University of Zagreb

Faculty of Mining, Geology and Petroleum Engineering

Nikolina Račić

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

DOCTORAL THESIS

Mentors:
Prof. Stanko Ružičić, PhD
Asst. prof. Mario Lovrić, PhD

Zagreb, 2026

University of Zagreb

Faculty of Mining, Geology and Petroleum Engineering

Nikolina Račić

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

DOCTORAL THESIS

Mentors:
Prof. Stanko Ružičić, PhD
Asst. prof. Mario Lovrić, PhD

Zagreb, 2026

Sveučilište u Zagrebu

Rudarsko-geološko-naftni fakultet

Nikolina Račić

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

DOKTORSKI RAD

Mentori:
prof. dr. sc. Stanko Ružičić
nasl. doc. dr. sc. Mario Lovrić, znanstveni suradnik

Zagreb, 2026.

This doctoral dissertation was made at Division of Environmental Hygiene, Institute for Medical Research and Occupational Health, under the supervision of Prof. Stanko Ružičić, PhD and Asst. Prof. Mario Lovrić, PhD, as part of the Postgraduate program of Geology at the Department of Mineralogy, Petrology and Mineral Resources, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb. The doctoral thesis was prepared as part of the scientific research projects "EnvironPollutHealth" carried out by Institute for Medical Research and Occupational Health and Horizon NextAIRE carried out by Institute for Anthropological Research.

Acknowledgements

First, I would like to express my gratitude to the Institute for Medical Research and Occupational Health, in particular to its Director, Ana Lucić Vrdoljak, PhD, and to the Head of the Division of Environmental Hygiene, Gordana Pehnec, PhD, for funding this study and for recognizing and supporting my ambition. I am equally grateful to all my colleagues at the Division for their support and assistance throughout the preparation of this thesis.

I would also like to thank the staff of the Faculty of Mining, Geology and Petroleum Engineering, who have accompanied me since my undergraduate days. Special thanks go to Tomislav Brenko, PhD, and Michaela Hruškova Hasan, PhD, for their help with the soil analyses.

My sincere thanks go to the members of the dissertation evaluation committee, Prof. Marta Mileusnić, PhD, Prof. Goran Durn, PhD, and Sanja Frka Milosavljević, PhD, for their valuable advice on how to improve this work.

I would also like to thank The Lisbon Council for the opportunity to work in an environment where I have had the privilege of learning from highly skilled professionals, and for giving me the chance to grow professionally while doing what I love – writing and implementing European scientific projects. Special thanks go to Francesco Mureddu and Jon Switters.

I am grateful to L'Oréal-UNESCO for the national For Women in Science fellowship awarded to me in 2026. This great recognition meant a lot to me in the continuation of my scientific career and was a tremendous encouragement.

My deepest gratitude goes to my two excellent mentors, Prof. Stanko Ružičić, PhD, from the Faculty of Mining, Geology and Petroleum Engineering, and Asst. Prof. Mario Lovrić, PhD, from the Institute for Anthropological Research. Their collaborative engagement and readiness to resolve any problem without delay were of great assistance throughout the course of this work. I look forward to our collaboration in the years ahead.

Mario, thank you for being the only one who recognized that a job and a doctorate alone were still not enough for me – that I would like to learn something more along the way. Thank you for the knowledge you have shared with me, for including me in your projects, for all the advice you have given me along the way, and for the excellent opportunities for professional development you have provided. I cannot thank you enough for all of it.

Special thanks to the best data processing team: Valentino, Dejan, and Mario.

Finally, I would like to thank my family and friends for their support and understanding.

“Do not go where the path may lead, go instead where there is no path and leave a trail.”

Ralph Waldo Emerson

Information about the mentors:

Prof. Stanko Ružičić, PhD

University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering

Department of mineralogy, petrology and mineral resources

Stanko Ružičić (born 1980) graduated in Geology from the Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb in 2007 and earned his PhD there in 2013. He has worked at the same Faculty since 2007: Scientific Novice (2007–2013), Assistant Professor (2014–2019), Associate Professor (2019–2024), and Professor (2024–present). He teaches one undergraduate course, five graduate courses, and one doctoral course, and has completed training in modern teaching methods and e-learning. From 2019 to 2023, he served on the Committee for the Amendment of Study Programs. He has also held multiple leadership and committee roles, including President of the Committee for Final and Graduation Exams (Geology and Geological Engineering Studies) and President of the Committee for Postgraduate Studies, and he has been a member of the Faculty Committee for Undergraduate and Graduate Studies since 2021/22. He participated in the InterRGN project working group focused on internationalising Faculty study programmes. His research experience includes participation in one Ministry-funded scientific project, one EU FP7 project, and four Croatian Science Foundation projects. He has authored 29 full-length journal articles indexed in Web of Science/Scopus and two book chapters. He is a member of the Croatian Geological Society and the Croatian Society of Soil Science.

Asst. prof. Mario Lovrić, PhD

Institute for Anthropological Research

Laboratory for Chemical and Biomedical Informatics

Mario Lovrić (born 1986) is a Scientific Associate and Head of the Laboratory for Chemical and Biomedical Informatics at the Institute for Anthropological Research (Zagreb, Croatia). He earned his degree in Applied Chemistry in 2012 from the Faculty of Chemical Engineering and Technology in Zagreb and completed his PhD at the Faculty of Science, University of Zagreb in 2021. Since 2022, he has served as an Adjunct Assistant Professor at the University of Rijeka, teaching Systems Biomedicine and AI in Life Sciences. Dr. Lovrić has coordinated and participated in multiple international research projects, including Horizon Europe initiatives such as NextAIRE and EDIAQI. His expertise includes applied machine learning, chemometrics, and bioinformatics, and he has authored over 70 scientific publications. His career includes research and leadership roles at the Research Center for Pharmaceutical Engineering (Austria), Know-Center (Austria), and Pliva (Croatia). He has also completed research stays at Harvard Medical School, the University of Copenhagen, and the University of Basel. He mentors several PhD and graduate students, has supervised Erasmus+ interns, and contributes to the organisation of international scientific conferences and scientific committees, with ongoing work focused on AI applications in environmental and biomedical research.

ABBREVIATIONS

AQ	Air Quality
ML	Machine Learning
PM	Particulate Matter
PAHs	Polycyclic Aromatic Hydrocarbons
NMF	Non-negative Matrix Factorization
PCA	Principal Component Analysis
RF	Random Forest
SHAP	SHapley Additive exPlanations
PTE	Potentially toxic elements
WHO	World Health Organization
EU	European Union
R^2	Coefficient of determination
RMSE	Root Mean Square Error
FA	Factor Analysis
TSP	Total Suspended Particles
LOD	Limit of Detection
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
XRF	X-ray Fluorescence
UV	Ultraviolet radiation
PM_{10}	Particulate matter with aerodynamic diameter $\leq$10 µm

LIST OF PUBLICATIONS

- **Račić, N.**; Ružičić, S.; Terzić, T.; Pehnec, G.; Jakovljević, I.; Sever Štrukil, Z.; Žužul, S.; Rinkovec, J.; Lovrić, M. (2024) Analyzing the relationship between gas consumption and airborne pollutants: case study of Zagreb, Croatia. Air Quality Atmosphere and Health, 18, 507-519. https://doi.org/10.1007/s11869-024-01655-7

- **Račić, N.**, Ružičić, S., Petrić, V., Terzić, T., Antunović, M., Škaro, I., Pehnec, G., Bešlić, I., Jakovljević, I., Sever Štrukil, Z., Rinkovec, J., Žužul, S., Lovrić, M. (2026) Assessment of contributors to airborne PAHs and heavy metals in PM_{10} using temporal, spatial, traffic and heating data in explainable machine learning models. Atmospheric Environment: X, 29, https://doi.org/10.1016/j.aeaoa.2026.100413

- **Račić, N.**; Ružičić, S.; Pehnec, G.; Jakovljević, I.; Sever Štrukil, Z.; Rinkovec, J.; Žužul, S.; Smoljo, I.; Zgorelec, Ž.; Lovrić, M. (2025) Linking Atmospheric and Soil Contamination: A Comparative Study of PAHs and Metals in PM10 and Surface Soil near Urban Monitoring Stations. Toxics 2025, 13, 866. https://doi.org/10.3390/toxics13100866

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 (PM_{10}) 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 ($R^2 \approx 0.60$–0.68) than for PTEs ($R^2 \approx 0.24$–0.42), with temperature and solar radiation emerging as dominant variables explaining PAH variability, while PM_{10} and NO_2 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, anthropogenic enrichment above geogenic baselines 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 studies to connect parallel urban air and soil measurements of PAHs and PTEs in Zagreb, and among the relatively few integrated studies in Europe.

PROŠIRENI SAŽETAK

Policiklički aromatski ugljikovodici (PAU) i potencijalno toksični elementi (PTE) predstavljaju skupinu postojanih i toksičnih onečišćujućih tvari koje su prisutne u različitim okolišnim sastavnicama, uključujući zrak i tlo. Njihovo porijeklo pretežno je povezano s ljudskim djelovanjem, osobito s procesima izgaranja fosilnih goriva, prometom, industrijskim emisijama te grijanjem. Zbog svoje sposobnosti dugotrajnog zadržavanja u okolišu i dokazano štetnih učinaka na ljudsko zdravlje, ovi spojevi su predmet intenzivnih znanstvenih istraživanja. U urbanim sredinama srednje i jugoistočne Europe, razine PAU-a i PTE-a u lebdećim česticama (PM, *engl. particulate matter*) u zraku tijekom zimskih mjeseci pokazuju izrazit porast koncentracija, što se povezuje s intenzivnijim grijanjem i meteorološkim uvjetima koji otežavaju raspršenje onečišćenja. Uz praćenje koncentracija ovih spojeva u zraku, nužno je razumjeti i njihovu prostornu raspodjelu, sezonske varijacije, kao i međusobnu povezanost s emisijskim pokazateljima poput potrošnje plina i prometnih obrazaca. S obzirom na kompleksnost okolišnih interakcija, fokus ovog istraživanja je na primjeni metoda strojnog učenja i statističkih analiza radi identifikacije izvora i boljeg razumijevanja odnosa onečišćujućih tvari u zraku i tlu.

Ciljevi i hipoteze

Ciljevi doktorskog rada uključuju:

- Kvantificirati sezonske i prostorno-vremenske varijacije koncentracija 11 PAU-a i 8 PTE-a u frakciji lebdećih čestica PM_{10} na mjernim postajama za praćenje kvalitete zraka u Zagrebu tijekom razdoblja 2017.–2020., te identificirati dugoročne trendove i prostorne razlike;
- Procijeniti utjecaj meteoroloških uvjeta, prometa i potrošnje plina na koncentracije PAU-a i PTE-a u PM_{10} primjenom naprednih statističkih i metoda strojnog učenja, kako bi se odredio relativni doprinos pojedinih čimbenika prostorno-vremenskim varijacijama;
- Ispitati koncentracije istih 11 PAU-a i 8 PTE-a u tlu na odabranim lokacijama u Zagrebu tijekom dvije sezone, analizirati sezonske i dubinske obrasce akumulacije te ispitati povezanost s koncentracijama u zraku;
- Identificirati izvore PAU-a i PTE-a u zraku i tlu primjenom statističkih metoda i metoda strojnog učenja, kako bi se razlučili antropogeni i prirodni izvori onečišćenja;

Hipoteze su temeljene na pretpostavci da:

(i) sezonski obrasci grijanja u Zagrebu određuju varijacije razina PAU-a i PTE-a u zraku, s povećanim koncentracijama tijekom razdoblja intenzivnog grijanja;

(ii) promet u Zagrebu predstavlja ključan izvor PAU-a i PTE-a u zraku i tlu, pri čemu lokacije s većim prometnim intenzitetom pokazuju povišene koncentracije;

(iii) tlo u Zagrebu akumulira PAU-e i PTE-e iz zraka, odražavajući sezonske i prostorne obrasce onečišćenja te djeluje kao rezervoar atmosferskih tvari.

