INTERPRETABLE MODELLING OF BENZENE VARIABILITY ACROSS EUROPE: FROM SHAP DEPENDENCE TO TEMPERATURE REGIMES

Timea Bezdan[1]*,
[0000-0001-6938-6974]

Gabriel Joseph Isibor[2,3],
[0000-0002-3414-3574]

Gordana Jovanović[4],
[0000-0001-8657-423X]

Andreja Stojić[1,4],
[0000-0002-5293-9533]

Mirjana Perišić[1,4]
[0000-0002-8287-4136]

[1]Singidunum University,
Belgrade, Serbia

[2]Faculty of Physics,
University of Belgrade,
Belgrade, Serbia

[3]Department of Radiology,
Medical Physics Unit,
Keffi, Nigeria

[4]Institute of Physics Belgrade,
Belgrade, Serbia

Correspondence:
Timea Bezdan

e-mail:
tbezdan@singidunum.ac.rs

Abstract:

This study presents an integrated machine learning (ML) and explainable artificial intelligence (XAI) framework for modelling daily benzene concentrations across 84 monitoring sites in Europe during February 2020-October 2021. A suite of ensemble tree-based algorithms was evaluated, with metaheuristic optimization used to refine model performance; the optimized LightGBM model achieved the best agreement with observations ($R2 \approx 0.85$). To interpret the model, SHAP (Shapley Additive Explanations) was applied at both global and local levels. In addition to conventional SHAP analysis, SHAP-based clustering was used to identify distinct environmental settings, and the temperature dependence of SHAP values was parameterized using a segment-wise linear approximation. The results show strong spatial heterogeneity in benzene levels and in the magnitude of temperature importance across Europe. At the same time, the signed SHAP dependence for 2 m air temperature (T02m) exhibits a remarkably coherent cross-site structure: positive contributions under cold conditions, a progressive decline toward a zero crossing around 9.5 °C, a broad negative regime through mild and warm conditions, and a weak rebound at the highest temperatures. This indicates that temperature acts less as an isolated driver than as a compact state variable that reflects seasonal accumulation, dispersion, oxidation, and temperature-sensitive emission processes. The main novelty of the study is therefore not the use of SHAP alone, but the extraction of transferable regime parameters from SHAP structure. The proposed framework combines predictive skill with interpretability and can be extended to other predictors and interactions in large-scale air-quality applications.

Keywords:

Machine Learning, Metaheuristics, Explainable Artificial Intelligence, Air Quality, Benzene.

INTRODUCTION

Machine learning (ML) offers a powerful framework for modelling complex environmental systems characterized by nonlinear responses, heterogeneous predictors, and strong interactions among variables [1, 2, 3, 4]. In air-quality applications, ensemble tree-based methods are especially attractive because they can accommodate mixed predictor types and capture thresholds and non-additive effects that are difficult to represent with conventional linear models [1, 3]. Their main limitation is interpretability.

DOI: 10.15308/Sinteza-2026-226-232

Explainable artificial intelligence (XAI) tools such as SHAP partially address this problem by decomposing each prediction into additive feature contributions and by enabling functional inspection of predictor-target relationships [4].

Benzene is a particularly relevant target for such analysis. It is a toxic and carcinogenic volatile organic compound emitted by traffic, fuel handling, solvent-related activities, combustion, and industrial processes [5, 6]. In ambient air, benzene resides predominantly in the gas phase and is removed mainly by reaction with hydroxyl radicals, with atmospheric residence times ranging from about one day to two weeks depending on conditions [6]. Its concentrations therefore reflect a combination of source strength, atmospheric mixing, and chemical loss, making benzene a suitable test case for interpretable ML in heterogeneous environmental settings.

The pandemic period created an unusual large-scale natural experiment in which changes in traffic, economic activity, mobility, and public-health restrictions occurred simultaneously across Europe [7, 8, 9, 10]. This setting is methodologically challenging because pollutant concentrations respond not only to direct changes in emissions but also to meteorology, regional transport, and site-specific context. Recent work has shown that the same predictor can have different statistical roles under different environmental settings, reinforcing the need for interpretable approaches that preserve contextual information [11, 12].

