, A., 2021. Modeling air quality prediction using a deep learning approach: method optimization and evaluation. *Sustain. Cities Soc.* 65, 102567.
- Marc, M., Namieśnik, J., Zabiegała, B., 2014. BTEX concentration levels in urban air in the area of the Tri-City agglomeration (Gdansk, Gdynia, Sopot), Poland. *Air Qual. Atmos. Health* 7, 489–504.
- Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., et al., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmos. Chem. Phys.* 15, 8889–8973.
- Pyae, T.S., Kallawicha, K., 2024. First temporal distribution model of ambient air pollutants (PM<sub>2.5</sub>, PM<sub>10</sub>, and O<sub>3</sub>) in Yangon City, Myanmar during 2019–2021. *Environ. Pollut.* 347, 123718.
- Qi, Y., Sun, C., Li, D., Zhang, H., Wang, H., Duan, Y., 2023. Does urban air pollution have an impact on public health? Empirical evidence from 288 prefecture-level cities in China. *Urban Clim.* 51, 101660.
- Reche, C., Moreno, T., Amato, F., Pandolfi, M., Pérez, J., et al., 2018. Spatio-temporal patterns of high summer ozone events in the Madrid Basin, Central Spain. *Atmos. Environ.* 185, 207–220.
- Salem, S.B., Naouali, S., Chtourou, Z., 2018. A fast and effective partitioning algorithm for large categorical datasets using a k-means based approach. *Comput. Electr. Eng.* 68, 463–483.
- Sassano, M., Seyyedsalehi, M.S., Boffeta, P., 2024. Occupational benzene exposure and colorectal cancer: a systematic review and meta-analysis. *Environ. Res.* 257, 119213.
- Schnitzhofer, R., Beauchamp, J., Dunkl, J., Wisthaler, A., Weber, A., et al., 2008. Long-term measurements of CO, NO, NO<sub>2</sub>, benzene, toluene and PM<sub>10</sub> at a motorway location in an Austrian valley. *Atmos. Environ.* 42, 1012–1024.
- Sharma, N., Taneja, S., Sagar, V., Bhatt, A., 2018. Forecasting air pollution load in Delhi using data analysis tools. *Procedia Comput. Sci.* 132, 1077–1085.
- Smith, M.T., 2010. Advances in understanding benzene health effects and susceptibility. *Annu. Rev. Public Health* 31, 133–148.
- Sowmya, H.N., Dinamani, M., Shivashankara, G.P., Ramaraju, H.K., Surendra, B.S., Shanbhag, V.V., Prajwal, R., 2024. A novel diurnal and seasonal variation analysis of BTEX in Bengaluru urban area. *Res. Cold Arid Reg.* 16, 14–19.
- Stettler, M.E.J., Eastham, S., Barrett, S.R.H., 2011. Air quality and public health impacts of UK airports. Part I: Emissions. *Atmos. Environ.* 45, 5415–5424.
- Su, J.G., Jerrett, M., Beckerman, B., Verma, D., Arain, M.A., Kanaroglou, P., et al., 2010. A land use regression model for predicting ambient volatile organic compound concentrations in Toronto. *Canada. Atmos. Environ.* 44, 3529–3537.
- Trebs, I., Lett, C., Krein, A., Junk, J., 2023. Air quality impacts of aviation activities at a mid-sized airport in Central Europe. *Atmos. Pollut. Res.* 14, 101696.
- Xu, Z., Nie, W., Chi, X., Huang, X., Zheng, L., et al., 2018. Ozone from fireworks: chemical processes or measurement interference? *Sci. Total Environ.* 633, 1007–1011.
- Yang, X., Cheng, S., Wang, G., Xu, R., Wang, X., et al., 2018. Characterization of volatile organic compounds and the impacts on the regional ozone at an international airport. *Environ. Pollut.* 238, 491–499.
- Zhai, H., Huang, L., Emery, C., Zhang, X., Wang, Y., et al., 2024. Recommendations on benchmarks for photochemical air quality model applications in China — NO<sub>2</sub>, SO<sub>2</sub>, CO and PM<sub>10</sub>. *Atmos. Environ.* 319, 120209.
- Zhang, H., Wang, J., Nie, Y., 2024. A novel optimization model based on fuzzy time series for short-term air quality index forecasting. *Knowl.-Based Syst.* 296, 111905.
- Zhao, L., Wang, X., He, Q., Wang, H., Sheng, G., Chan, L.Y., et al., 2004. Exposure to hazardous volatile organic compounds, PM<sub>10</sub> and CO while walking along streets in urban Guangzhou. *China. Atmos. Environ.* 38, 6177–6184.
- Zhao, H., Ge, Y., Tan, J., Yin, H., Guo, J., et al., 2011. Effects of different mixing ratios on emissions from passenger cars fueled with methanol/gasoline blends. *J. Environ. Sci.* 23, 1831–1838.
- Zhu, B., Cao, L.M., Xia, S.Y., Niu, Y.B., Man, H.Y., et al., 2023. Identifying the airport as a key urban VOC source in the Pearl River Delta, China. *Atmos. Environ.* 301, 119721.