Znanstveni doprinos

Ovo istraživanje predstavlja jedno od prvih sustavnih istraživanja u Hrvatskoj, a jedno od rijetkih u Europi koje integrirano pristupa prostorno-vremenskoj analizi PAU-a i PTE-a u zraku i tlu primjenom metoda strojnog učenja, poput faktorizacije nenegativne matrice (NMF), analize nasumičnih šuma (*engl. Random Forest*) i SHAP (*engl. SHapley Additive exPlanations*) analize. Korištenjem višegodišnjih podataka s gradskih mjernih postaja za praćenje kvalitete zraka, analize su omogućile preciznije razumijevanje izvora onečišćenja i njihovu povezanost s lokalnim emisijskim pokazateljima. Dodatno, analiza tla omogućuje razumijevanje akumulacijskih procesa i interakcije onečišćujućih tvari iz zraka u tlo. S geokemijskog aspekta, analiza tla omogućava razlučivanje geogenog (prirodnog) doprinosa elemenata, vezanog uz sastav matičnog supstrata, od antropogenog obogaćenja uzrokovanog atmosferskom depozicijom, čime se doprinosi i području okolišne i urbane geokemije. Radi se o jedinstvenim integriranim setovima podataka o onečišćujućim tvarima, prometu, potrošnji plina i meteorologiji koji do sada nisu analizirani zajedno. Rezultati ovog istraživanja imaju praktičnu primjenu u unaprjeđenju urbanog planiranja, prostorne politike i javnozdravstvenih mjera kontrole onečišćenja zraka.

Metode i postupci

U ovom istraživanju korišteni su podaci o koncentracijama 11 PAU-a i osam PTE-a (As, Cd, Pb, Mn, Fe, Cu, Zn, Ni) u frakciji lebdećih čestica PM_{10} tijekom četiri godine (2017.–2020.) na mjernim postajama IMI, SIG, ZG-1 i ZG-3. Nikal (Ni) je mjeren na svim postajama, međutim koncentracije su pretežno bile ispod granice detekcije analitičke metode te je stoga isključen iz daljnjih statističkih analiza i modela. Usporedno su prikupljeni meteorološki podaci (temperatura, tlak, padaline, vjetar), prometni podaci (broj vozila s glavnih ulaza u Zagreb s autocesta) te podaci o dnevnoj potrošnji plina kao pokazatelj grijanja. Za analizu podataka korištene su metode strojnog učenja (Random Forest, SHAP), faktorska analiza (NMF) i linearna regresija. Uzorkovanja tla provedena su u blizini mjernih postaja za praćenje kvalitete zraka; PAU-i u uzorcima tla analizirani su tehnikom plinske kromatografije vezane na spektrometar masa s trostrukim kvadrupolom (GC–MS/MS) nakon ubrzane ekstrakcije otapalom uz povišeni tlak i temperaturu (ASE), a PTE-i su određeni rendgenskom fluorescentnom spektrometrijom (XRF), sukladno validiranim protokolima. U zraku (PM_{10}), PAU-i su nakon ultrazvučne ili ubrzane ekstrakcije otapalom analizirani tekućinskom kromatografijom visoke djelotvornosti uz fluorescentni detektor (HPLC-FLD). PTE-i u PM_{10} određeni su tehnikom spektrometrije masa uz induktivno spregnutu plazmu (ICP-MS) nakon mikrovalne razgradnje u dušičnoj kiselini pod povišenim tlakom i temperaturom.

Rezultati i zaključci

Istraživanje provedeno u urbanom području grada Zagreba jasno je pokazalo izražene sezonske i prostorne varijacije koncentracija PAU-a i PTE-a u frakciji lebdećih čestica PM_{10}, s višim razinama tijekom zimskih mjeseci. Povišene koncentracije u zimskom razdoblju bile su povezane s povećanom potrošnjom energije za grijanje, što je potvrđeno analizom službenih podataka o potrošnji plina kao vremenskog pokazatelja sezonskih emisija. Rezultati ukazuju da je grijanje jedan od dominantnih izvora onečišćenja zraka zimi, posebno za određene PAU-e, dok promet ima izraženu ulogu tijekom cijele godine, osobito za PTE-e i PAU-e s većim brojem aromatskih prstenova. Od analiziranih PAU-a, benzo(a)antracen, krizen i benzo(k)fluoranten identificirani su kao pokazatelji emisija iz prometa, dok su fluoranten i piren uglavnom bili vezani uz izgaranje biomase i fosilnih goriva u kućnim ložištima. PAU-i viših molekulskih masa poput benzo(a)pirena, benzo(b)fluorantena, benzo(ghi)perilena i dibenzo(ah)antracena pokazali su mješovite doprinose prometnih i sezonskih izvora izgaranja. Ova podjela u skladu je i s analizom sezonskih varijacija, pri čemu je većina PAU-a imala povišene koncentracije zimi, ali su razine PAU-a nižih molekulskih masa (poput fluorantena i pirena) bile značajnije povišene u područjima sa starijim kućanstvima i individualnim ložištima.

Što se tiče PTE-a, olovo (Pb), cink (Zn), kadmij (Cd) i arsen (As) pokazali su snažnu međusobnu povezanost, a njihova prisutnost upućuje na kombinirani utjecaj prometa, grijanja i mogućih industrijskih izvora. S druge strane, željezo (Fe), mangan (Mn) i bakar (Cu) su pokazali obrasce koji odgovaraju zagađenju iz prometa koje nije vezano uz izgaranje goriva, uključujući trošenje guma, kočnica i komponenti vozila. Njihove koncentracije nisu bile toliko podložne sezonskim oscilacijama, što dodatno podupire njihovo porijeklo u prometu, a ne u sezonskim razlikama u izgaranju goriva. Pritom treba uzeti u obzir da Fe, Mn i Cu mogu dijelom imati i geogeno podrijetlo, s obzirom na njihovu prirodnu zastupljenost u mineralnom sastavu tla i resuspendirane prašine, što dodatno otežava jasno razlučivanje prometnog i prirodnog doprinosa ovih elemenata.

Prostorna analiza pokazala je razlike među mjernim postajama, pri čemu su na mjernim postajama SIG i ZG-3 zabilježene najviše koncentracije PAU-a i PTE-a, što se može pripisati njihovoj blizini frekventnim prometnicama, visokoj gustoći naseljenosti, mikroklimatskim uvjetima i mogućem utjecaju ostalih izvora poput industrije. Mjerne postaje IMI i ZG-1, smještene u stambenim i mješovitim urbanim područjima, pokazale su relativno niže razine većine onečišćujućih tvari, uz izraženiji utjecaj sezonskih izvora (grijanje). Ova razlika dodatno objašnjava važnost lokacije kao prediktora u modelima zagađenja zraka. Primjenom metoda strojnog učenja (Random Forest) i statističke analize (regresija, PCA analiza), utvrđeni su glavni prediktori promjena u koncentracijama onečišćujućih tvari. Za PAU-e, to su bili temperatura, sunčevo zračenje i masena koncentracija PM_{10}, dok su za PTE-e veći značaj imali pokazatelji vezani uz promet, NO_2 i tlak zraka. Rezultati SHAP analize dodatno upućuju na to da varijable poput temperature i sunčevog zračenja snažno utječu na razgradnju PAU-a, dok vjetar i tlak utječu na raspršenje PTE-a. Prediktivna snaga modela bila je veća za PAU-e (R^2 do 0,68) nego za PTE-e (R^2 do 0,42), što upućuje na stabilnije obrasce sezonskih emisija kod organskih spojeva, dok su koncentracije PTE-a podložnije epizodnim ili prostorno lokaliziranim izvorima.

Analizom uzoraka tla u blizini mjernih postaja za kvalitetu zraka utvrđene su značajne korelacije između koncentracija u zraku i tlu, osobito za olovo (Pb) i bakar (Cu), te za PAU-e većih molekulskih masa poput benzo(a)pirena i benzo(ghi)perilena. Za PAU-e je odnos složeniji i ovisan o svojstvima pojedinog spoja: korelacije su bile slabije ili negativne pri kratkim vremenskim periodima, što odražava utjecaj hlapljivosti, fotokemijske razgradnje i sorpcije na organsku tvar tla na koncentracije u tlu. Korelacije su bile izraženije kada su koncentracije u zraku promatrane kao tromjesečni ili šestomjesečni prosjeci, što ukazuje na kumulativnu prirodu taloženja. Ukupne koncentracije svih 11 mjerenih PAU (Σ_{11}PAU) u tlu kretale su se u rasponu od 1,2 do 524 µg/g, s višim vrijednostima u središnjim urbanim područjima. Koncentracije PTE-a bile su uobičajene za gradske zone, no lokalni porasti Cu i Zn mogu ukazivati na utjecaj prometa i resuspenziju čestica s površine. Dublji slojevi tla (5–10 cm) pokazivali su niže vrijednosti, što potvrđuje površinsku akumulaciju kroz suho i mokro taloženje. Geokemijska interpretacija ovih obrazaca pokazuje da povišene koncentracije Pb, Cu i Zn iznad geogene pozadinske razine odražavaju antropogeno obogaćenje tla. Koncentracije Fe i Mn ostale su unutar očekivanog raspona za litogenu pozadinu zagrebačkog aluvijalnog tla, no jake pozitivne korelacije s atmosferskim koncentracijama ukazuju na dodatni antropogeni doprinos vezan uz resuspenziju i prometne emisije.

Zaključno, integracijom višegodišnjih podataka o kvaliteti zraka, meteorologiji, potrošnji plina, prometu i analizama onečišćujućih tvari u tlu, te primjenom naprednih statističkih metoda i metoda strojnog učenja, ovaj rad je omogućio detaljno razumijevanje raspodjele, izvora i ponašanja PAU-a i PTE-a u urbanom okolišu. Dobiveni rezultati ističu potrebu za ciljanom regulacijom sezonskih i lokaliziranih izvora, osobito tijekom zimskih mjeseci, ali i za jačim povezivanjem sustava praćenja onečišćenja zraka i tla u kontekstu javnozdravstvene i okolišne politike.