In this study, daily benzene concentrations from 84 monitoring sites across Europe were modelled using a suite of ensemble ML algorithms, with metaheuristic optimization applied to the best-performing models. SHAP values were then used to quantify predictor effects and to explore their nonlinear structure. Within the broad set of results, this conference paper concentrates on one representative and physically meaningful example: the relationship between 2 m air temperature (T02m) and benzene. The methodological contribution is deliberately framed in a narrow but useful way. The novelty is not the use of SHAP by itself, but the conversion of a SHAP dependence pattern into a compact segment-wise regime model with interpretable transition temperatures and fitted parameters. In other words, the dependence plot is treated not as a final visualization, but as an object for secondary modelling.

2. METHODOLOGY

2.1.DATA

The dataset comprises daily averaged benzene concentrations from 84 air-quality monitoring sites across Europe for the period 15 February 2020 to 10 October 2021. The network spans diverse climatic, urban, and emission contexts. Predictors include co-pollutants (NO_2, O_3, SO_2, PM_{10}), meteorological variables from the GDAS1 reanalysis product [13], spatial descriptors (e.g. latitude and longitude), and proxies of anthropogenic activity such as policy indices, epidemiological indicators, and mobility-related metrics [8, 9, 10]. This design allows the model to represent not only immediate pollutant-meteorology relations but also broader societal modulation during the pandemic period.

2.2. MODELLING AND INTERPRETABILITY WORKFLOW

Benzene concentrations were modelled using tree-based ensemble algorithms including AdaBoost, Extra Trees, Gradient Boosting, Histogram-based Gradient Boosting, CatBoost, LightGBM, and XGBoost [1, 2, 3]. Model performance was evaluated by cross-validation, with R2 used as the main selection criterion. The best-performing algorithms were further refined using Sine Cosine Algorithm (SCA) and Harris Hawks Optimization (HHO) to explore the hyperparameter space [14, 15].

Interpretation relied on SHAP values [4]. Two different SHAP-derived views are important in this paper and must be distinguished explicitly. First, the right panel of Figure 2 shows a site-aggregated mean absolute SHAP magnitude for T02m and therefore represents the spatial magnitude of temperature importance only; it does not encode whether temperature increases or decreases the benzene prediction. Second, Figure 3 shows observation-level signed SHAP values and therefore captures the direction and functional form of the T02m effect. This distinction is essential for correct interpretation.

To identify broader contextual structure, the full SHAP representation was reduced using PaCMAP and clustered using HDBSCAN [16, 17], yielding groups of observations with similar contribution patterns across all predictors. These clusters are interpreted as environmental settings rather than strict source classes.

Finally, the signed T02m-SHAP dependence was approximated by a segment-wise linear model of the form $\varphi_T(T) = a_s \cdot T + b_s$, for T within segment s.

In the exported plots, the original meteorological variable label T02M is retained and the x-axis is shown in the native GDAS temperature scale (K). In Table 1, the corresponding segment limits are reported in °C, and the intercepts are converted so that the tabulated parameters are algebraically consistent with that Celsius scale. Because a shift from K to °C is translational, slopes are unchanged whereas intercepts must be adjusted.

3. RESULTS AND DISCUSSION

3.1. PREDICTIVE PERFORMANCE

Figure 1 summarizes the predictive behaviour of the tested models. The optimized LightGBM model gives the highest R^2 and the most convincing agreement between observed and predicted benzene values, confirming its ability to reproduce both the spread and the central tendency of the dataset. The optimization trajectories also show stable convergence for the selected LightGBM configurations.

