


## Article

# The PM<sub>2.5</sub>-Bound Polycyclic Aromatic Hydrocarbon Behavior in Indoor and Outdoor Environments, Part III: Role of Environmental Settings in Elevating Indoor Concentrations of Benzo(a)pyrene

Gordana Jovanović<sup>1,2</sup>, Mirjana Perišić<sup>1,2</sup> , Timea Bezdán<sup>2</sup> , Svetlana Stanišić<sup>2</sup> , Kristina Radusin<sup>3</sup> , Aleksandar Popović<sup>3</sup>  and Andreja Stojić<sup>1,2,\*</sup>

- <sup>1</sup> Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia; gordana.jovanovic@ipb.ac.rs (G.J.); mirjana.perisic@ipb.ac.rs (M.P.)
- <sup>2</sup> Software and Information Engineering, Singidunum University, Danijelova 32, 11000 Belgrade, Serbia; tbezdán@singidunum.ac.rs (T.B.); sstanisic@singidunum.ac.rs (S.S.)
- <sup>3</sup> Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia; kristina.radusin@zdravlje.org.rs (K.R.); apopovic@chem.bg.ac.rs (A.P.)
- \* Correspondence: andreja.stojic@ipb.ac.rs

**Abstract:** This study aims to investigate the impact of indoor sources and outdoor air on indoor PM<sub>2.5</sub>-bound benzo(a)pyrene, with a focus on identifying emission sources and understanding the influence of environmental variables. For this purpose, we collected indoor and outdoor data on PM<sub>2.5</sub>-bound PAHs, inorganic gaseous pollutants, trace metals, ions, radon, and meteorological parameters, resulting in a comprehensive dataset of 100 variables from an urban site in Belgrade, Serbia. We applied seven regression tree ensemble algorithms to interrelate the variables alongside six metaheuristic optimization algorithms to refine model accuracy and robustness. Subsequently, we explained the best-performing model locally using Shapley additive explanations and clustered variables with similar impacts into distinct groups. These groups were systematically characterized, defining them as environmental settings that shape benzo(a)pyrene dynamics. The setting resulting in the highest indoor benzo(a)pyrene concentrations (197% to 297% relative to the expected value) was dominated by outdoor emissions associated with residential heating and traffic (up to 140%) and indoor source identified as cooking. This integrated approach uniquely enables a quantitative assessment of the contributions from both indoor and outdoor emission sources to pollutant concentrations in indoor spaces, underscoring the importance of both in shaping indoor air quality. Unlike traditional source apportionment methods that assume linear source mixing, our approach integrates nonlinear interactions and contextual variables, such as meteorological conditions and outdoor pollutants, to better capture indoor air quality dynamics. The results also highlight the need for further studies to explore broader contextual factors and refine source attribution in complex urban settings.

**Keywords:** benzo(a)pyrene; machine learning; metaheuristics; explainable artificial intelligence; AI-based source apportionment



**Citation:** Jovanović, G.; Perišić, M.; Bezdán, T.; Stanišić, S.; Radusin, K.; Popović, A.; Stojić, A. The PM<sub>2.5</sub>-Bound Polycyclic Aromatic Hydrocarbon Behavior in Indoor and Outdoor Environments, Part III: Role of Environmental Settings in Elevating Indoor Concentrations of Benzo(a)pyrene. *Atmosphere* **2024**, *15*, 1520. <https://doi.org/10.3390/atmos15121520>

Academic Editor: Célia dos Anjos Alves

Received: 13 November 2024

Revised: 16 December 2024

Accepted: 18 December 2024

Published: 19 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Polycyclic aromatic hydrocarbons (PAHs), and their representative benzo(a)pyrene (B[a]P), are ubiquitous environmental pollutants recognized as a substantial health risk for their carcinogenic and mutagenic properties. With individuals spending most of their time indoors, exposure to PAHs in enclosed environments is a growing concern. Recent studies have underscored the dual origin of indoor pollutants in general, as well as PAHs, whose concentrations are shaped by a complex interplay of outdoor and indoor pollution sources, environmental conditions, housing characteristics, and occupant behaviors [1,2].

The primary indoor sources of PAHs include cooking, residential heating, and tobacco smoking, with their impact on indoor air quality being especially pronounced in settings with limited ventilation [3]. Seasonal variations further influence the dynamics between outdoor and indoor pollutant levels, with colder months showing elevated indoor PAH concentrations due to restricted ventilation and increased use of heating fuels, which facilitate the infiltration and accumulation of pollutants indoor [4]. While coal combustion for cooking in developing countries is a major contributor to global premature mortality due to indoor air pollution, practices in those countries also adversely impact indoor air quality. For example, although gas stoves are considered cleaner than biomass stoves in many countries, they still contribute to indoor PAH emissions and associated health risks, particularly respiratory issues such as childhood asthma. A study by Gruenwald et al. [5] in the United States estimated that 12.7% of current childhood asthma cases could be attributed to domestic gas stove use, with variations observed across different states.

Also, Vardoulakis et al. [6] observed that indoor PAH levels can exceed outdoor levels, especially in homes with smoking, kerosene heating, or specific cooking practices. Thereby, specific emission sources contribute differently to indoor PAH levels, based on their molecular weight. It has been shown that low-molecular-weight PAHs are often associated with smoking, cooking and moth repellents, while high-molecular-weight PAHs predominantly originate from outdoor sources.

Even during significant outdoor pollution events, indoor PAH concentrations can surpass outdoor levels, as indoor environments trap pollutants due to limited ventilation and the persistence of indoor sources. As shown by the study of Ghetu et al. [2], who investigated vapor-phase PAHs in indoor and outdoor air before, during, and after wildfire events, indoor PAH concentrations were higher in 77% of samples across all sampling events, with 58% of locations showing increased indoor PAH levels even during wildfires.

Research on pollutant dynamics has conventionally centered on analyzing pollutant concentrations through Principal component analysis (PCA), Unmix, and Positive matrix factorization (PMF), to identify primary emission sources. It is widely acknowledged, however, that numerous factors beyond emissions, such as meteorology, human activities, topography, and various environmental aspects, significantly influence air quality and related health effects.

Traditional source apportionment methods, while useful, face notable limitations. These include the assumption of linear mixing, which oversimplifies the complexity of pollutant interactions, and the inability to capture nonlinear relationships or interactions between variables. Moreover, these methods fail to incorporate critical contextual variables, such as meteorological parameters and outdoor air pollutants, that profoundly affect indoor air quality.