KEYWORDS

Polycyclic Aromatic Hydrocarbons

Potentially Toxic Elements

Particulate Matter

Machine Learning

Soil Pollution

Urban Geochemistry

KLJUČNE RIJEČI

Policiklički aromatski ugljikovodici

Potencijalno toksični elementi

Lebdeće čestice

Strojno učenje

Onečišćenje tla

Urbana geokemija

TABLE OF CONTENTS

LIST OF FIGURES:

LIST OF TABLES:

1 INTRODUCTION

1.1 Air pollution: overview and public health relevance

Air pollution has become an increasingly relevant topic in environmental and health-related research, as well as in public discourse, due to its widespread presence and measurable impact on human health and ecological systems. In recent years, more scientists, policymakers and members of the public have shown interest in understanding how air pollution affects daily life, especially in cities where pollution levels often exceed recommended limits. The World Health Organization (WHO) has classified air pollution as the single largest environmental health risk globally, responsible for approximately 7 million premature deaths annually due to ambient and indoor exposure to air pollutants, particularly in urban and industrial regions (**WHO, 2021**).

Particulate matter (PM), nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), carbon monoxide (CO), volatile organic compounds (VOCs), polycyclic aromatic hydrocarbons (PAH) and various trace elements constitute the primary components of concern in ambient air. These pollutants originate from a combination of anthropogenic sources (traffic, residential heating, industry, agriculture, waste burning) and natural sources (desert dust, wildfires, sea spray, pollen, volcanic eruptions) (**Alolayan et al., 2023**), as illustrated in Figure 1. Fine and coarse particulate matter fractions, especially fractions PM_1 (particles with an aerodynamic diameter ≤ 1 µm), $PM_{2.5}$ (≤ 2.5 µm), and PM_{10} (≤ 10 µm), are of particular concern due to their capacity to penetrate the human respiratory system to varying depths (**Polichetti et al., 2009; Schraufnagel, 2020**). Smaller particles (e.g., PM_1 and $PM_{2.5}$) can reach the alveolar region of the lungs, whereas coarser particles typically deposit in the upper respiratory tract (**Wan Mahiyuddin et al., 2023**). These particles can act as carriers of toxic constituents such as PTEs, PAHs, and biological agents, posing substantial risks to human health (**Anderson et al., 2012; Bizualem et al., 2023**).

Numerous epidemiological studies have demonstrated a strong association between elevated concentrations of air pollutants and increased morbidity and mortality due to respiratory, cardiovascular, and cerebrovascular diseases (**Alahmad et al., 2023; Bhatnagar, 2022; Jedrychowski et al., 2015; Konduracka & Rostoff, 2022; Mehta et al., 2013; Raaschou-Nielsen et al., 2011**). Long-term exposure to ambient air pollution is also linked to neurodevelopmental disorders in children, cognitive decline in adults, and the exacerbation of chronic conditions such as asthma and chronic obstructive pulmonary disease (COPD) (**Albayrak et al., 2023; Bhatnagar, 2022; Sørensen et al., 2003**). Particularly concerning are the effects of persistent, low-level exposure to complex pollutant mixtures that include carcinogenic compounds such as PAHs and neurotoxic metals like lead and cadmium (Błaszczyk et al., 2017; Jakovljević et al., 2015; Potter et al., 2021).

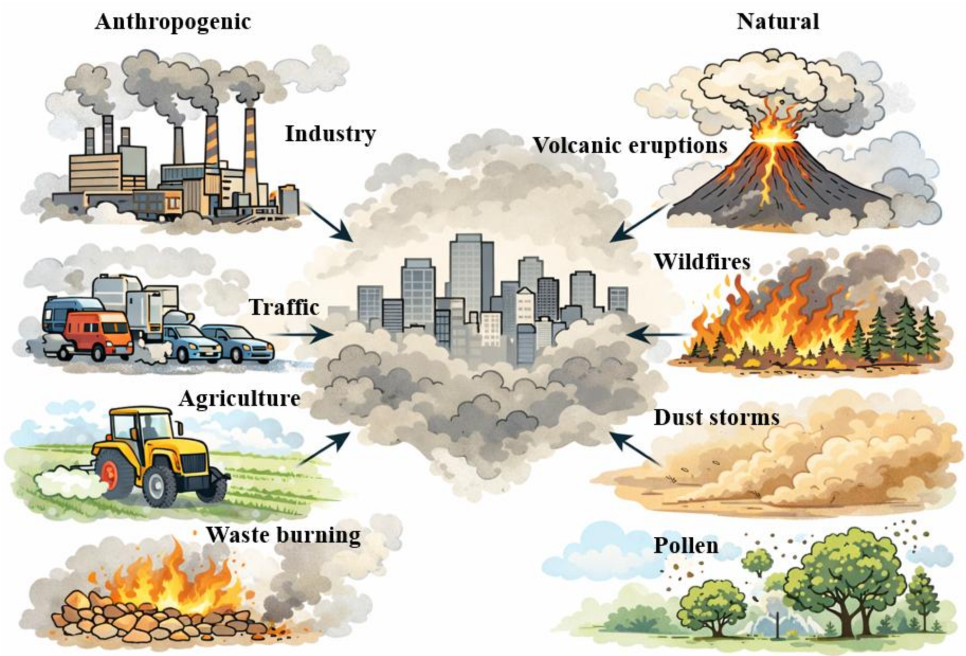


Figure 1. Main anthropogenic and natural sources of air pollution (schematic illustration; created using ChatGPT, version 5.2 and Adobe Illustrator, version 29.5.1)

The adverse effects of air pollution extend beyond human health. They include ecosystem acidification, loss of biodiversity and contributions to climate change through the emission of short-lived climate pollutants (SLCPs) such as black carbon (BC) and tropospheric ozone (**Echendu et al., 2022**). As such, air pollution represents not only a public health problem but also a cross-sectoral issue relevant to sustainable development, urban planning, energy policy, and climate governance. Despite significant regulatory efforts in the European Union (EU) and globally, many urban areas continue to exceed the recommended air quality thresholds, particularly for PM and NO_2. EU air quality legislation sets PM_{10} limit values of 50 µg/m³ (24-hour, not to be exceeded more than 35 times per year) and 40 µg/m³ (annual mean), and a NO_2 annual limit value of 40 µg/m³. However, with the aim to align more closely with the latest WHO guidelines (**WHO, 2021**), the Directive 2024/2881/EC on ambient air quality and cleaner air for Europe (recast) introduces stricter air quality standards to be met by January 1, 2030. In Croatia, data from national and regional monitoring networks indicate that PM_{10}

concentrations in several urban areas, including Zagreb, frequently exceed the annual and daily limit values prescribed by current EU and national legislation. Seasonal peaks during winter are consistently observed, reflecting increased heating demand, reduced atmospheric dispersion, and the accumulation of combustion-related emissions **(Jakovljević et al., 2018; Račić et al., 2024**).

Given the multidimensional nature of air pollution and its far-reaching consequences, research into the composition, sources, and environmental behaviour of the main particulate-phase air pollutants, such as PAHs and PTEs, is essential. Such knowledge not only supports evidence-based air quality management but also informs public health interventions and risk assessment frameworks at both local and regional levels.

1.2 Polycyclic aromatic hydrocarbons (PAHs) and potentially toxic elements (PTEs) in the urban atmosphere

Among the many harmful substances found in urban air, PAHs and PTEs are of particular concern due to their toxicity, persistence, and potential to cause long-term health effects. PAHs are a group of organic compounds made up of two or more fused aromatic rings (**Ambade et al., 2021**), as shown in Figure 2. They are mainly released into the atmosphere through incomplete combustion of organic materials, such as coal, wood, oil, petrol, and biomass (**Morillo et al., 2007**). In cities, the most common sources include traffic emissions, residential heating, industrial processes, and open burning (**Błaszczyk et al., 2017; Lee, 2010**). PAHs can be found in both the gas phase and adsorbed onto particles, especially fine particles like $PM_{2.5}$ and the fine fraction of PM_{10} (**Maharaj Kumari & Lakhani, 2018**). Their behavior in the air depends on their molecular weight: lighter PAHs tend to be in the gas phase, while heavier ones (such as benzo[a]pyrene (BaP) are more often attached to particles (**Bukowska et al., 2022; Guerreiro et al., 2016**). These heavier compounds are particularly dangerous because of their carcinogenic and mutagenic properties. Once inhaled, PAHs can trigger oxidative stress, inflammation, and DNA damage, contributing to respiratory and cardiovascular diseases, as well as cancer (**Castel et al., 2023; Cirillo et al., 2006**).

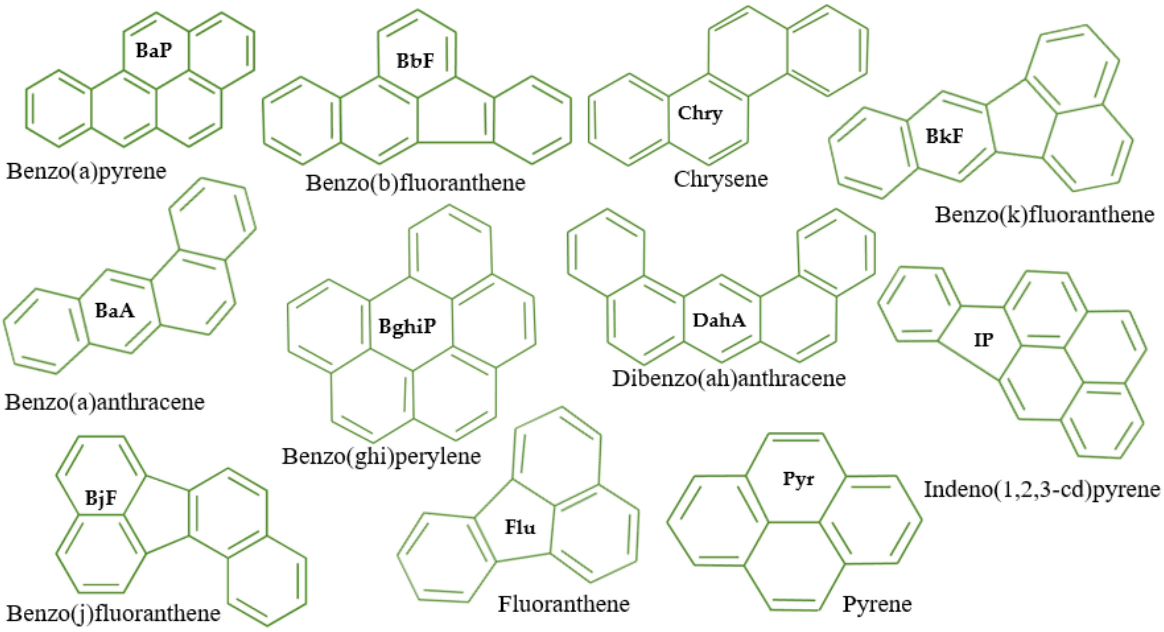


Figure 2. Polycyclic aromatic hydrocarbons (PAHs) included in this study

PTEs in PM also represent a significant health concern. They can originate from both natural sources, such as soil dust or sea spray, and anthropogenic sources, such as vehicle exhaust, brake and tire wear, industrial emissions, and waste incineration (**Chen et al., 2022; Guo et al., 2022**). From a geochemical perspective, the natural (geogenic) fraction of PTEs in atmospheric particles derives primarily from the weathering and resuspension of crustal minerals, and its contribution can be estimated using lithogenic reference elements such as Fe, Al, or Ti; concentrations exceeding geogenic baselines are then attributed to anthropogenic enrichment (**Wedepohl, 1995**). Some elements commonly found in urban air include arsenic (As), cadmium (Cd), lead (Pb), manganese (Mn), iron (Fe), copper (Cu), nickel (Ni) and zinc (Zn). In air pollution research these species are often referred to simply as "metals." However, the term potentially toxic elements (PTEs) is more accurate because it also includes metalloids (e.g., arsenic), and because toxicity depends on concentration, chemical speciation, and bioavailability (**Duffus, 2002, 2003**). Therefore, the term PTE is used throughout this thesis. Many of these elements are toxic even at low concentrations and can accumulate in the human body over time. Like PAHs, PTEs are often bound to particles and can remain suspended in the air for long periods (**Popoola et al., 2018**). Their composition and concentration can vary by season and location, depending on weather conditions and local sources. For example, in colder months, PTEs levels in PM_{10} often rise due to increased heating activities and reduced atmospheric dispersion (**Yang et al., 2015**).