The comparative bar chart reveals an important nuance: metaheuristic optimization is helpful, but not uniformly so. The gains are strongest for LightGBM and noticeable for HistGradientBoosting, whereas Extra Trees changes little relative to its default configuration. For this reason, the appropriate conclusion is not that metaheuristics improve every model, but that they can provide model-specific gains when the base learner already has sufficient structural flexibility. This nuance strengthens, rather than weakens, the choice of optimized LightGBM as the model used for interpretation.

3.2. SPATIAL HETEROGENEITY AND THE MAGNITUDE OF TEMPERATURE IMPORTANCE

The left panel of Figure 2 confirms pronounced spatial heterogeneity in benzene concentrations across Europe. Elevated average levels are concentrated in parts of Central and Southeastern Europe and at selected urban-industrial sites, while lower concentrations are more common in parts of Western and Southwestern Europe. This pattern is consistent with mixed control by local emissions, urbanization, source composition, and meteorological context rather than by a single continental-scale driver.

The right panel of Figure 2 does not show the sign of the temperature contribution, but the mean absolute magnitude of the T02m SHAP effect aggregated by site. This distinction changes the interpretation substantially. The map indicates that temperature importance is spatially heterogeneous, but it does not justify statements about positive or negative temperature influence at the site level. Several central and southeastern sites show both elevated benzene and strong temperature importance, yet the correspondence is incomplete; some benzene hotspots do not coincide with equally strong temperature importance. This mismatch is informative, because it suggests that temperature is not simply a surrogate for concentration level. Instead, it modulates benzene more strongly in some environmental settings than in others.

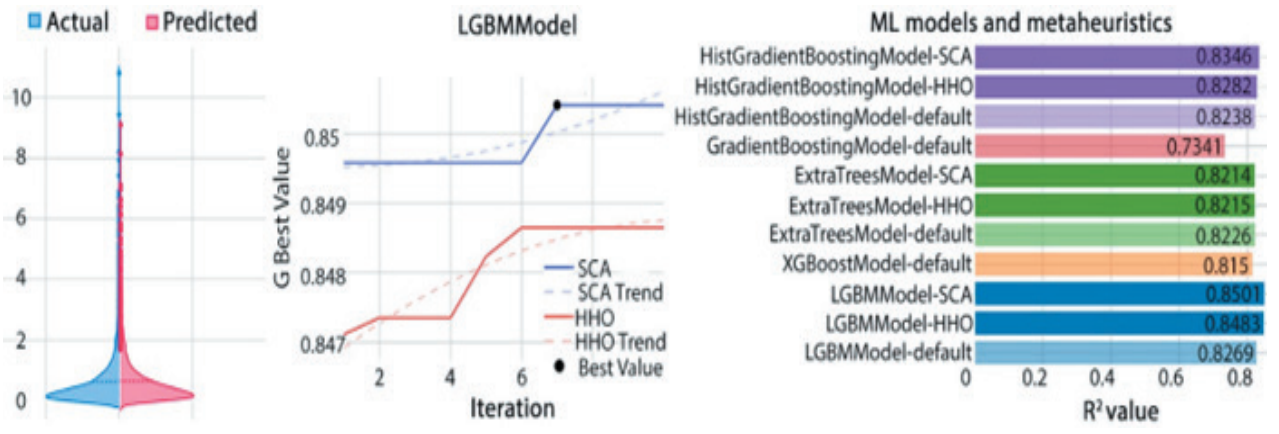

Figure 1. Comparison of model performance and prediction characteristics: distribution of observed vs. predicted values (left), convergence of the LightGBM model under SCA and HHO optimization (centre), and comparative performance of ML models with and without metaheuristic tuning in terms of R^2 (right)

3.3. INTERPRETING THE SIGNED T02M-SHAP RELATIONSHIP

Figure 3 provides the key scientific result of this paper. Unlike Figure 2, the left panel of Figure 3 contains signed SHAP values and therefore expresses whether a given T02m value raises or lowers the model prediction relative to the model baseline. The overall pattern is clear: cold conditions are associated with positive SHAP contributions, the effect declines progressively with warming, the curve crosses approximately zero near 9.5 °C, a broad negative basin follows through much of the mild-to-warm range, and the warmest conditions show a modest rebound toward zero and slightly positive values.