To address this complexity, we introduced the concept of “environmental settings” which encompasses a broad range of natural and anthropogenic factors, which independently, interactively, or collectively impact the distribution and behavior of pollutants, offering a more comprehensive approach to understanding air quality dynamics beyond conventional source apportionment methods. The concept combines advanced machine learning optimized by metaheuristics applied to contextualize variables [7,8] and explainable artificial intelligence and clustering of the resulting local impacts, to group variables by similar impact characteristics that determine distinct environmental settings [9–11]. By resolving issues inherent to linear mixing assumptions and incorporating meteorological and outdoor factors, our approach contextualizes pollutant dynamics to reflect real-world conditions. The approach centered on environmental settings enables a more nuanced understanding of pollutant behavior under varying ambient conditions, whereby the same factor may exert different effects on pollutant depending on the surrounding environmental context [12,13].

Building on our prior research into PAH emission sources across indoor and outdoor environments [14,15], this study narrows its focus to B[a]P dynamics, particularly regarding the complex and under-researched interactions between indoor and outdoor environments.

While outdoor air pollution has been extensively studied, the exchange and behavior of B[a]P across these spaces remain insufficiently understood.

## 2. Methodology

A three-month measurement campaign (1 March–31 May 2016) was conducted at indoor and outdoor sites at Singidunum University in Belgrade, Serbia (44°45′33.8″ N, 20°29′47.6″ E). The experimental setup and all relevant methodological details are thoroughly explained in our previously published papers [14,15].

### 2.1. Data

Briefly, concentrations of inorganic gaseous pollutants, radon, PM<sub>2.5</sub>, and particle constituents (trace metals, ions, and PAHs) were measured. Meteorological data, including temperature, humidity, air pressure, wind, and rain characteristics, were recorded, with additional 24-parameter data from the Global data assimilation system (GDAS1). Air sampling was conducted on the rooftop and inside the university building, which hosts around 4000 students. Indoor sampling took place in an amphitheater with a 350-person capacity, where student numbers ranged from 50 to 80. The university is surrounded by residential areas with individual fireboxes (W, SW, NE) and nearby small-scale industries, including the Road Institute of Belgrade, a building company, and a beverage factory stockroom. Additional sources include a confectionery factory, footwear factory, and chemical plants (600 m NW and S), as well as a district heating plant and a fuel oil heating plant used by an urban forestry organization (800 m W and SW). A boulevard with moderate traffic lies 250 m SW, and a high-traffic road is 500 m W-NW. The old city center and river confluence are over 2 km NW, with occasional air quality impacts from emissions across the river.

Outdoor PM<sub>2.5</sub> and meteorological measurements were conducted at the rooftop (10 m above ground), while indoor air sampling was set 6 m from the amphitheater floor, with the PM<sub>2.5</sub> device positioned 2 m high. The air sampling system included a Pfeiffer MVP diaphragm vacuum pump (Pfeiffer Vacuum GmbH, Wetzlar, Germany) and manifolds with ports for inorganic gaseous pollutant measurement (O<sub>3</sub>, CO, SO<sub>2</sub>, NO<sub>x</sub>) using Horiba 370 series analyzers (Horiba, Ltd., Kyoto, Japan), and electronically controlled valves for alternating indoor/outdoor sampling in ten-minute cycles. PM<sub>2.5</sub> sampling was performed with Sven Leckel LVS6-RV devices (Sven Leckel, Ingenieurbüro GmbH, Berlin, Germany) at a flow rate of 2.3 m<sup>3</sup>/h over 24 h, while the concentrations and constituents (trace metals, ions, PAHs) were analyzed at the Institute of Public Health Belgrade. Outdoor meteorological data were collected via a Vaisala WXT530 station (Vaisala, Vantaa, Finland) on the rooftop, while indoor radon levels, temperature, humidity, and air pressure were monitored using an SN1029 radon monitor and sensors (Sun Nuclear Corporation, Melbourne, FL, USA), placed centrally in the amphitheater at a 1 m height.

PM<sub>2.5</sub> samples were collected daily on pre-fired quartz filters, with concentrations determined by gravimetric analysis following SRPS EN 12341:2015 [16]. Ion concentrations (Cl<sup>-</sup>, Na<sup>+</sup>, Mg<sup>2+</sup>, Ca<sup>2+</sup>, K<sup>+</sup>, NO<sub>3</sub><sup>-</sup>, SO<sub>4</sub><sup>2-</sup>, NH<sub>4</sub><sup>+</sup>) were measured via ion chromatography, while trace elements (As, Cd, Cr, Mn, Ni, Pb) were quantified by ICP-MS following SRPS EN 14902:2008/AC:2013 [17]. Sixteen US EPA PAHs were analyzed using GC-MS in accordance with SRPS ISO 12884:2010 [18] after microwave extraction, while inorganic gaseous pollutants O<sub>3</sub>, CO, SO<sub>2</sub>, NO<sub>x</sub> were monitored indoors and outdoors according to the standards SRPS EN 14625:2013, SRPS EN 14626:2013, SRPS EN 14212:2013/AC:2015, and SRPS EN 14211:2013 [19–22].

The only difference from the dataset used in our previous studies is that the instance recorded on 19 March 2016 was excluded from the analysis conducted in this study. This adjustment was made because the indoor concentrations of benzo(a)anthracene (B[a]A), chrysene (Chr), benzo(b)fluoranthene (B[b]F), benzo(k)fluoranthene (B[k]F), B[a]P, indeno(1,2,3-cd)pyrene (I[cd]P), dibenzo(a,h)anthracene (D[ah]A), and benzo(g,h,i)perylene (B[ghi]P) measured on that day were identified as outliers. In total, 100 indoor and outdoor variables

were included in the analysis to ensure comprehensive assessment of B[a]P dynamics across varying environmental conditions.

## 2.2. Data Analysis

Seven regression tree ensemble algorithms (AdaBoost, CatBoost, ExtraTrees, Gradient Boosting, Histogram Gradient Boosting, LightGBM, and XGBoost) were used to analyze data. These algorithms enhance prediction accuracy and robustness by combining multiple models to form a more precise ensemble, effectively reducing overfitting [23]. AdaBoost [24] refines model performance by focusing on previously misclassified instances. CatBoost [25] enhances gradient boosting by efficiently handling categorical features and minimizing overfitting with ordered boosting. ExtraTrees builds multiple decision trees by randomly selecting splits, thereby increasing model robustness and reducing variance [26]. LightGBM [27] uses gradient-based one-side sampling and exclusive feature bundling for computational efficiency. XGBoost [28] constructs trees in parallel with regularization to prevent overfitting, supporting various loss functions. Additionally, Gradient boosting and histogram gradient boosting (from Python's sklearn package [29]) improve modeling efficiency: gradient boosting builds models in a forward stage-wise manner, while histogram gradient boosting applies a binning method to reduce memory usage and accelerate training.