Because both PAHs and PTEs are commonly found together in urban PM, they are often studied together when analyzing air quality (**P. Pandey et al., 2013**). Among other pollutants, NO_2 is often used as an indicator for traffic emissions but can originate from other combustion sources (**Anttila et al., 2011; Carslaw, 2005**). Understanding their concentrations, seasonal trends, and sources is essential for identifying pollution hotspots and planning effective air quality management strategies. All pollutants measured and analyzed in this study, including PAHs, PTEs, and other air quality indicators, are listed in **Table 1**.

Table 1. Air pollutants included in this study

PAHs*	PTEs*	Other Pollutants
Fluoranthene (Flu)	Cadmium (Cd)	Nitrogen dioxide (NO_2)
Indeno(1,2,3-cd)pyrene (IP)	Manganese (Mn)	PM_{10}
Benzo(a)anthracene (BaA)	Copper (Cu)	
Benzo(a)pyrene (BaP)	Iron (Fe)	
Benzo(b)fluoranthene (BbF)	Zinc (Zn)	
Benzo(ghi)perylene (BghiP)	Lead (Pb)	
Benzo(j)fluoranthene (BjF)	Arsenic (As)	
Benzo(k)fluoranthene (BkF)	Nickel (Ni)	
Dibenzo(ah)anthracene (DahA)		
Chrysene (Chry)		
Pyrene (Pyr)		

*PM_{10}-bound

1.3 Soil as a receptor of atmospheric pollutants

Soil has an important role in the environmental fate of airborne pollutants, acting not only as a natural filter but also as a long-term sink for various contaminants emitted into the atmosphere. Urban soils represent a complex geochemical reservoir that integrates inputs from both geogenic (natural, lithogenic) and anthropogenic sources, making their interpretation central to understanding urban pollution dynamics (**Alloway, 2013; Kabata-Pendias, 2010**). Among the most persistent of these are PAHs and PTEs, which can accumulate in the upper soil layers through both dry and wet atmospheric deposition (**D. Li et al., 2023; Ozaki et al., 2006**). Once deposited, these pollutants interact with soil particles, organic matter, and microbial communities, thereby affecting their mobility, bioavailability, and potential for re-emission (**Cachada et al., 2018**). PAHs, due to their hydrophobic nature and low volatility, tend to adsorb strongly to soil organic matter, making them relatively stable in surface soils (**Cofield et al., 2007**). However, their persistence raises concerns about long-term ecological toxicity and the potential for food chain contamination, especially in urban and industrial areas where deposition rates are elevated (**Wilcke, 2000**). Similarly, PTEs such as cadmium, lead, arsenic, and zinc can accumulate in soil matrices, where they are not subject to degradation. Their geochemical behaviour in soil is governed by a combination of factors including pH, redox conditions, organic matter content, clay mineralogy, and cation exchange capacity, all of which influence speciation, mobility, and bioavailability (**Alloway, 2013; Kabata-Pendias, 2010**). Elements such as Fe and Mn are largely present in soils as components of primary and secondary minerals of lithogenic origin, and their background concentrations reflect the geochemical composition of the parent material; elevated concentrations above these geogenic baselines therefore indicate anthropogenic enrichment (**Reimann, 1998**).

In urban environments, proximity to emission sources, such as traffic, industry, and residential heating, contributes to localized enrichment of these pollutants in soils above geogenic background levels. This anthropogenic overprint on the natural geochemical baseline is a defining feature of urban geochemistry and makes it necessary to distinguish geogenic from anthropogenic contributions when interpreting soil pollution data. Research by **Różański et al. (2017)** has shown that soils near roads and industrial zones often exhibit higher concentrations of specific PTEs and PAHs, reflecting their origin from combustion processes, tire and brake wear, and fossil fuel use. This accumulation poses direct or indirect risks to human health through dust resuspension, dermal contact, or plant uptake.

Soil analysis not only complements air quality monitoring but also helps identify persistent hotspots of contamination that may not be apparent through airborne measurements alone. In the context of environmental and urban geochemistry, the depth distribution of PTEs and PAHs in soil profiles provides additional evidence on the relative importance of recent atmospheric deposition versus longer-term geochemical redistribution processes, including vertical migration, bioturbation, and interaction with soil mineral phases. Monitoring the levels and spatial distribution of PAHs and PTEs in soil is therefore essential for evaluating the long-term impacts of atmospheric pollution.

1.4 Analytical approaches in air and soil pollution studies: from receptor models to explainable ML

Understanding the sources and drivers of PAHs and PTEs in urban environments commonly relies on a combination of (i) receptor-oriented source apportionment methods and (ii) predictive models that relate concentrations to emission proxies and meteorology (**European Commission. Joint Research Centre., 2019; Hopke, 2003; Subramaniam et al., 2022**). In PM studies, multivariate receptor models such as principal component analysis (PCA) and factor analysis (FA) have long been used to identify co-varying pollutant groups that can be interpreted as source-related patterns (**Bro & Smilde, 2014; Souza et al., 2018**). More quantitative source apportionment is often conducted using Nonnegative Matrix Factorization (NMF), which estimates source profiles and their contributions; however, its performance depends on data completeness, robust uncertainty estimates, and the degree of chemical distinctness among sources (**Lekinwala & Bhushan, 2022; Žibert et al., 2016**).

In parallel, supervised ML has gained prominence in air pollution research for modelling complex, non-linear relationships between pollutant concentrations and predictors such as meteorology, traffic intensity, heating demand, and site characteristics (**Alolayan et al., 2023; Essamlali et al., 2024)**. Tree-based ensemble models (e.g., Random Forest) are particularly attractive because they capture interactions and non-linearities without requiring strong distributional assumptions (**Kamińska, 2018; Rubal & Kumar, 2018)**. However, their use in environmental applications increasingly requires transparency: explainable ML tools, such as SHAP (SHapley Additive exPlanations) (**Vega García & Aznarte, 2020**), provide a consistent framework to quantify how predictors contribute to model outputs across seasons, locations, and concentration ranges (**Castronuovo et al., 2023; Pan et al., 2024**). Beyond RF, gradient-boosting models (e.g., XGBoost/LightGBM/CatBoost) and deep learning approaches (CNN/LSTM) have also been widely explored for air quality prediction, and more recently, graph-based neural networks (GNNs) have been proposed to learn spatial dependencies across monitoring networks. However, increasing model complexity can reduce transparency and may require larger datasets and careful validation to avoid overfitting, which strengthens the motivation for explainable ML in environmental applications (**Han et al., 2023; Liao et al., 2020**).

In both air and soil pollution research, analytical workflows commonly combine descriptive statistics (seasonality, trend analysis, exceedances), multivariate pattern recognition (PCA/FA, clustering), and source-apportionment frameworks (**Agudelo-Castañeda & Teixeira, 2014; Souza et al., 2018; Žibert et al., 2016)**. For PAHs, diagnostic ratios are also frequently used as qualitative indicators of dominant combustion sources, although their interpretation can be sensitive to atmospheric processing and therefore requires caution (**Dvorská et al., 2011; Finardi et al., 2017**). In soil studies, contamination and enrichment metrics and geostatistical mapping are often applied to characterise hotspots and potential risk relevance, while multivariate methods support source signature interpretation (**Kowalska et al., 2018**).

In soil pollution studies, multivariate methods (PCA, clustering, factorization) are widely used to interpret spatial trends, co-occurrence of contaminants, and potential sources, while geostatistical mapping and enrichment metrics are often applied to identify hotspots and evaluate ecological or human-health relevance (**J. Liu et al., 2023; Mostert et al., 2010; Soffianian et al., 2014**). Integrating soil with atmospheric observations is particularly informative because soils can act as selective, time-integrated receptors of deposition, retaining some pollutants more effectively than others depending on their physicochemical properties. Building on these methodological procedures, this doctoral thesis combines receptor-style factorization (PCA/NMF) with explainable ML (RF + SHAP) to interpret PAHs and PTEs in PM_{10} across multiple urban stations and seasons, and links atmospheric trends to pollutant accumulation in nearby soils.

1.5 Previous investigation in the study area

This research was conducted in Zagreb, Croatia, an urban area where air quality shows seasonal contrasts and winter pollution episodes are frequently associated with increased heating demand and meteorological conditions that increase pollutant accumulation. The study area covers multiple urban environments within Zagreb, including locations with high traffic intensity and sites representative of urban background/residential settings, to capture intra-urban spatial heterogeneity. The spatial distribution of monitoring station locations is shown in Figure 3.

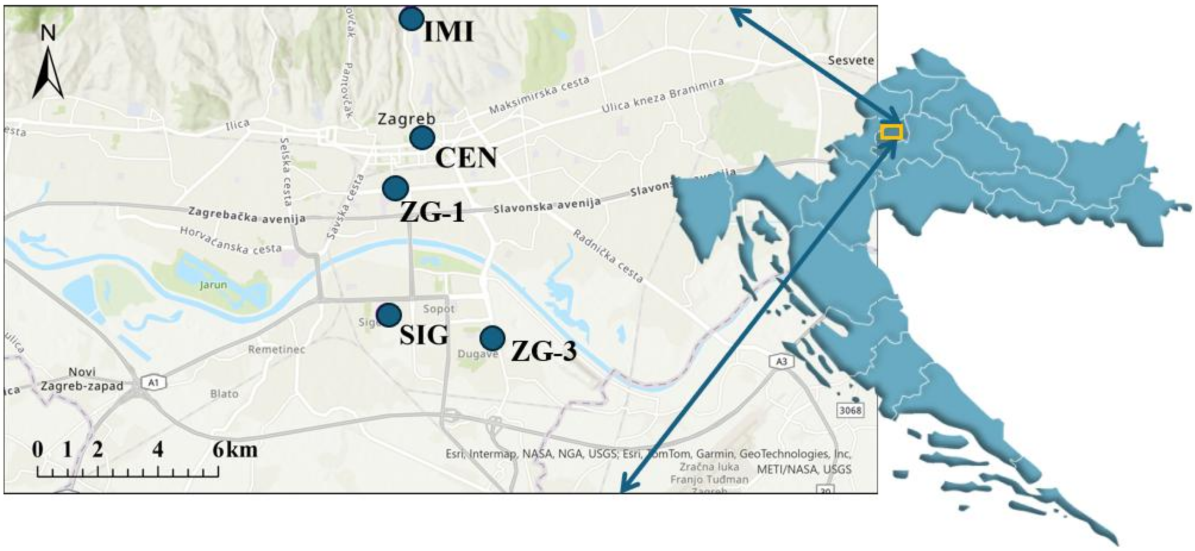


Figure 3. Monitoring station locations. The figure was created using Google Maps and Adobe Illustrator. IMI – (monitoring station Kasverska cesta, altitude 116 m a.s.l., 45°50’7” N, 15°58’42” E); SIG – (monitoring station Siget, altitude 116 m a.s.l., 45°46′25” N, 15°59’04” E), ZG-1 – (monitoring station Zagreb-1, altitude 113 m a.s.l., 45°48’18” N, 15°58’27” E), ZG-3 – (monitoring station Zagreb-3, altitude 116 m a.s.l., 45°40’46” N, 16°0’18” E), CEN – (monitoring station Đorđićeva ulica, altitude 113 m a.s.l., 45°48’37” N, 15°59’4” E).