This structure is physically plausible. Under cold conditions, benzene can accumulate because combustion-related emissions are higher, atmospheric mixing is often weaker, and chemical removal is slower. These factors are consistent with the long gas-phase lifetime of benzene relative to highly reactive VOCs and with the dominant role of OH as a sink [6]. European and urban non-methane hydrocarbons (NMHC) observations likewise show stronger winter accumulation under shallow boundary layers and low-wind conditions, often accompanied by contributions from domestic heating in addition to traffic [18, 19]. The positive SHAP values at low T02m are therefore best interpreted as a combined signal of winter accumulation processes rather than as a simple direct thermal effect.

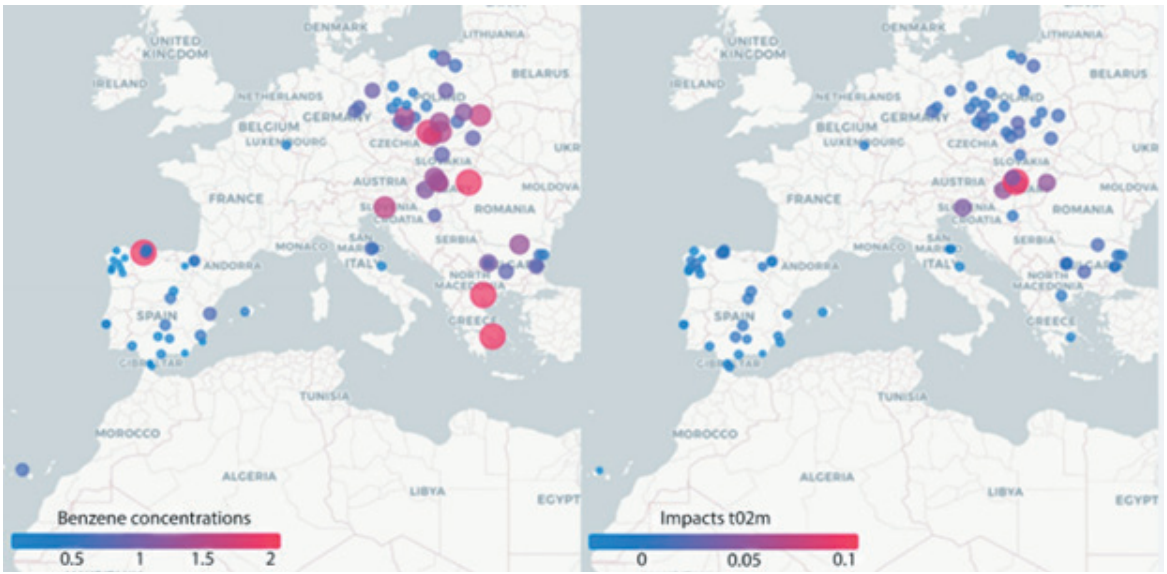

Figure 2. Spatial distribution of site-mean benzene concentrations across Europe (left) and site-aggregated mean absolute SHAP magnitude for T02m (right). The right panel describes the magnitude of temperature importance only, not the direction of the effect