To ensure robust evaluation, we applied 5-fold cross-validation, maximizing dataset utilization for both training and validation to reduce overfitting risk. The top three performing models, identified based on r-squared values, underwent further refinement through metaheuristic optimization.

To optimize machine learning method hyperparameters, six metaheuristic algorithms were employed: Firefly Algorithm [30], Artificial Bee Colony [31], Harris Hawks Optimization [32], Sine Cosine Algorithm [33], Slime Mould Algorithm [34], and Quantum Search Algorithm [35]. Each algorithm efficiently navigates the search space to identify near-optimal hyperparameter values. The best-performing model was selected based on its r-squared value after hyperparameter optimization.

After identifying the best-performing model, we applied SHAP (SHapley Additive exPlanations [36]) to interpret the model's predictions. SHAP assigns each feature an importance value for a particular prediction, providing a unified measure of feature importance across a model, enabling comprehensive analysis of feature impact across the dataset. We further refined SHAP values into relative and normalized for clearer interpretation. Relative SHAP values represent the proportionate impact of each feature relative to all other features [14], while normalized SHAP values (adjusted to the expected value) simplify the understanding of impact magnitude.

To deepen the analysis of variable impacts and relationships, we applied dimensionality reduction with UMAP and clustering using HDBSCAN to the obtained SHAP values. UMAP [37] preserved both local and global data structures, making it suitable for complex datasets, while HDBSCAN [38] allowed for hierarchical clustering and the identification of clusters with varying densities. This combination enabled effective identification and categorization of localized impacts.

Based on data and observed impacts, we defined three ranges of normalized predictor levels—low (0–33%), medium (33–66%), and high (66–100%) of the variable's absolute value—and three ranges for normalized impacts: low (1–5%), medium (5–15%), and high (15–100%). Variables with high or medium impacts in a given cluster were identified as key determinants of that cluster, ensuring the meaningful grouping of variables. This ranking system enables the identification of variables that dominate specific environmental settings, offering a systematic approach to differentiate their roles in influencing pollutant dynamics.

## 3. Results and Discussion

According to the evaluation statistics, out of the top three best performing models, the extra trees algorithm optimized by the sine cosine algorithm provided the best MAPE-to-r-

squared ratio (Table 1). The model metrics include mean absolute error 0.05, mean squared error 0.01, root mean squared error 0.09, mean absolute percentage error 0.24, explained variance 0.98, max error 0.24, and r-squared 0.98.

**Table 1.** The top three best performing model evaluation statistics.

| Metrics                               | Gradient Boosting/<br>Harris Hawks Optimization | Extra Trees/<br>Sine Cosine Algorithm | XGBoost/<br>Firefly Algorithm |
|---------------------------------------|---|---------------------------------------|-------------------------------|
| Mean absolute error (MAE)             | 0.048   | 0.051                                 | 0.058                         |
| Mean squared error (MSE)              | 0.005   | 0.008                                 | 0.011                         |
| Root mean squared error (RMSE)        | 0.070   | 0.088                                 | 0.107                         |
| Mean absolute percentage error (MAPE) | 0.315   | 0.240                                 | 0.281                         |
| Explained variance                    | 0.986   | 0.979                                 | 0.969                         |
| Max error                             | 0.213   | 0.238                                 | 0.332                         |
| r-squared                             | 0.985   | 0.976                                 | 0.965                         |

### 3.1. Environmental Settings

In previous studies, we separately examined the impact of indoor and outdoor sources on B[a]P concentrations [14,15]. However, given that indoor concentrations of B[a]P averaged  $0.50 \text{ ng m}^{-3}$ , closely matching outdoor levels of  $0.48 \text{ ng m}^{-3}$ , and that there is a limited number of emission sources indoors, the question arose as to the extent to which pollutants from outdoor sources contribute to observed indoor B[a]P levels. To address this, we applied our methodology to a combined dataset from both outdoor and indoor environments, with the aim of distinguishing the respective contributions of outdoor and indoor sources to indoor B[a]P concentrations, and identifying the seasonal and meteorological conditions influencing the compound's dynamics.

By clustering the obtained local impacts, four distinct indoor environmental settings (E1–E4) responsible for the environmental fate of B[a]P were identified (Table 2). Unclustered instances, which may result from the combined influence of multiple sources and sinks, outliers, or transitional regimes, require more detailed analysis and are thus fall outside the scope of this study.

**Table 2.** Local impact clustering statistics.

| Environmental Setting | Mean Impact | Mean Normalized Impact [%] | Mean Absolute Impact | Population Percentage [%] |
|-----------------------|-------------|----------------------------|----------------------|---------------------------|
| Unclustered           | 0.01        | 2.9                        | 0.7                  | 27.0                      |
| E1                    | 1.18        | 262.9                      | 2.4                  | 16.8                      |
| E2                    | −0.10       | −22.8                      | 0.5                  | 24.7                      |
| E3                    | −0.27       | −59.7                      | 0.6                  | 11.0                      |
| E4                    | −0.34       | −76.2                      | 0.7                  | 20.5                      |

The mean absolute impacts, reflecting the overall influence of variables within each environmental setting on B[a]P dynamics, decreased in the following order: E1 ( $2.4 \text{ ng m}^{-3}$ ), E4 ( $0.7 \text{ ng m}^{-3}$ ), E3 ( $0.6 \text{ ng m}^{-3}$ ), and E2 ( $0.5 \text{ ng m}^{-3}$ ). Of the four identified settings, only E1 was associated with elevated B[a]P levels ( $1.18 \text{ ng m}^{-3}$ , 262.9% normalized to the expected value  $0.55 \text{ ng m}^{-3}$ ). Although E1 accounts for a relatively small portion of data instances (approximately 17%), its significance lies in revealing the specific conditions under which B[a]P concentrations notably increase. The elevated B[a]P concentrations observed in E1 were attributed to the combined influence of high indoor levels of B[b]F, B[k]F, I[cd]P, and B[a]A and infiltration of outdoor B[a]P from fossil fuel combustion. These conditions were likely exacerbated during colder periods due to increased heating activities in the vicinity, resulting in enhanced emissions of PAHs from residential heating and traffic sources. The elevated B[a]P levels in E1 are particularly concerning due to their implications for indoor air quality and potential health risks. Long-term exposure to B[a]P,



a known carcinogen, is associated with increased risks of respiratory illnesses and cancer. Therefore, E1 will be examined in detail in this study, as it offers valuable insights into the environmental dynamics and potential sources linked to elevated B[a]P levels. The remaining settings, comprising the other 55.2% of instances, contribute to the reduction of B[a]P levels normalized to the expected value  $0.55 \text{ ng m}^{-3}$ : E4 ( $-0.34 \text{ ng m}^{-3}$ ,  $-76.2\%$ ), E3 ( $-0.27 \text{ ng m}^{-3}$ ,  $-59.7\%$ ), and E2 ( $-0.1 \text{ ng m}^{-3}$ ,  $-22.8\%$ ).