Air pollution research and monitoring in Croatia have a long tradition, beginning with routine measurements in Zagreb in the early 1960s. Early programmes focused on classical combustion-related pollutants such as sulphur dioxide (SO_2) and smoke, reflecting the dominant energy and industrial context of that period. Over subsequent decades, monitoring expanded to include total suspended particles (TSP) and later the chemical characterisation of particle-bound species. By the 1990s and 2000s, measurements increasingly targeted other gaseous pollutants (NO_2, O_3), regulated particulate matter fractions (PM_{10} and $PM_{2.5}$) and their chemical composition, consistent with the growing public-health focus of air quality regulation and the need to understand toxicity drivers beyond bulk PM mass.

The first Croatian studies reporting PAHs in the atmosphere appeared in the early 1970s, including measurements of carcinogenic PAHs such as benzo(a)pyrene (**Cigula et al., 1972**). With the adoption of chromatographic techniques from the late 1980s onward, PAH research expanded in both scope and analytical robustness. Studies increasingly examined PAHs in a range of exposure-relevant environments, including outdoor urban air, indoor environments (homes, offices, vehicles), and personal exposure (**S. K. Pandey et al., 2011; Poster et al.,**

2006). In Zagreb specifically, earlier work documented seasonal increases of particulate-bound PAHs during winter, consistent with intensified combustion activity and atmospheric conditions favouring accumulation. Other investigations compared PAH burdens between particulate size fractions (PM_{10} vs $PM_{2.5}$) and assessed elevated PAH concentrations at traffic-influenced locations and near specific industrial or energy-related sources (**Šišović et al., 2002, 2005**). Collectively, these studies established that PAHs represent a persistent urban exposure concern and that their temporal behaviour is strongly shaped by seasonality and source mixture.

Over the past fifteen years, Croatian PAH research has developed along several parallel lines: investigation of spatial and temporal distribution, source identification using diagnostic ratios or linear multivariate statistical methods, and characterisation of PAHs across different particle size fractions. The primary objective of most of these studies was to establish population exposure levels and assess the carcinogenic potential of particle-bound PAHs and the associated health risk. Studies focused on spatial and temporal distribution consistently demonstrated typical seasonal variations, with elevated winter concentrations more pronounced in the continental part of Croatia compared to coastal areas, and high values recorded not only in urban but also in rural settings (**Jakovljević et al., 2016, 2025; Šišović et al., 2012**). The distribution of PAHs across particle size fractions and their associated carcinogenic potential were also investigated (**Jakovljević et al., 2015, 2018; Pehnec & Jakovljević, 2018**). More recent studies have placed particular emphasis on health impacts, health risk assessment, and associations with biomarkers of exposure (**Gerić et al., 2024; Jakovljević et al., 2023; Matković et al., 2025**).

Monitoring of different elements such as Pb, Cd, Zn and Mn in airborne particles in Croatia began in the early 1970s with TSP measurements (**Vadjić et al., 2010**). From 2006, in line with EU air quality directives, measurements focused on the PM_{10} fraction, targeting health-relevant metals such as Pb, Cd, As, and Ni. Long-term records show a marked decline in some elements, most notably Pb, reflecting the phase-out of leaded petrol and broader emission controls. Zinc has been measured for decades and is frequently found predominantly in the PM_{10} fraction, consistent with contributions from traffic-related and industrial processes (**Vadjić et al., 2010**). More recent work expanded the analytical scope to include emerging traffic tracers, such as platinum group elements (Pt, Pd, Rh) measured in PM_{10} in Zagreb, supporting the impact of traffic emissions (including catalytic converters) in urban areas (**Rinkovec et al., 2018**). It should be noted, however, that all the above-mentioned studies relied on linear statistical methods and typically examined PAHs in isolation or alongside only basic meteorological variables, without incorporating advanced statistical tools or ML approaches. To date, no Croatian study has simultaneously integrated PAHs and PTEs with a broad set of environmental predictors. Furthermore, data on PAHs in soil remain very scarce in Croatia; to the authors' knowledge, the only available urban-adjacent soil PAH data concern the Raša area in Istria, where historical coal mining, processing, and transport activities, as well as a coal-fired power plant, have elevated PAH levels in the surrounding environment (**Jakovljević et al., 2022**).

Although prior Croatian studies have provided valuable evidence on PAH seasonality, traffic hotspots, and long-term trends of particle-bound elements, important gaps remain for Zagreb, This doctoral thesis represents comprehensive research, which at the same time combine: (i) multi-station, multi-year PM_{10} chemical composition, (ii) explicit heating proxies (e.g., gas consumption) alongside meteorology and traffic indicators, (iii) source-resolving multivariate methods and explainable ML for driver attribution, and (iv) a cross-media perspective linking atmospheric pollution to nearby urban soils as long-term receptors. This dissertation addresses the gaps of previous studies by integrating long-term monitoring across Zagreb with factorisation methods (PCA/NMF), explainable modelling (RF–SHAP), and targeted soil sampling to interpret pollution dynamics across both time scales and environmental compartments.

1.6 Objectives and hypotheses of research

The main objectives of this research are:

(1) Quantify seasonal and spatio-temporal variability of 11 PAHs and 8 potentially toxic elements (PTEs) in the PM_{10} fraction across Zagreb air quality monitoring stations during 2017–2020, including the assessment of long-term trends and spatial differences;

(2) Evaluate the influence of meteorological conditions, traffic, and gas consumption (as a proxy for residential heating) on PM_{10}-bound PAHs and PTEs using advanced statistical methods and ML, to determine the relative contribution of key predictors to observed spatio-temporal trends;

(3) Assess concentrations of the same 11 PAHs and 8 PTEs in urban soils at selected Zagreb locations across two seasons and two depth intervals, describe seasonal and depth-related accumulation trends, and examine their relationship with atmospheric concentrations; and

(4) Identify and interpret dominant sources of PAHs and PTEs in air and soil using statistical and ML approaches, with the aim of distinguishing anthropogenic and natural contributions;

These objectives were based on the following three main hypotheses:

(i) residential heating seasonality in Zagreb is a major driver of PAHs and PTEs variability in PM_{10}, leading to higher concentrations during the heating period;

(ii) traffic-related emissions are a key source of PAHs and PTEs in both air and soil, with higher levels at locations influenced by greater traffic intensity;

(iii) urban soils in Zagreb accumulate atmospheric PAHs and PTEs over time, reflecting seasonal and spatial pollution patterns and acting as a reservoir of deposited contaminants.

1.7 Data and methodology used

This dissertation integrates three complementary datasets and a combined analytical workflow to investigate the spatio-temporal variability and sources of PM_{10}-bound PAHs and PTEs, and to assess their relationship with accumulation in nearby urban soils.

1.7.1 Data sources and study design

The atmospheric data is based on 24-hour PM_{10} sampling and chemical characterisation from Zagreb's air quality monitoring stations over the period 2017–2020. Publication I uses a single long-term station (IMI-Ksaverska cesta) measurements to examine seasonal behaviour and the influence of heating-related activity, while Publication II extends the analysis to four monitoring stations (SIG, IMI, ZG-1 and ZG-3) to quantify intra-urban spatial contrasts and identify main drivers of variability of pollutant concentrations. Air pollutant dataset included mass concentrations of 11 PAHs in PM_{10} (benzo(a)pyrene (BaP), benzo(a)anthracene (BaA), benzo(ghi)perylene (BghiP), dibenzo(ah)anthracene (DahA), chrysene (Chry), benzo(k)fluoranthene (BkF), benzo(b)fluoranthene (BbF), benzo(j)fluoranthene (BjF), fluoranthene (Flu), pyrene (Pyr) and indeno(1,2,3-cd)pyren (IP)) and 8 PTEs in PM_{10} (arsenic (As), cadmium (Cd), lead (Pb), manganese (Mn), iron (Fe), copper (Cu), nickel (Ni) and zinc (Zn)), together with PM_{10} and NO_2 mass concentrations (NO_2 is often used as an additional indicator of combustion and traffic-related conditions). Nickel concentrations were frequently below the limit of detection (LOD) and were therefore excluded from most further analyses.

To interpret atmospheric variability, chemical data were integrated with meteorological parameters (e.g., temperature, pressure, wind, precipitation, and solar radiation) and with emission proxies. Residential heating intensity was approximated using gas consumption (temporal proxy), while traffic volume was represented through available traffic indicators and/or station characteristics reflecting proximity to major roads and urban morphology. Meteorological data were obtained from the ERA5 reanalysis and/or national meteorological observations, depending on the publication and the availability of variables.

The soil component (Publication III) consists of targeted sampling of urban soils near selected air quality monitoring stations, specifically chosen within a defined radius of the IMI, CEN and SIG stations to ensure spatial correspondence with the atmospheric measurements and to allow direct evaluation of deposition-related coherence between the two media. Soil samples were collected in two seasons (spring and autumn) and at two depth intervals (0–5 cm and 5–10 cm) over three years (2018–2020), resulting in 36 composite samples (3 stations × 2 depths × 2 seasons × 3 years). At each sampling location, five individual subsamples were collected and homogenised into a single composite sample per depth and season, in order to reduce small-scale spatial variability and obtain a representative estimate of local soil pollutant concentrations.

The choice of the two depth intervals was based on the expectation that atmospherically deposited pollutants accumulate preferentially in the uppermost soil horizon, while the deeper interval (5–10 cm) provides a basis for assessing the extent of downward redistribution and longer-term, historical accumulation. This depth-resolved design therefore allows the relative contribution of recent versus cumulative atmospheric deposition to be distinguished, an aspect

that is not directly accessible from PM_{10} measurements alone. Prior to chemical analysis, soil samples were air-dried, homogenised, and finely ground using a vibratory disc mill to ensure analytical consistency and to reduce particle size variability in subsequent extraction and instrumental analysis.
Soil samples were analysed for the same suite of PAHs and PTEs measured in PM_{10}, which is essential for enabling a direct, pollutant-by-pollutant comparison between the atmospheric and soil compartments. By pairing soil sampling campaigns with time-aggregated atmospheric concentration windows, this design supports the assessment of atmospheric deposition as a contributing pathway to soil pollutant accumulation, while also allowing the identification of pollutants for which soil concentrations are decoupled from recent atmospheric levels, suggesting alternative or additional sources.