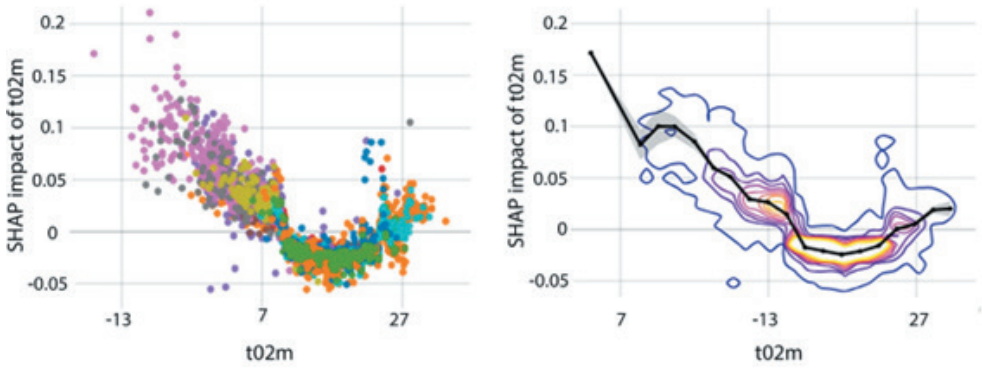

Figure 3. Observation-level signed SHAP dependence of T02m (left) and segment-wise linear approximation of the same relationship (right). Points are coloured by SHAP-based clusters representing environmental settings. The x-axis in the plots is shown in the native GDAS scale (K); Table 1 reports equivalent segment bounds in °C

The transition toward negative SHAP values with warming is equally informative. In this range, temperature should not be read as a lone causal driver but as a compact marker of seasonal process state: greater boundary-layer development, more effective dilution, a stronger oxidative environment, and a reduced relative role of heating-related emissions all push the prediction downward. The broad negative regime is especially important because it shows that, within the multivariate model, typical mild and warm conditions tend to suppress predicted benzene relative to the dataset average.

The slight warm-tail rebound requires the most careful wording. It should not be overinterpreted as evidence that high temperature robustly increases benzene everywhere. A more defensible interpretation is that the strong negative temperature effect weakens at the warmest end of the distribution. This is consistent with partial compensation between faster chemical loss and temperature-sensitive non-combustion emissions such as evaporative and fuel-system-related VOC releases [19]. In a pan-European dataset, this rebound may also reflect site-specific summer activity patterns and interactions with other predictors such as NO_2, O_3, wind speed, radiation, and mobility.

The clustering visible in Figure 3 adds a second layer of interpretation. At the same temperature, the SHAP contribution can still vary substantially, particularly in the central and warm parts of the distribution.

This spread shows that temperature acts conditionally, not universally: similar T02m values can contribute differently depending on the co-occurring pollutant mixture, local activity, site characteristics, and meteorological context. The ordered arrangement of cluster colours along the curve suggests that temperature partly organizes the environmental settings, but the overlap among clusters shows that temperature alone is insufficient to determine benzene behaviour. In this sense, the segment-wise fit provides the common backbone of the T02m effect, while the clusters represent context-specific deviations around that backbone.

3.4. SEGMENT-WISE PARAMETERIZATION AND ITS INTERPRETIVE VALUE

The fitted piecewise-linear representation converts a visually intuitive but still qualitative SHAP dependence pattern into a compact numerical object. This is the main methodological added value of the paper. Instead of merely stating that the T02m effect is nonlinear, the analysis identifies where the slope changes, where the contribution crosses zero, and where the relationship flattens or rebounds. For heterogeneous, continent-scale data, such parameterization is useful because it creates a transferable summary that can later be compared across pollutants, site groups, or interaction contexts.

Table 1. Segment-wise linear model parameters for the T02m-SHAP relationship

Segment	T02m min (°C)	T02m max (°C)	Slope	Intercept
1	-17.1564	-10.4666	-0.0133	-0.0481
2	-10.4666	-8.0444	0.0074	0.1653
3	-8.0444	-5.8275	-0.0001	0.1009
4	-5.8275	-3.1604	-0.0055	0.0690
5	-3.1604	-0.5753	-0.0099	0.0486
6	-0.5753	1.7127	-0.0037	0.0635
7	1.7127	4.2653	-0.0085	0.0572
8	4.2653	6.7953	-0.0013	0.0339
9	6.7953	9.2770	-0.0046	0.0541
10	9.2770	11.7738	-0.0130	0.1233
11	11.7738	14.2459	-0.0015	-0.0048
12	14.2459	16.7444	-0.0013	0.0112
13	16.7444	19.2484	0.0012	-0.0505
14	19.2484	21.7479	0.0022	-0.0647
15	21.7479	24.2262	0.0066	-0.1641
16	24.2262	26.6468	0.0019	-0.0459
17	26.6468	29.1159	0.0056	-0.1411
18	29.1159	31.4320	0.0005	0.0030

Note: the exported SHAP plot uses the native GDAS temperature scale (K), whereas the bounds in this table are expressed in °C. The slopes are therefore identical to the original fit, but the intercepts are converted so that the parameters are internally consistent with the Celsius bounds reported here.