It is important to note the distinction between an environmental setting and a single emission source, or more generally, between the methodological approach presented here and traditional source apportionment techniques. Compounds from the same source typically show temporal correlations, at least over short periods following their emission into the atmosphere. However, an environmental setting represents a broader context where a set of variables, beyond concentrations alone, jointly influences the dynamics of the target compound—in this case, B[a]P.

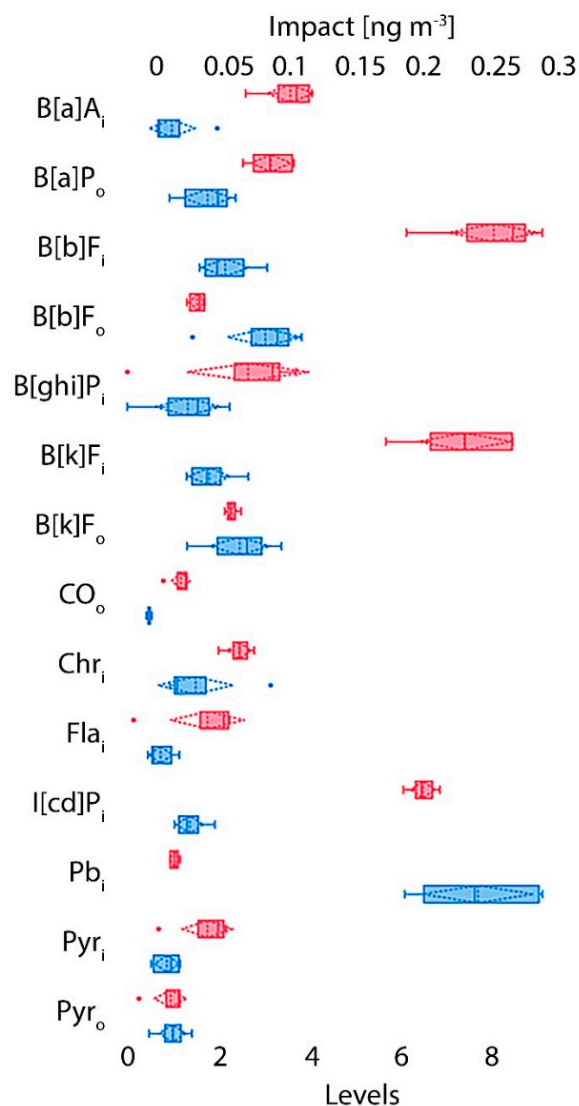
Moreover, individual environmental settings usually comprise a synergy of various sub-settings, with each one contributing differently to the observed pollutant levels. To elucidate B[a]P dynamics within a specific environmental sub-setting, it is essential to analyze the relationships and impacts of the most important variables in detail. The impacts of variables defining a sub-setting should correlate; that is, any change in the sub-setting's overall impact should be mirrored by changes in the impacts of its constituent variables. The term "constituent variables" refers to variables classified as having moderate or high impacts, according to the categorization in Section 2. In contrast, variable levels may not necessarily correlate, as they can exhibit independent relationships within the sub-setting, arising from shared emission sources, coexisting sources, or common physico-chemical interactions. If both impacts and levels correlate, however, this suggests that the sub-setting itself represents an emission source, as stable relationships between compound levels and impacts indicate a single, identifiable source. This setting-based approach allows for the identification of consistent patterns in target compound behavior, shaped by the combined effects of multiple variables linked to various sources within a shared environmental context.

### 3.2. Setting Associated with Elevated Indoor B[a]P Concentrations—E1

The distribution of the most important variable levels and impacts, which collectively constitute the environmental setting E1, is presented in Figure 1. Fluctuations in B[a]P levels relative to the expected value ranged from  $0.9 \text{ ng m}^{-3}$  (normalized impact 197.4%) to  $1.3 \text{ ng m}^{-3}$  (296.9%). These fluctuations were predominantly associated with both indoor and outdoor compounds, with notable correlations to indoor species, including B[b]F (relative impact 20.7%), B[k]F (18.9%), I[cd]P (16.6%), B[a]A (8.1%), B[ghi]P (5.7%), Chr (5.0%), pyrene (Pyr) (2.9%), fluoranthene (Fla) (2.8%), and Pb (1.0%). Outdoor-origin compounds, primarily B[a]P (7.0%), B[k]F (4.6%), B[b]F (2.4%), CO (1.4%), and Pyr (0.7%), also show a significant influence on these fluctuations. The influence of other variables was considered negligible.

The degree of impact on indoor B[a]P closely corresponded with the concentrations of five key variables, revealing a consistent alignment between high impacts and concentration levels for indoor B[b]F, B[k]F, I[cd]P, B[a]A, and outdoor-origin B[a]P. Conversely, moderate impacts were associated with high levels of indoor B[ghi]P, Chr, Fla, and Pyr, as well as outdoor-origin B[k]F and B[b]F. Low impacts were paired with high levels of outdoor Pyr and CO, and indoor Pb. The variety of these relationships suggests the presence of several sub-settings within E1, some of which may be attributed to specific emission sources.

Strong correlations in levels of indoor B[b]F, B[k]F, B[a]A, and Chr, with indoor B[a]P (Table 3), along with correlated impacts of these compounds on indoor B[a]P (Table 4), indicate that they originate from a shared emission source, likely linked to the incomplete combustion of organic materials [39]. This combination of high correlations in both levels and impacts points to the significant influence of this emission source on indoor air quality, with the potential to raise indoor B[a]P concentrations by up to 140%.



**Figure 1.** Statistical distribution (box-plot) of variable levels (blue) and impacts (red) characterizing the environmental setting E1 (medium and high impact variables only). The indexes next to the compound abbreviations indicate indoor (i) or outdoor (o) pollutant origin.

**Table 3.** Correlation of the most important parameter levels in E1. The indexes next to the compound abbreviations indicate indoor (i) or outdoor (o) pollutant origin.