1.7.2 Analytical framework

A combined set of statistical and ML methods was applied across the three publications to address complementary questions on source-related trends and environmental drivers of the variability of pollutant concentrations. Prior to multivariate analysis, the distributional properties of pollutant concentration data were examined. As is typical for urban air quality datasets, concentrations of PAHs and PTEs exhibited right-skewed, often log-normal distributions, with occasional high-concentration episodes reflecting episodic emission events. Consequently, non-parametric approaches were preferred where appropriate. Exploratory analysis included descriptive statistics and Spearman correlation to characterise variation among pollutants and predictors and to quantify air–soil associations. PCA was used to reduce dimensionality, identify co-varying pollutant groups, and address multicollinearity among predictors. NMF was applied as an unsupervised, receptor-style approach to decompose pollutant matrices into additive components, thereby supporting the interpretation of source-related profiles and their temporal behaviour.

To model pollutant variability and quantify the relative importance of drivers in the presence of non-linear and interactive relationships, RF regression was developed for all pollutants. Model interpretation used SHAP analysis to understand the direction and magnitude of predictor effects and to identify main variables controlling pollutant concentrations. For the air–soil component, atmospheric concentrations were additionally aggregated over multi-month windows (1, 3, and 6 months prior to each soil sampling campaign) to better reflect cumulative deposition processes and to evaluate whether longer integration times strengthened the association with soil concentrations. Overall, this integrated data structure and multi-method workflow (Figure 4) enabled robust interpretation of urban pollution dynamics across time scales and different environmental media, combining source-oriented multivariate decomposition with explainable ML methods.

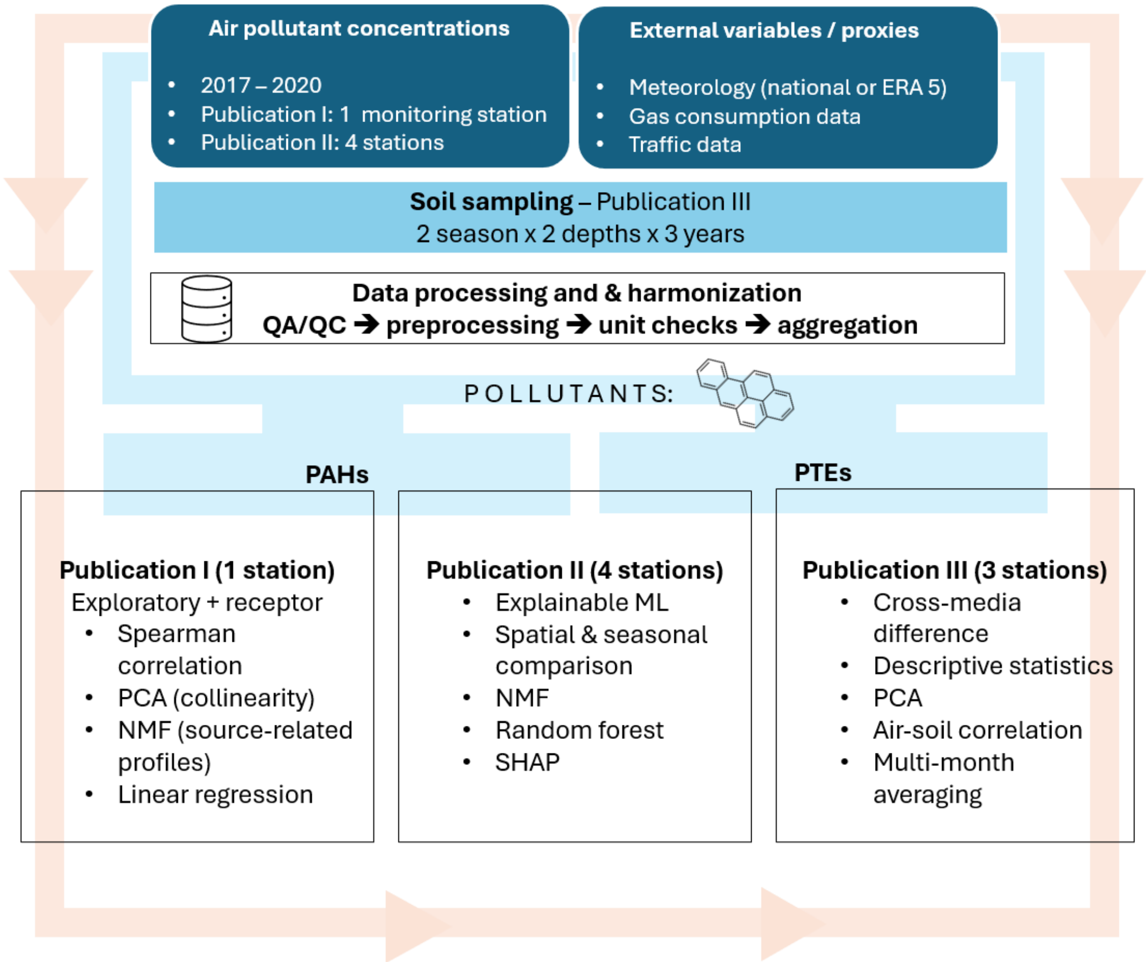


Figure 4. Integrated workflow across publications

1.8 Overview of the studies forming this dissertation

This dissertation is based on three peer-reviewed publications that together provide a coherent, stepwise investigation of PAHs and PTEs in the urban environment of Zagreb, Croatia. The research is structured to progressively increase complexity and integration: it begins with a focused assessment of seasonal trends and heating-related effects in PM_{10} at a single monitoring location, then expands to a multi-station analysis that quantifies spatial differences and identifies main spatio-temporal drivers using explainable ML. Finally, it extends the interpretation beyond the atmosphere by connecting PM_{10} composition with the accumulation of the same pollutants in nearby urban soils.

1.8.1 Publication I overview

Publication 1 provides the conceptual and methodological foundation for the dissertation by characterising PM_{10}-bound PAHs and PTEs at an urban monitoring site in Zagreb and examining their seasonal behaviour in relation to meteorology, NO_2, and residential heating activity using gas consumption as a proxy. Using correlation analysis and multivariate methods (including PCA and NMF) alongside regression modelling, the publication identifies pollutant groupings and interprets them in terms of source-related trends and seasonality.

NMF revealed two interpretable pollutant groupings, with a third component showing minimal temporal variation and limited interpretability. For PAHs, the factor structure differentiated a group characterised primarily by lower molecular weight PAHs (Pyr and Flu, with contributions from BjF) consistent with heating-related and biomass combustion emissions, from a profile enriched in higher molecular weight (5–6 ring) compounds including BaA, BkF, BaP, BghiP, BbF and DahA, which are more commonly associated with traffic-related high-temperature combustion processes. This separation supports the interpretation of at least two major influences: a combustion-driven component with a seasonal effect (consistent with heating-related emissions during the cold season) and a component dominated by heavier PAHs typically associated with traffic-related high-temperature combustion and proximity to road sources.

For PTEs, NMF separated elements into two main groups. One grouping comprised Fe–Mn–Cu, interpreted as a traffic-related non-exhaust source (consistent with vehicle component wear, brake/tyre wear, and resuspended road dust), while a second cluster including Pb–Zn–Cd–As indicated a broader mixed anthropogenic influence, reflecting combined contributions from combustion-related activities and urban traffic influences. PCA highlighted strong collinearity among temperature, gas consumption and NO_2, demonstrating that a single "heating-season axis" captures much of the shared variability in these predictors and is therefore critical for interpreting wintertime concentration elevations. Regression modelling using the extracted NMF components showed that the predictability differs across source-related trends: the strongest performance was obtained for the Fe–Mn–Cu (traffic/wear) component ($R^2 \approx 0.82$), whereas the remaining NMF components showed notably lower explanatory power ($R^2 \approx 0.36$ and $R^2 \approx 0.27$), indicating additional complexity in PAH variability beyond the heating proxy and basic meteorological controls. Overall, Publication I establishes the main interpretative framework applied throughout the dissertation: PM_{10}-bound PAHs and PTEs exhibit structured seasonal behaviour explained by emission intensity and meteorology, and multivariate factorisation can robustly separate source-related pollutant trends in an urban setting.

1.8.2 Publication II overview

This publication expands the scope from a single monitoring site to a city-scale framework based on four stations across Zagreb and forms the central spatio-temporal driver analysis of the dissertation. Using daily PM_{10} composition data, the study quantifies spatial contrasts among stations and evaluates which predictors best explain variability in PAHs and PTEs. The publication integrates multiple predictor classes, including meteorological conditions (with ERA5 data also), traffic indicators, gas consumption as a proxy for residential heating, and station-related factors that capture local context and proximity to sources.

Methodologically, the main advancement is the use of supervised ML (RF) combined with SHAP to identify and interpret the dominant predictors of pollutant variability under complex, non-linear and interacting effects. The results show spatial variability across the monitoring network, with some stations exhibiting higher PAH levels and distinct PTE trends reflecting local environments and source proximity. Source-related interpretation is supported by factorisation, with components that separate pollutant mixtures into profiles consistent with heating-related combustion (Flu, Pyr) versus traffic-influenced PAHs enriched in higher-ring species. A main quantitative finding is that model performance is systematically higher for PAHs than for PTEs: RF achieved R^2 values of approximately 0.60–0.68 for PAHs, compared with approximately 0.24–0.42 for PTEs, suggesting that PAH variability was better captured by the predictor set, while PTE variability was more influenced by localised and episodic contributions. SHAP interpretation identified temperature and solar radiation as main drivers of PAH variability across stations (consistent with seasonal emissions and atmospheric processing), whereas PM_{10}, NO_2, and station effects were major predictors for PTEs, highlighting the combined influence of particle load, co-pollutant indicators, and site-specific conditions. Collectively, Publication II provides the dissertation's strongest evidence on how the relative importance of meteorology, heating, traffic, and location differs between PAHs and PTEs and across the different monitoring stations.

1.8.3 Publication III overview

Third publication examines cross-media pollutions by connecting atmospheric observations with levels in nearby urban soils, addressing whether trends observed in PM_{10} particles are reflected in longer-term deposition and accumulation in soils. The study is built on the concept of soil as a selective, time-integrated receptor that can preserve deposition over longer periods, with retention efficiency varying by pollutant class. Surface soils collected at selected locations near air quality monitoring stations are analysed for the same PAHs and PTEs across two seasons and two depth intervals (0–5 cm and 5–10 cm), for 3 years, enabling assessment of both seasonal accumulation and vertical concentration profiles.