The 18 fitted segments can be read at two levels. At a fine level, they describe local changes in slope. At a coarser and more scientifically useful level, they collapse into four macro-regimes: (1) a cold positive-influence regime below roughly -8 °C, (2) a cold-to-mild transition regime from about -8 °C to around 10 °C, (3) a broad negative regime from about 10 °C to about 24 °C, and (4) a weak rebound regime above about 24 °C. The two most interpretable transition temperatures are therefore the first zero crossing near 9.5 °C and the return toward zero near 24.2 °C.

Not every short segment deserves equal physical interpretation. The brief positive slope in Segment 2 and the nearly flat Segment 3 should be treated cautiously, because the tails of the temperature distribution contain fewer observations and are more vulnerable to cluster mixing and local anomalies. The same caution applies to the hottest segments. In practical terms, the slope sign and the transition points are more informative than the intercept alone, and the macro-regime interpretation is more robust than assigning a separate mechanism to each of the 18 segments.

4. CONCLUSION

This study shows that the combination of ensemble ML, SHAP-based interpretation, clustering, and segment-wise approximation can yield both predictive skill and scientifically meaningful explanation for a heterogeneous European benzene dataset. The optimized LightGBM model provides strong predictive performance, but the scientific value of the workflow lies in what happens after prediction: spatial importance mapping, signed dependence analysis, and the extraction of interpretable regime parameters from SHAP structure.

For the temperature case study, the main message is twofold. First, the magnitude of temperature importance varies substantially across Europe, confirming that T02m operates within a context-sensitive environmental setting. Second, the signed T02m-SHAP dependence follows a coherent large-scale structure: positive influence under cold conditions, a decline to a first zero crossing near 9.5 °C, a negative middle regime, and a weak warm-tail rebound above about 24 °C. This common backbone, together with cluster-specific deviations, offers a compact explanation of how temperature participates in benzene variability across sites.

Further work should now move beyond temperature alone. The same regime-extraction strategy should be applied to wind speed, relative humidity, solar radiation, NO_2, O_3, and mobility-related variables. In addition, SHAP interaction values should be explored for pairs such as T02m-NO2, T02m-O3, T02m-wind speed, and T02m-mobility, because the spread visible at equal temperature strongly suggests interaction effects. Methodologically, the robustness of the extracted regimes should be tested using bootstrap confidence intervals for breakpoints and slopes, and the SHAP results should be compared with accumulated local effects (ALE), which are less vulnerable to unrealistic extrapolation under correlated predictors [20, 21]. Finally, validation should be strengthened with blocked temporal and spatial cross-validation, because row-wise splits can overestimate generalization in multi-site environmental.

5. ACKNOWLEDGEMENTS

The authors acknowledge funding provided by the Institute of Physics Belgrade through support from the Ministry of Science, Technological Development and Innovation 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.

REFERENCES

[1] J. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001. doi: 10.1214/aos/1013203451

[2] P. Geurts, Ernst, D. and Wehenkel, L., "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3-42, 2006. doi: 10.1007/s10994-006-6226-1

[3] G. Ke, Meng, Q., Finley, T., Wang, T., Chen, W. and Ma, W., "LightGBM: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, p. 30, 2017. https://dl.acm.org/doi/10.5555/3294996.3295074

[4] S. Lundberg and Lee, S.-I., "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, p. 30, 2017. doi: 10.48550/arXiv.1705.07874