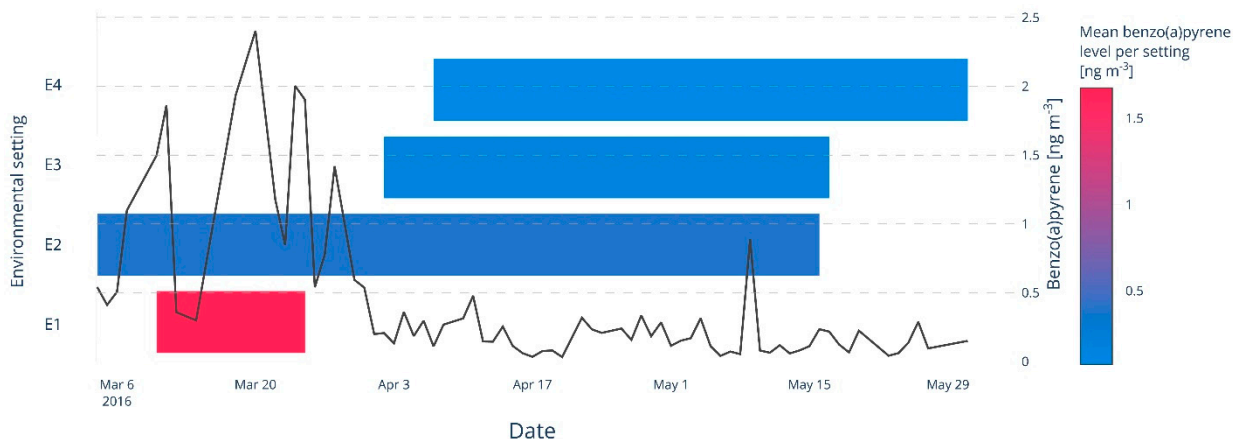
|                      | B[b]F <sub>i</sub> | B[k]F <sub>i</sub> | I[cd]P <sub>i</sub> | B[ghi]P <sub>i</sub> | B[a]A <sub>i</sub> | Chr <sub>i</sub> | B[a]P <sub>o</sub> | Fla <sub>i</sub> | Pyr <sub>i</sub> | B[k]F <sub>o</sub> | Pyr <sub>o</sub> | B[b]F <sub>o</sub> | CO <sub>o</sub> | Pb <sub>i</sub> |
|----------------------|--------------------|--------------------|---------------------|----------------------|--------------------|------------------|--------------------|------------------|------------------|--------------------|------------------|--------------------|-----------------|-----------------|
| B[b]F <sub>i</sub>   |                    |                    |                     |                      |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| B[k]F <sub>i</sub>   | 0.98               |                    |                     |                      |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| I[cd]P <sub>i</sub>  | 0.05               | −0.09              |                     |                      |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| B[ghi]P <sub>i</sub> | 0.26               | 0.11               | 0.53                |                      |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| B[a]A <sub>i</sub>   | 0.93               | 0.95               | −0.23               | 0.22                 |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| Chr <sub>i</sub>     | 0.92               | 0.95               | −0.28               | 0.17                 | 1                  |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| B[a]P <sub>o</sub>   | −0.54              | −0.58              | 0.17                | −0.18                | −0.75              | −0.73            |                    |                  |                  |                    |                  |                    |                 |                 |
| Fla <sub>i</sub>     | −0.33              | −0.26              | −0.56               | −0.59                | −0.35              | −0.3             | 0.7                |                  |                  |                    |                  |                    |                 |                 |
| Pyr <sub>i</sub>     | 0.4                | 0.5                | −0.81               | −0.46                | 0.49               | 0.54             | −0.05              | 0.63             |                  |                    |                  |                    |                 |                 |
| B[k]F <sub>o</sub>   | −0.71              | −0.72              | −0.01               | −0.3                 | −0.83              | −0.81            | 0.96               | 0.76             | −0.02            |                    |                  |                    |                 |                 |
| Pyr <sub>o</sub>     | 0.37               | 0.45               | −0.25               | −0.67                | 0.2                | 0.24             | 0.24               | 0.56             | 0.66             | 0.16               |                  |                    |                 |                 |
| B[b]F <sub>o</sub>   | −0.74              | −0.82              | 0.51                | 0.08                 | −0.92              | −0.93            | 0.82               | 0.24             | −0.59            | 0.81               | −0.23            |                    |                 |                 |
| CO <sub>o</sub>      | 0.82               | 0.86               | −0.42               | −0.12                | 0.79               | 0.82             | −0.24              | 0.25             | 0.84             | −0.35              | 0.67             | −0.72              |                 |                 |
| Pb <sub>i</sub>      | −0.71              | −0.59              | −0.34               | −0.7                 | −0.54              | −0.51            | 0.02               | 0.18             | −0.13            | 0.24               | −0.06            | 0.18               | −0.54           |                 |

**Table 4.** Correlation of the most important parameter impacts in E1. The indexes next to the compound abbreviations indicate indoor (i) or outdoor (o) pollutant origin.

|                      | B[b]F <sub>i</sub> | B[k]F <sub>i</sub> | I[cd]P <sub>i</sub> | B[ghi]P <sub>i</sub> | B[a]A <sub>i</sub> | Chr <sub>i</sub> | B[a]P <sub>o</sub> | Fla <sub>i</sub> | Pyr <sub>i</sub> | B[k]F <sub>o</sub> | Pyr <sub>o</sub> | B[b]F <sub>o</sub> | CO <sub>o</sub> | Pb <sub>i</sub> |
|----------------------|--------------------|--------------------|---------------------|----------------------|--------------------|------------------|--------------------|------------------|------------------|--------------------|------------------|--------------------|-----------------|-----------------|
| B[b]F <sub>i</sub>   |                    |                    |                     |                      |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| B[k]F <sub>i</sub>   | 0.91               |                    |                     |                      |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| I[cd]P <sub>i</sub>  | 0.41               | 0.46               |                     |                      |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| B[ghi]P <sub>i</sub> | −0.08              | −0.13              | 0.77                |                      |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| B[a]A <sub>i</sub>   | 0.99               | 0.94               | 0.39                | −0.12                |                    |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| Chr <sub>i</sub>     | 0.98               | 0.95               | 0.48                | −0.09                | 0.98               |                  |                    |                  |                  |                    |                  |                    |                 |                 |
| B[a]P <sub>o</sub>   | 0.59               | 0.45               | −0.09               | −0.09                | 0.62               | 0.44             |                    |                  |                  |                    |                  |                    |                 |                 |
| Fla <sub>i</sub>     | 0.93               | 0.82               | 0.07                | −0.35                | 0.94               | 0.87             | 0.76               |                  |                  |                    |                  |                    |                 |                 |
| Pyr <sub>i</sub>     | 0.94               | 0.79               | 0.1                 | −0.36                | 0.93               | 0.9              | 0.63               | 0.97             |                  |                    |                  |                    |                 |                 |
| B[k]F <sub>o</sub>   | 0.57               | 0.63               | 0.12                | −0.01                | 0.65               | 0.48             | 0.87               | 0.66             | 0.48             |                    |                  |                    |                 |                 |
| Pyr <sub>o</sub>     | 0.92               | 0.86               | 0.05                | −0.44                | 0.94               | 0.89             | 0.65               | 0.98             | 0.97             | 0.6                |                  |                    |                 |                 |
| B[b]F <sub>o</sub>   | 0.22               | 0.12               | −0.44               | −0.25                | 0.27               | 0.06             | 0.9                | 0.49             | 0.33             | 0.76               | 0.4              |                    |                 |                 |
| CO <sub>o</sub>      | 0.97               | 0.85               | 0.22                | −0.21                | 0.97               | 0.92             | 0.73               | 0.99             | 0.97             | 0.65               | 0.97             | 0.42               |                 |                 |
| Pb <sub>i</sub>      | −0.04              | 0.14               | 0.79                | 0.57                 | −0.06              | 0.11             | −0.66              | −0.39            | −0.31            | −0.35              | −0.32            | −0.85              | −0.27           |                 |