To compare air and soil levels, in the publication time-aggregated air concentration windows (1, 3 and 6-month rolling medians prior to each soil sampling campaign) were used, knowing that soil concentrations reflect cumulative inputs rather than day-to-day variability. The results show evidence of deposition-related coherence between air and soil for a subset of pollutants, with the strongest and most consistent air–soil associations observed for Pb and Cu. This is evident particularly when atmospheric concentrations are averaged over longer windows (approximately 3–6 months), supporting a cumulative deposition interpretation. Soil PAH concentrations covered a wide range (Σ11PAHs ≈ 1.2–524 µg/g), with higher levels at more urbanised locations. Depth analysis indicated that concentrations were often lower in the 5–10 cm layer than in the 0–5 cm layer, consistent with surface enrichment, while several elements, most notably Pb (and partly Cu), showed trends consistent with longer-term accumulation and depth-dependent behaviour, suggesting that vertical redistribution and local site processes may contribute alongside deposition. For PTEs, soil levels were within the range expected for urban environments, while local enrichments, particularly for Cu and Zn, are consistent with traffic-related non-exhaust inputs and resuspension. Fe and Mn concentrations remained within the expected range for the lithogenic background of Zagreb's alluvial substrate, although their strong positive air–soil correlations suggest a measurable anthropogenic contribution from traffic-related resuspension. Overall, Publication III shows that combining atmospheric monitoring with soil measurements strengthens interpretation of urban pollution across different environmental matrices and provides an empirical basis for identifying which PAHs and PTEs show the clearest long-term deposition signals in the urban environment.

1.9 Scientific contribution

This dissertation advances the scientific understanding of urban environmental pollution by integrating multiple long-term datasets: measured concentrations of PAHs and PTEs in PM_{10}, NO_2, traffic density data, gas consumption data as a proxy for heating activity and meteorological parameters from both local stations and ERA5 reanalysis data. The study addresses a particularly complex urban pollution setting in which residential heating during cold seasons and year-round traffic emissions represent competing and temporally overlapping sources, whose relative contributions are difficult to understand using conventional approaches alone. Using a combination of statistical and ML approaches, including NMF, PCA, and RF models, the study identifies dominant pollution sources and quantifies their seasonal and spatial influences. The soil component of this research is particularly relevant from the perspective of environmental and urban geochemistry: by examining the geochemical distribution of PTEs and PAHs across urban soil profiles at two depth intervals, and by distinguishing anthropogenic enrichment from geogenic background contributions, this dissertation provides new insights into the geochemical behaviour of these elements in a continental urban setting. Additionally, the inclusion of PAHs and PTEs in soil samples collected near air monitoring stations offers new insights into soil as a long-term receptor of atmospheric pollutants. The multi-matrix design, simultaneously examining PAHs and PTEs in atmospheric PM_{10} and in urban soils collected near monitoring stations, further extends the scope of the analysis beyond short-term air quality assessment toward an evaluation of long-term pollutant accumulation and inter-media transfer. To date, such a comprehensive and multidisciplinary approach, linking atmospheric measurements, source apportionment, and soil pollution, has not been applied in Croatia, and has been applied only in a limited number of studies across Europe, making this research a novel and valuable contribution to the field of air quality and environmental science.

Beyond the Croatian and European context, several methodological aspects of this dissertation are not widely reported in the global air pollution literature. Most existing studies apply either receptor-style decomposition methods or ML-based driver attribution separately; the systematic combination of NMF source profiling with RF–SHAP predictor attribution alongside parallel soil sampling within a single analytical framework is, to the authors' knowledge, rarely applied. The use of gas consumption as an explicit, quantitative proxy for residential heating intensity offers a more direct link between heating demand and pollutant variability than the calendar-based or temperature-based season definitions typically used in international studies. The time-aggregated air concentration windows applied in the air–soil comparison represent a methodological approach that goes beyond simple seasonal matching and enables evaluation of cumulative deposition at multiple time scales, an aspect that is not commonly addressed in the existing cross-media literature, where air and soil are typically compared using concurrent or annual mean concentrations. Finally, the measurement of the same suite of 11 PAHs and 7 PTEs in both atmospheric PM_{10} and urban soils using validated analytical protocols enables direct, pollutant-by-pollutant comparison across environmental matrices, which is rarely achieved in practice due to differences in monitoring programmes, analytical methods, or target compound lists between air and soil studies.

This methodology, which effectively bridges atmospheric chemistry, data science, and soil science, offers a novel and valuable contribution to environmental science, providing a robust, data-driven framework for targeted air quality management and the assessment of long-term geochemical, ecological and human health risks in continental urban environments.

2 DISCUSSION

2.1 Overview of main findings and overall interpretation

Before addressing the broader implications and comparison with previous studies, it is important to interpret the main findings obtained across the three parts of this doctoral research. The results provide a comprehensive understanding of the occurrence, behaviour, and sources of PAHs and PTEs in PM_{10} and urban soils in the Zagreb area, explaining complex seasonal, spatial, and cross-media pollution trends.

The atmospheric component revealed seasonal variations in PAHs concentrations with elevated levels during the heating season (Figure 5), which is common in urban areas **(Balcioğlu et al., 2021; Siudek, 2022; Y. Zhang & Tao, 2008**) and was observed in previous studies at the area (**Jakovljević et al., 2018; Račić et al., 2025**). These findings corresponded closely with increased gas consumption and adverse meteorological conditions, highlighting the importance of domestic heating and reduced dispersion in explaining wintertime air quality (**Ayoobi et al., 2022; L. Zhang et al., 2001)**. Unlike PAHs, PTEs showed a less pronounced seasonal pattern (Figure S1), with generally higher concentrations observed during the colder months. This weaker seasonality suggests that while heating emissions may contribute to PTEs levels during winter, traffic-related sources such as brake wear, tyre abrasion, and road dust resuspension represent a more continuous year-round input (**Arruti et al., 2010; Truong et al., 2022**).

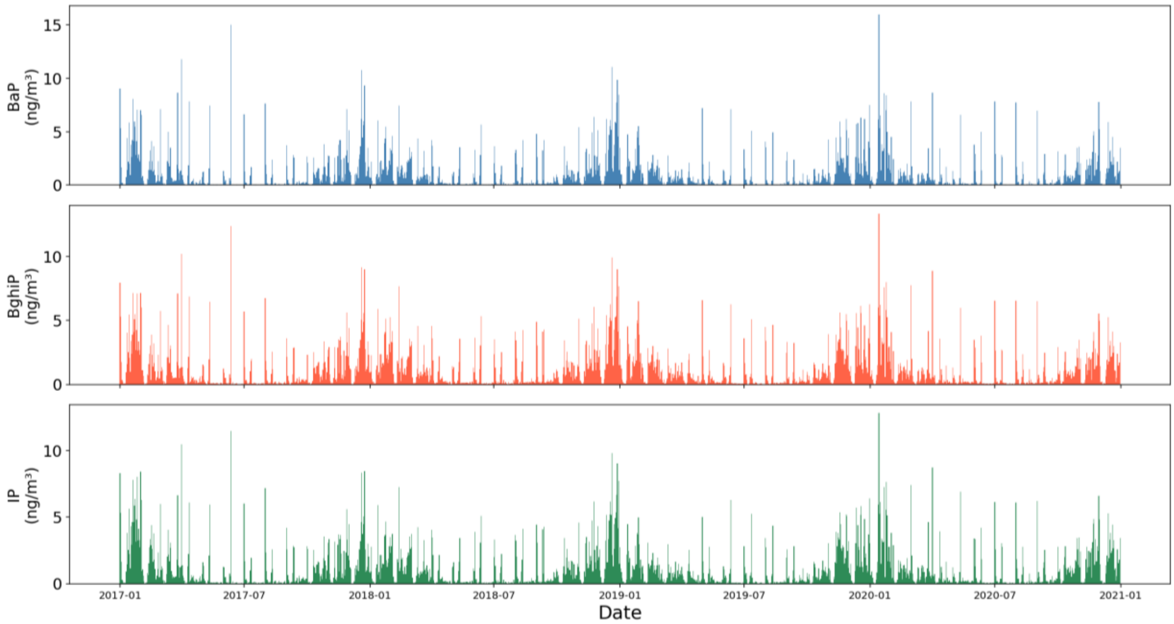


Figure 5. Daily concentrations of selected PAHs (BaP, BghiP, and IP) in PM_{10} at the IMI monitoring station during the period 2017–2020 (ng/m^3).

Overall, the atmospheric results indicate that wintertime PAHs increases are driven by the combined effects of seasonal emissions and meteorology, whereas PTEs variability shows a stronger contribution from traffic-related inputs and station-dependent influences. Consistent with this, predictive modelling captured PAHs variability more robustly than PTEs variability, indicating clearer seasonally structured drivers for PAHs and a larger role of localised processes for PTEs. The soil component showed heterogeneous pollutant distributions across Zagreb, with higher Σ_{11}PAHs and PTEs concentrations near main traffic areas and older residential zones (Figure 6).

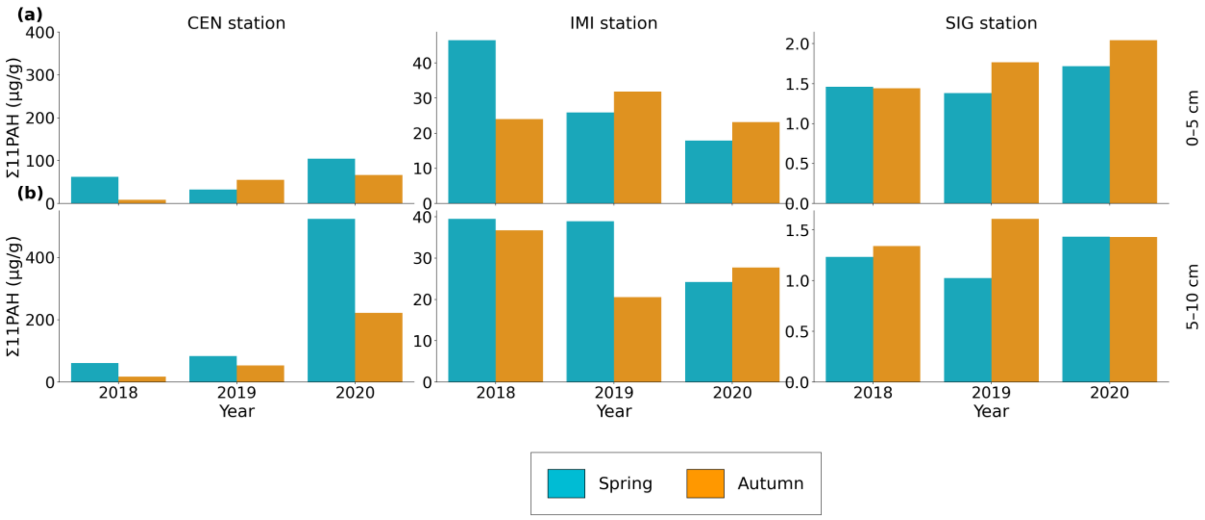


Figure 6. Annual variation of sum PAHs (Σ_{11}PAHs) concentrations in soil at depths (a) 0-5 cm; (b) 5-10 cm at monitoring sites IMI, CEN, and SIG, in µg/g.

It should be noted that the composite sampling approach, in which five subsamples were combined into a single sample per depth and season, was chosen to reduce the influence of small-scale spatial heterogeneity inherent to urban soils, following standard practice in urban soil research. However, this design does not yield within-site replicates or standard deviations, meaning that apparent seasonal difference, particularly at SIG, where absolute concentrations are low, may fall within the range of natural soil heterogeneity and analytical uncertainty. The seasonal patterns discussed here are therefore treated as indicative trends rather than statistically confirmed differences.

