Understanding the drivers of differences in PAH compositions of PM_1 and PM_{10} – a machine learning study in Zagreb, Croatia

Nikolina Račić[1*], Mario Lovrić[2,3,4], Michael Forsmann[2], Ivana Jakovljević[1], Zdravka Sever Štrukil[1], Gordana Pehnec[1]

[1]Institute for Medical Research and Occupational Health, Ksaverska cesta 2, Zagreb, Croatia; nracic@imi.hr
[2]Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, 2820 Gentofte, Denmark
[3]The Lisbon Council, Rue de la Loi 155, 1040 Brussels, Belgium
[4]Centre for Bioanthropology, Institute for Anthropological Research, Gajeva 32, HR-10000 Zagreb, Croatia

Keywords: air pollution, contributions, PAHs, PM_1, PM_{10}

Polycyclic aromatic hydrocarbons (PAHs) are of significant interest due to their known association with health effects as constituents of particulate matter (PM). Understanding the differences in PAH composition between PM_1 and PM_{10} is crucial, particularly because PM_1 particles, given their size, can penetrate into the respiratory system.

This study presents an analysis of the seasonal variation and factors affecting the composition of PAHs in PM_1 and PM_{10} particles in Zagreb, Croatia. Using machine learning techniques, including Principal Component Analysis (PCA) and Random Forest (RF) regression, alongside statistical methods such as t-tests and Cohen's d calculations, this study explores differences in the PAH-related composition of PM. To understand contributions in the models, techniques such as SHAP for post-hoc model interpretation were utilized. The results exhibited seasonal trends, with PM_1 consistently showing a higher PAH contribution compared to PM_{10}, suggesting a greater potential health risk. Seasonal boxplots revealed that PAH concentrations were notably higher in winter and autumn across both particulate sizes, reflecting the impact of temperature inversion and increased emissions from heating. In the analysis conducted using Random Forest regression models and permutation importance analysis, the temperature maximum emerged as the most significant variable impacting PAH concentrations across pollutants. This was likely due to its impact on the volatilization and atmospheric chemistry of these compounds, underscoring the importance of meteorological conditions in air pollution research. Our findings offer valuable insights into the seasonally affected distributions of PAHs in urban environments, thereby contributing to the understanding of pollution dynamics and public health implications.