Among indoor sources containing these compounds, the university kitchen adjacent to the amphitheater, although the only significant indoor source of B[a]P, has been assessed to have a relatively minor contribution, as high-temperature cooking techniques, such as grilling and frying, release PAH compounds [40].

Given the absence of other indoor combustion sources and the building’s use of electric boilers and heat pumps, it can be inferred that the origin of these compounds lies in fossil fuel combustion in the surrounding outdoor environment, with pollutants subsequently infiltrating the indoor space. Outdoor emissions can permeate through ventilation systems, structural cracks, and other openings, thereby influencing indoor air quality. During colder seasons, such as the period associated with E1 (Figure 2), heightened use of heating fuels in the vicinity correlates with increased indoor levels of these PAHs.



**Figure 2.** Environmental settings timeline.

Correlations between indoor and outdoor levels of B[k]F, B[b]F, and B[a]P in E1 further substantiate this conclusion, suggesting that outdoor sources contribute substantially to indoor B[a]P concentrations. This indoor-outdoor association might be more readily apparent if not for the behavior of these “infiltrated” compounds, which likely reflect residual concentrations from previous episodes of outdoor air infiltration, with a lag period equal or exceeding one day (measurement frequency). Namely, a shorter lag would likely result in stronger correlations between indoor and outdoor levels for each compound individually; however, the observed dynamics suggest a gradual accumulation of about one day as confirmed by the cross-correlation analysis rather than an immediate reflection of outdoor levels (correlation coefficients for indoor and outdoor B[k]F, B[b]F, and B[a]P



in E1 provided in Table 3 were approximately equal to their cross-correlation coefficients for lag 1). This gradual accumulation implies that outdoor B[a]P contributions to indoor concentrations are moderated by the infiltration process, which delays the direct impact of outdoor emissions and emphasizes the importance of indoor retention and re-suspension of particles.

The degradation rate of PAHs in outdoor environments varies significantly based on the specific PAH compound, the presence of oxidants such as ozone (O<sub>3</sub>), hydroxyl radicals (•OH), and nitrogen dioxide (NO<sub>2</sub>), as well as environmental factors like UV radiation, temperature, and humidity [41]. On average, the half-life of PAHs in outdoor environment ranges from a few hours to several days. Lower molecular weight PAHs, such as naphthalene, are more reactive and can degrade within hours to a day, while higher molecular weight PAHs, such as B[a]P, tend to be more stable, with half-lives extending from one to several days. The fastest degradation occurs during sunny conditions with high levels of UV radiation and oxidants, as these factors accelerate the transformation of PAHs into oxidized derivatives. However, in indoor environments, PAH compounds tend to remain stable and unaltered due to the absence of key reactive factors present in outdoor settings. Firstly, oxidants such as ozone (O<sub>3</sub>), hydroxyl radicals (•OH), and nitrogen dioxide (NO<sub>2</sub>), which facilitate PAH transformations in the atmosphere, are typically present at much lower concentrations indoors, significantly reducing the potential for oxidation and chemical transformation of PAHs. Secondly, photolytic reactions, which are common outdoors due to UV radiation from sunlight, rarely occur indoors due to limited exposure to UV light. In enclosed spaces, any UV light exposure is minimal unless there is direct sunlight through windows, further limiting photolytic transformations. As a result, while outdoor environments expose PAHs to reactive and photochemical conditions that can lead to transformations, indoor spaces lack these factors, allowing PAH compounds to persist largely unchanged, aside from gradual physical processes such as deposition on surfaces.

Additionally, in indoor environments, deposition processes play a crucial role in the accumulation and persistence of PAHs. When outdoor pollutants infiltrate indoor spaces, higher molecular weight PAH compounds, such as Pyr, tend to adhere more readily to various surfaces, including walls, furniture, and ventilation ducts, due to their hydrophobic nature and lower volatility. This creates a “reservoir effect”, where deposited PAHs gradually accumulate on indoor surfaces and, over time, can re-enter the air influenced by factors such as temperature fluctuations, airflow, and mechanical disturbances. In contrast, CO, due to its gaseous state and chemical stability, remains predominantly in the air and does not contribute to surface deposition. Consequently, while CO levels decline as ventilation or infiltration varies, PAHs bound to surfaces continue to be re-emitted, sustaining elevated indoor concentrations over extended periods and contributing to prolonged exposure risks in indoor settings.

The correlation of indoor B[a]P with I[cd]P and B[ghi]P suggests an influence from high-temperature combustion processes, particularly fossil fuel burning. These specific compounds are commonly associated with the incomplete combustion of heavy fossil fuels, such as fuel oil, frequently used for heating in nearby buildings during colder seasons. The presence of I[cd]P and B[ghi]P, in conjunction with B[a]P, is characteristic of emissions from industrial boilers, heating systems, and diesel engines, indicating contributions from both local heating activities and potential traffic sources. This pattern of PAH compounds supports the conclusion that nearby fossil fuel combustion is a significant contributor to the observed indoor B[a]P levels. The observed correlations and cross-correlation analysis highlight how outdoor B[a]P contributions are moderated by infiltration dynamics, with retention and surface deposition further sustaining elevated indoor concentrations beyond immediate outdoor influences.

