

SPATIO-TEMPORAL ANALYSIS AND SOURCE APPORTIONMENT OF POLYCYCLIC AROMATIC HYDROCARBONS AND METALS IN PARTICULATE MATTER AND SOIL USING MACHINE LEARNING

Abstract:

Understanding the distribution and sources of polycyclic aromatic hydrocarbons (PAHs) and potentially toxic elements (PTEs) is essential for assessing long-term risks to human health and ecosystems. This dissertation investigates spatio-temporal variability and source-related variations of particulate matter-bound PAHs and PTEs and evaluates their relationship with levels in nearby urban soils, using multi-year monitoring data, receptor-oriented multivariate methods, and explainable machine learning (ML). The atmospheric analysis integrates particulate matter (PM₁₀) composition from four monitoring stations in Zagreb (Croatia) over a four-year period, complemented by soil sampling at three locations during two seasons at two depth intervals.

To capture anthropogenic influences, ERA5 meteorology, gas consumption (as a residential-heating proxy), and traffic-volume data were incorporated. The data were analysed using correlation analysis, principal component analysis (PCA), non-negative matrix factorization (NMF), and Random Forest (RF) models with SHapley Additive exPlanations (SHAP) interpretation. NMF separated a heating-related PAH component enriched in 4-ring compounds from a traffic-associated profile characterised by 5–6-ring PAHs. Across stations and seasons, RF models achieved substantially higher skill for PAHs ($R^2 \approx 0.60\text{--}0.68$) than for PTEs ($R^2 \approx 0.24\text{--}0.42$), with temperature and solar radiation emerging as dominant variables explaining PAH variability, while PM₁₀ and NO₂ were among the strongest predictors for PTEs variability, consistent with mixed contributions from traffic-related non-exhaust processes and local site effects.

To complement the atmospheric component, soil samples collected near air quality stations were used to assess deposition and accumulation trends. Air–soil relationships were strongest for Pb and Cu and became more pronounced when airborne pollutant concentrations were averaged over longer periods (approximately 3–6 months), consistent with cumulative deposition processes and different environmental persistence across pollutants.

By integrating spatio-temporal modelling with explainable ML, this work provides a coherent, data-driven framework for interpreting urban pollution dynamics across different environmental matrices and time scales and supports targeted mitigation strategies that address both short-term air quality and longer-term soil pollution. To our knowledge, this study represents one of the first systematic research that connect parallel urban air and soil measurements of PAHs and PTEs in Zagreb, and among the relatively few integrated studies in Europe.

PROSTORNO-VREMENSKA ANALIZA I ODREĐIVANJE IZVORA POLICIKLIČKIH AROMATSKIH UGLJIKOVODIKA I METALA U LEBDEĆIM ČESTICAMA I TLU PRIMJENOM METODA STROJNOG UČENJA

Sažetak:

Razumijevanje raspodjele i izvora policikličkih aromatskih ugljikovodika (PAU-a) i potencijalno toksičnih elemenata (PTE-a) ključno je za procjenu dugoročnih rizika za ljudsko zdravlje i ekosustave. Ova disertacija istražuje prostorno-vremenski odnos i varijacije povezane s izvorima PAU-a i PTE-a vezanih uz lebdeće čestice te procjenjuje njihov odnos s razinama u obližnjim urbanim tlima, koristeći višegodišnje podatke monitoringa, multivarijatne statističke metode i metode strojnog učenja. Analiza integrira koncentracije lebdećih čestica (PM₁₀) s četiri mjerne postaje u Zagrebu (Hrvatska) tijekom četverogodišnjeg razdoblja i uzorkovanje tla na tri lokacije tijekom dvije sezone, na dvije dubine.

Kako bi se obuhvatio utjecaj antropogenih čimbenika, analizirani su meteorološki podaci ERA5, potrošnja plina (kao pokazatelj grijanja) i podaci o intenzitetu prometa. Podaci su analizirani primjenom korelacijske analize, analize glavnih komponenti (PCA), ne-negativne faktorizacije matrice (NMF) i modela Random Forest (RF) uz interpretaciju SHAP (engl. *SHapley Additive exPlanations*). NMF je izdvojio komponentu PAU-a povezanu s grijanjem, obogaćenu spojevima s četiri prstena, od onih povezanih s prometom, karakteriziranog PAU-ima s 5 do 6 prstenova. Na razini postaja i sezona, RF modeli pokazali su se puno bolji za PAU-e ($R^2 \approx 0,60\text{--}0,68$) nego za PTE-e ($R^2 \approx 0,24\text{--}0,42$), pri čemu su temperatura i sunčevo zračenje izdvojeni kao dominantne varijable koje objašnjavaju varijabilnost PAU-a, dok su PM₁₀ i NO₂ među najjačim prediktorima varijabilnosti PTE-a, što je u skladu s doprinosima iz prometa i lokalnih specifičnosti lokacije.

Kao dopuna atmosferskoj komponenti, uzorci tla prikupljeni u blizini postaja za mjerenje kvalitete zraka korišteni su za procjenu trendova taloženja i akumulacije. Odnosi zrak–tlo bili su najizraženiji za Pb i Cu te su postajali uočljiviji kada su koncentracije onečišćujućih tvari u zraku usrednjene kroz dulja razdoblja (približno 3 do 6 mjeseci), što je u skladu s kumulativnim procesima taloženja i različitom postojanošću onečišćujućih tvari u okolišu.

Integracijom prostorno-vremenskog modeliranja s metodama strojnog učenja, ovaj rad pruža sveobuhvatan, podatkovno utemeljen okvir za interpretaciju dinamike urbanog onečišćenja kroz različite okolišne matrice i vremenske skale, te podupire ciljne strategije smanjenja onečišćenja koje obuhvaćaju kvalitetu zraka i dugoročnije onečišćenje tla. Prema našim saznanjima, ovo istraživanje predstavlja jedno od prvih sustavnih istraživanja koje povezuje paralelna urbana mjerenja PAU-a i PTE-a u zraku i tlu u Zagrebu, te je među malobrojnim studijama u Europi.