[5] I. A. f. R. o. C. (IARC), "Benzene. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 120.," IARC, Lyon, 2018. ISBN-13 978-92-832-0158-8

[6] W. H. O. R. O. f. Europe, "Benzene," in *WHO Guidelines for Indoor Air Quality: Selected Pollutants*, Copenhagen, WHO Regional Office for Europe, 2010. ISBN 978 92 890 0213 4

[7] M. Adam, Tran, P.T.M. and Balasubramanian, R., "Air quality changes in cities during the COVID-19 lockdown: A critical review," *Atmospheric Research*, vol. 264, p. 105823, 2021. doi: 10.1016/j.atmosres.2021.105823

[8] T. Hale, Angrist, N., Goldszmidt, R., Kira, B. and Petherick, A., "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)," *Nature*, 2021. doi: 10.1038/s41562-021-01079-8

[9] 2. Google LLC, "Google COVID-19 Community Mobility Reports," Google LLC, 2022. [Online]. Available: https://google.com/covid19/mobility/. [Accessed 2022].

[10] O. W. i. Data, "Coronavirus (COVID-19) dataset," 2023. [Online]. Available: https://ourworldindata.org/coronavirus. [Accessed 2022].

[11] N. Radić, Perišić, M., Jovanović, G., Bezdan, T., Stanišić, S., Stanić, N. and Stojić, A., "An AI-Based Framework for Characterizing the Atmospheric Fate of Air Pollutants Within Diverse Environmental Settings," *Atmosphere*, vol. 16, no. 2, p. 231, 2025. doi: 10.3390/atmos16020231

[12] G. Jovanović, Perišić, M., Bezdan, T., Stanišić, S., Radusin, K., Popović, A. and Stojić, A., "The PM2.5-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*, vol. 15, p. 1520, 2024. doi: 10.3390/atmos15121520

[13] N. A. R. Laboratory, "Global Data Assimilation System (GDAS1) meteorological data," 2023. [Online]. Available: https://ready.noaa.gov/gdas1.php. [Accessed 2023].

[14] S. Mirjalili, "Sine cosine algorithm: A metaheuristic for solving optimization problems," *Knowledge-Based Systems*,, vol. 96, pp. 120-133, 2016. doi: 10.1016/j.knosys.2015.12.022

[15] A. Heidari, Mirjalili, S., Faris, H., Aljarah, I. and Mafarja, M., "Harris hawks optimization: Algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849-872. doi: 10.1016/j.future.2019.02.028

[16] Y. Wang, Huang, H., Rudin, C. and Shaposhnik, Y., "Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMAP, and PaCMAP for data visualization," *Journal of Machine Learning Research*, vol. 22, pp. 1-73, 2021. doi: 10.48550/arXiv.2012.04456

[17] L. McInnes, Healy, J. and Astels, S., "HDBSCAN: Hierarchical density-based clustering," *Journal of Open Source Software,* vol. 2, no. 11, p. 205, 2017. doi: 10.21105/joss.00205

[18] X. Liu, Schnelle-Kreis, J., Zhang, X., Liakakou, E. and Mihalopulos, N., "Measurement report: Exploring the variations in ambient BTEX in urban Europe and their environmental health implications," *Atmospheric Chemistry and physics*, vol. 25, pp. 625-638, 2025. doi: 10.5194/acp-25-625-2025

[19] E. E. Agency, "EMEP/EEA air pollutant emission inventory guidebook 2019: 1.A.3.b.v Gasoline evaporation," EEA, Copenhagen, 2019.

[20] K. Aas, Jullum, M. and Løland, A., "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," *Artificial Intelligence*, vol. 298, p. 103502, 2021. doi: 10.1016/j.artint.2021.103502

[21] D. Apley and Zhu, J., "Visualizing the effects of predictor variables in black box supervised learning models," *Journal of the Royal Statistical Society: Series B,* vol. 82, no. 4, pp. 1059-1086, 2020. doi: 10.48550/arXiv.1612.08468