#### 4. Limitations

This study has several limitations. Its findings are based on data collected under specific environmental conditions over a defined period in Belgrade, Serbia, and further

research is needed to validate these results across different contexts for a broader understanding of B[a]P dynamics. Furthermore, extending the research to other locations and incorporating longer time series data, covering various seasons and a wider range of emission sources, would provide a more comprehensive insight into seasonal variations and source contributions. While the study accounts for various factors represented by 100 variables, it does not incorporate other significant influences, such as long-range pollutant transport and regional topographical differences, which could impact air quality and B[a]P environmental fate. Expanding the dataset to include these variables could provide a more comprehensive analysis. The discrepancies observed among certain variables within E1 indicate the existence of several sub-settings, each potentially associated with specific emission sources. While further sub-clustering could provide more detailed insights, the limited amount of data in this study restricts this possibility. Future research with expanded datasets could enable a more granular analysis, allowing for the identification of distinct sub-settings and their corresponding sources. Additionally, although the use of ensemble machine learning algorithms, metaheuristics, dimensionality reduction, and clustering is an innovative approach, the study does not compare these methods with other modeling techniques—such as recurrent neural networks, hybrid metaheuristics, or alternative dimensionality reduction and clustering methods—which could enhance the generalizability of the findings. Lastly, further refining the identified emission sources and sinks with advanced statistical and artificial intelligence techniques would strengthen the robustness of the conclusions.

## 5. Conclusions

This study provides a detailed assessment of elevated indoor B[a]P levels within the context of PM<sub>2.5</sub>, highlighting the complex interactions that shape its environmental fate. Among the machine learning algorithms applied, the ExtraTrees method optimized by the Sine cosine algorithm emerged as the best performer, achieving a high r-squared value alongside low mean absolute percentage error and root mean squared error values, demonstrating robust predictive capability for air quality data.

By clustering local impacts, four distinct indoor environmental settings (E1–E4) associated with B[a]P dynamics were identified, each characterized by varying mean absolute impacts on B[a]P levels. The only setting linked with elevated B[a]P concentrations, accounted for 17% of events, indicating a limited yet significant influence of these conditions on indoor air quality.

The environmental setting E1 is characterized by specific variable levels and impacts, resulting in notable fluctuations in B[a]P concentrations from 0.9 to 1.3 ng m<sup>-3</sup>. These fluctuations are influenced by both indoor and outdoor compounds, particularly high levels of indoor B[b]F, B[k]F, I[cd]P, B[a]A, and outdoor B[a]P, suggesting shared sources potentially linked to incomplete combustion of organic materials. Although the university kitchen is the only significant indoor source of B[a]P, its contribution remains minor, with PAH levels primarily attributed to infiltration from outdoor fossil fuel combustion sources. This influence is more pronounced during colder periods, corresponding with increased local heating activities. Additionally, interrelations between indoor and outdoor levels of B[a]P, B[k]F, and B[b]F support the conclusion that outdoor sources substantially impact indoor air quality. This association may reflect the accumulation of residual PAH concentrations indoors due to infiltration with a lag period exceeding one day. Outdoors, PAHs degrade relatively quickly due to UV exposure and oxidants, but indoor PAH levels remain stable in the absence of these reactive agents, allowing them to persist through surface deposition. Unlike PAHs, CO levels decline more rapidly indoors, highlighting the sustained correlation of PAHs with outdoor sources. Furthermore, the correlation of indoor B[a]P with I[cd]P and B[ghi]P indicates significant contributions from nearby fossil fuel combustion used for heating, underlining the influence of local combustion activities on indoor PAH levels.

This integrated approach, which incorporates machine learning, metaheuristics, explainable artificial intelligence, and further treatment of the obtained impacts for interpreting complex environmental settings, underscores the critical role of comprehensive contextual analysis in understanding pollutant dynamics and offers a framework for future studies on indoor air quality. The flexibility of this methodology allows its application to diverse geographical and structural environments, provided the contextual variables appropriately describe the pollutant of interest, thus enabling meaningful clustering and variable-impact analyses. Further research with extended datasets and additional contextual variables is recommended to refine these insights and validate findings across broader settings.

**Author Contributions:** G.J.: conceptualization, formal analysis, funding acquisition, investigation, writing—original draft, review, and editing. M.P.: conceptualization, data curation, formal analysis, funding acquisition, investigation, writing—original draft, review, and editing, project administration. T.B.: methodology, resources, software. S.S.: investigation, writing—original draft, review, and editing. K.R.: investigation, writing—original draft, review, and editing. A.P.: conceptualization, writing—original draft, review. A.S.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, review, and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, as well as by the Science Fund of the Republic of Serbia, Grant No. #7373, Characterizing crises-caused air pollution alternations using an artificial intelligence-based framework—crAIRsis.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zhang, J.; Wang, Z.; Wei, Y.; Yang, S.; Yao, S.; Yang, B.; Yang, L. Characteristics, sources, and health risks of PAHs and their derivatives in indoor dust in Zhengzhou. *Atmos. Pollut. Res.* **2024**, *15*, 102246. [[CrossRef](#)]
2. Ghetu, C.C.; Rohlman, D.; Smith, B.W.; Scott, R.P.; Adams, K.A.; Hoffman, P.D.; Anderson, K.A. Wildfire impact on indoor and outdoor PAH air quality. *Environ. Sci. Technol.* **2022**, *56*, 10042–10052. [[CrossRef](#)] [[PubMed](#)]
3. WHO. *Human Health Effects of Polycyclic Aromatic Hydrocarbons as Ambient Air Pollutants: Report of the Working Group on Polycyclic Aromatic Hydrocarbons of the Joint Task Force on the Health Aspects of Air Pollution*; World Health Organization Regional Office for Europe: Copenhagen, Denmark, 2021.
4. Li, C.; Bai, L.; Wang, H.; Li, G.; Cui, Y. Characteristics of indoor and outdoor Polycyclic Aromatic Hydrocarbons (PAHs) pollution in TSP in rural Northeast China: A case study of heating and non-heating periods. *J. Environ. Health Sci. Eng.* **2022**, *20*, 899–913. [[CrossRef](#)] [[PubMed](#)]
5. Gruenewald, T.; Seals, B.A.; Knibbs, L.D.; Hosgood, H.D. Population attributable fraction of gas stoves and childhood asthma in the United States. *Int. J. Environ. Res. Public Health* **2023**, *20*, 75. [[CrossRef](#)]
6. Vardoulakis, S.; Giagloglou, E.; Steinle, S.; Davis, A.; Sleenwenhoek, A.; Galea, K.S.; Dixon, K.; Crawford, J.O. Indoor exposure to selected air pollutants in the home environment: A systematic review. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8972. [[CrossRef](#)]
7. Bacanin, N.; Perisic, M.; Jovanovic, G.; Damaševićus, R.; Stanisic, S.; Simic, V.; Zivkovic, M.; Stojic, A. The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere. *Sci. Total Environ.* **2024**, *929*, 172195. [[CrossRef](#)]
8. Bezdan, T.; Perišić, M.; Jovanović, G.; Bačanin, N.; Stojić, A. Artificial Intelligence-Based Framework for Analyzing Crises-Caused Air Pollution. In *Sinteza 2024—International Scientific Conference on Information Technology, Computer Science, and Data Science*; Singidunum University: Beograd, Serbia, 2024; pp. 281–287.
9. Stojić, A.; Vuković, G.; Perišić, M.; Stanišić, S.; Šošarić, A. Urban air pollution: An insight into its complex aspects. In *A Closer Look at Urban Areas*; Nova Science Publishers: New York, NY, USA, 2018.
10. Stojić, A.; Stanić, N.; Vuković, G.; Stanišić, S.; Perišić, M.; Šošarić, A.; Lazić, L. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Sci. Total Environ.* **2019**, *653*, 140–147. [[CrossRef](#)] [[PubMed](#)]

11. Stanišić, S.; Jovanović, G.; Perišić, M.; Herceg Romanić, S.; Milićević, T.; Stojić, A. *Explaining the Environmental Fate of PAHs in Indoor and Outdoor Environments by the Use of Artificial Intelligence*; Nova Science Publishers: New York, NY, USA, 2022.
12. Jovanovic, G.; Perisic, M.; Bacanin, N.; Zivkovic, M.; Stanisic, S.; Strumberger, I.; Alimpic, F.; Stojic, A. Potential of coupling metaheuristics-optimized-xgboost and shap in revealing pahs environmental fate. *Toxics* **2023**, *11*, 394. [CrossRef] [PubMed]
13. Jovanovic, L.; Jovanovic, G.; Perisic, M.; Alimpic, F.; Stanisic, S.; Bacanin, N.; Zivkovic, M.; Stojic, A. The explainable potential of coupling metaheuristics-optimized-xgboost and shap in revealing vocs' environmental fate. *Atmosphere* **2023**, *14*, 109. [CrossRef]
14. Stojić, A.; Jovanović, G.; Stanišić, S.; Romanić, S.H.; Šoštarić, A.; Udovičić, V.; Perišić, M.; Milićević, T. The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo [a] pyrene levels. *Chemosphere* **2022**, *289*, 133154. [CrossRef] [PubMed]
15. Stanišić, S.; Perišić, M.; Jovanović, G.; Milićević, T.; Romanić, S.H.; Jovanović, A.; Šoštarić, A.; Udovičić, V.; Stojić, A. The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: Emission sources. *Environ. Res.* **2021**, *193*, 110520. [CrossRef] [PubMed]
16. SRPS EN 12341:2015; Ambient Air—Standard Gravimetric Measurement Method for the Determination of the PM10 or PM2,5 Mass Concentration of Suspended Particulate Matter. Available online: <https://iss.rs/en/project/show/iss:proj:49389> (accessed on 29 May 2015).
17. SRPS EN 14902:2008/AC:2013; Ambient Air—Standard Method for the Measurement of Pb, Cd, As and Ni in the PM10 Fraction of Suspended Particulate Matter. Available online: <https://iss.rs/en/project/show/iss:proj:46826> (accessed on 27 August 2013).
18. SRPS ISO 12884:2010; Ambient air—Determination of Total (Gas and Particle-Phase) Polycyclic Aromatic Hydrocarbons—Collection on Sorbent-Backed Filters with Gas Chromatographic/Mass Spectrometric Analyses. Available online: <https://iss.rs/en/project/show/iss:proj:24983> (accessed on 29 January 2010).
19. SRPS EN 14625:2013; Ambient Air—Standard Method for the Measurement of the Concentration of Ozone by Ultraviolet Photometry. Available online: <https://iss.rs/en/project/show/iss:proj:40991> (accessed on 31 May 2013).
20. SRPS EN 14626:2013; Ambient Air—Standard Method for the Measurement of the Concentration of Carbon Monoxide by Non-Dispersive Infrared Spectroscopy. Available online: <https://iss.rs/en/project/show/iss:proj:40993> (accessed on 31 May 2013).
21. SRPS EN 14212:2013/AC:2015; Ambient Air—Standard Method for the Measurement of the Concentration of Sulphur Dioxide by Ultraviolet Fluorescence. Available online: <https://iss.rs/en/project/show/iss:proj:54201> (accessed on 31 August 2015).
22. SRPS EN 14211:2013; Ambient Air—Ambient Air—Standard Method for the Measurement of the Concentration of Nitrogen Dioxide and Nitrogen Monoxide by Chemiluminescence. Available online: <https://iss.rs/en/project/show/iss:proj:40985> (accessed on 31 May 2013).
23. Dietterich, T.G. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2000.
24. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
25. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*. Available online: <https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html> (accessed on 12 November 2024).
26. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
27. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html> (accessed on 12 November 2024).
28. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 118–1232. [CrossRef]
29. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
30. Yang, X.S. Firefly algorithms for multimodal optimization. In *International Symposium on Stochastic Algorithms*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 169–178.
31. Karaboga, D. *An Idea Based on Honey Bee Swarm for Numerical Optimization*; Technical Report-tr06; Erciyes University, Engineering Faculty, Computer Engineering Department: Kayseri, Türkiye, 2005; Volume 200, pp. 1–10.
32. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [CrossRef]
33. Mirjalili, S. SCA: A sine cosine algorithm for solving optimization problems. *Knowl.-Based Syst.* **2016**, *96*, 120–133. [CrossRef]
34. Li, S.; Chen, H.; Wang, M.; Heidari, A.A.; Mirjalili, S. Slime mould algorithm: A new method for stochastic optimization. *Future Gener. Comput. Syst.* **2020**, *111*, 300–323. [CrossRef]
35. Zhang, J.; Xiao, M.; Gao, L.; Pan, Q. Queuing search algorithm: A novel metaheuristic algorithm for solving engineering optimization problems. *Appl. Math. Model.* **2018**, *63*, 464–490. [CrossRef]
36. Lundberg, S. A unified approach to interpreting model predictions. *arXiv* **2017**, arXiv:1705.07874.
37. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
38. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density-based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [CrossRef]

39. Abdel-Shafy, H.I.; Mansour, M.S. A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egypt. J. Pet.* **2016**, *25*, 107–123. [[CrossRef](#)]
40. Sampaio, G.R.; Guizzellini, G.M.; da Silva, S.A.; de Almeida, A.P.; Pinaffi-Langley, A.C.C.; Rogero, M.M.; de Camargo, A.C.; Torres, E.A. Polycyclic aromatic hydrocarbons in foods: Biological effects, legislation, occurrence, analytical methods, and strategies to reduce their formation. *Int. J. Mol. Sci.* **2021**, *22*, 6010. [[CrossRef](#)] [[PubMed](#)]
41. Gbeddy, G.; Goonetilleke, A.; Ayoko, G.A.; Egodawatta, P. Transformation and degradation of polycyclic aromatic hydrocarbons (PAHs) in urban road surfaces: Influential factors, implications and recommendations. *Environ. Pollut.* **2020**, *257*, 113510. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.