Vertical gradients (0–5 cm vs. 5–10 cm) generally indicated surface enrichment consistent with atmospheric deposition, while spatial trends reflected proximity to sources identified in the air dataset. Correlation analysis indicated statistically significant relationships between airborne and soil concentrations of Pb and Cu across all temporal windows, while PAH associations, including BaP and BghiP, became more apparent when atmospheric concentrations were integrated over longer periods (3–6 months), consistent with their cumulative deposition behaviour.

Taken together, these findings suggest how traffic and residential heating act as major pollution sources in Zagreb, while meteorological conditions modulate their behaviour. The integration of soil and air data further highlights the cumulative but pollutant-specific nature of urban pollution, where persistent, particle-bound species show the clearest evidence of long-term accumulation in surface soils, while more volatile compounds exhibit weaker or temporally lagged deposition signals. These results form the basis for the subsequent discussion, which examines the identified emission sources and environmental drivers in greater depth, compares the observed patterns with those reported in other European cities, and explores the broader environmental and policy implications of the findings.

2.2 Sources and drivers of PAHs and PTEs in the urban atmosphere

The identification of sources and drivers of PM_{10}-bound PAHs and PTEs in Zagreb requires distinguishing (i) emission-related sources linked to urban activities and (ii) meteorological factors that modulate dispersion, partitioning, and atmospheric processing. In this dissertation, source-related trends were analysed primarily through multivariate decomposition (PCA/NMF), while the relative influence of predictors was quantified using explainable ML (RF with SHAP interpretation). Together, these approaches provide complementary evidence on both source-related groupings and predictor attribution. The methodological combination applied here aligns with a broader trend in air-pollution studies, in which explainable ML is increasingly being combined with receptor-style decomposition to move beyond descriptive source profiles toward quantitative driver attribution (**Jain et al., 2022; Pan et al., 2024; Peng et al., 2024; Rybarczyk & Zalakeviciute, 2018**), an approach that has not previously been applied to PAHs and PTEs datasets in Croatia.

Across the atmospheric datasets, multivariate grouping revealed at least two major urban influences (Figure 7). First, PAH groupings enriched in higher molecular weight compounds (5–6 ring PAHs), including BaP, BbF, BkF, DahA and BghiP, were interpreted as traffic-influenced high-temperature combustion. This association between high-ring PAHs and traffic-related combustion has been widely reported in urban PM studies **(Gunawardena et al., 2012;**

X. Liu et al., 2019). In contrast, PAHs with stronger representation of lower molecular weight compounds, most notably Flu and Pyr, showed pronounced seasonal behaviour and were interpreted as indicators of cold-season heating-related combustion, including residential biomass and fossil fuel burning. For PTEs, the grouping of Fe–Mn–Cu was consistent with non-exhaust traffic sources and resuspension processes, whereas Pb–Zn–Cd–As reflected a broader mixed anthropogenic signal potentially influenced by both traffic and combustion-related emissions. Fe and Mn are major crustal elements whose presence in PM_{10} may partly reflect resuspension of lithogenic soil material alongside non-exhaust traffic emissions; their dual origin complicates straightforward source attribution and is consistent with the strong air–soil correlations observed for these elements. Similarly, As concentrations in PM_{10} may include a geogenic component linked to the regional geological substrate, in addition to combustion-derived inputs.

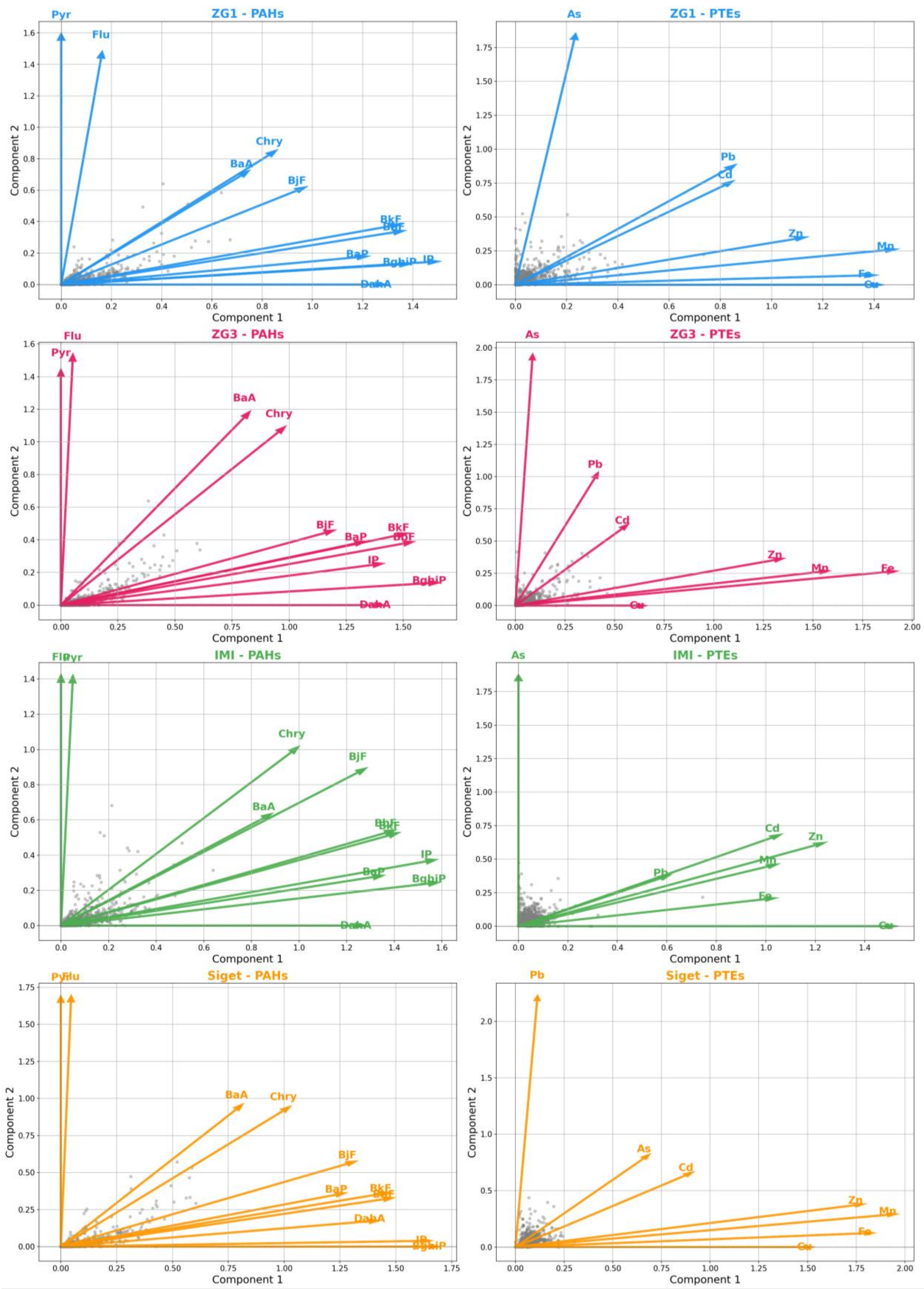


Figure 7. NMF bi-plots for PAHs and PTEs in PM_{10} at four monitoring stations

This separation is consistent with the expected emission mechanisms: PAHs are strongly affected by combustion intensity and atmospheric processing, whereas several PTEs are emitted through more mechanically driven processes and show stronger site dependence. Taken together, these PAHs and PTEs groupings align with source-apportionment outcomes widely reported in European and global studies (**Dvorská et al., 2011; Fussell et al., 2022; Grigoratos & Martini, 2015; Ribeiro et al., 2012; Vardar et al., 2008**). Figure 8 shows main PAHs and PTEs sources in urban environments.

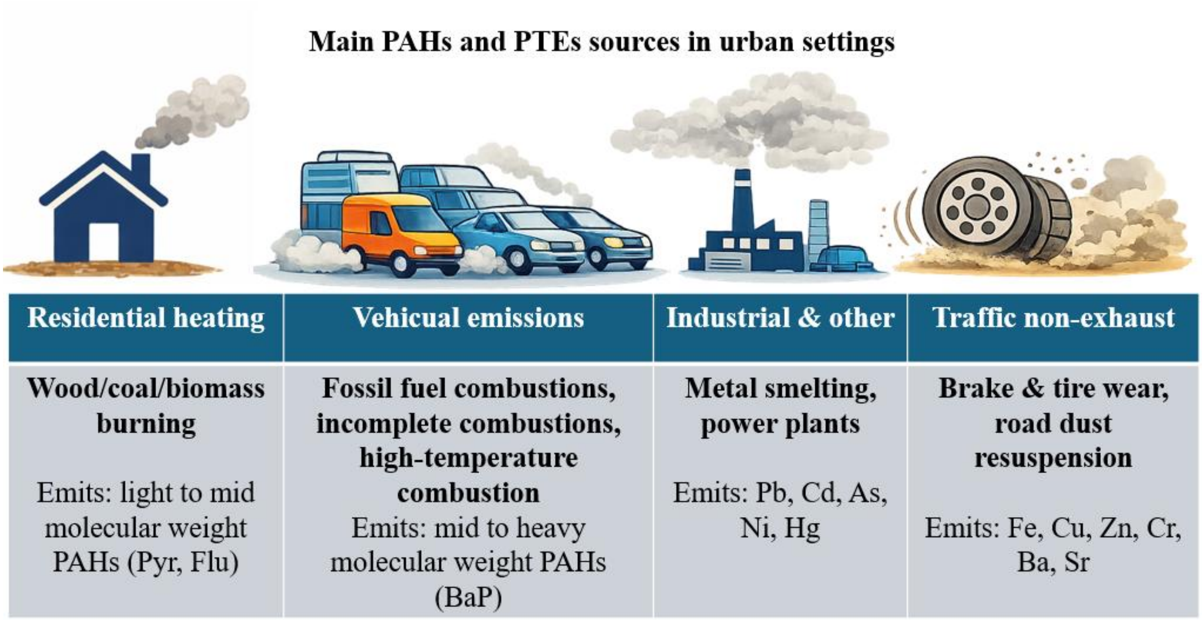


Figure 8. Main PAHs and PTEs sources in urban environments (adapted according to **Pacyna & Pacyna, 2001; Ravindra et al., 2008**)

Explainable ML strengthened these interpretations by quantifying the relative importance and directionality of predictors. Model performance was systematically higher for PAHs than for PTEs (PAHs R^2 up to ~0.68 vs. PTEs up to ~0.42), indicating that PAH variability is more consistently captured by the predictor set (meteorology + proxies for heating/traffic), while PTEs variability is more influenced by localised and episodic processes. SHAP analysis (Figure 9) highlighted temperature and solar radiation as main drivers for PAHs, with mechanistic interpretation discussed in Section 3.2.1.

In contrast, PTEs variability was more strongly associated with PM_{10}, NO_2, and site-related factors, supporting the interpretation that particle load, traffic intensity, co-pollutant conditions, and local source proximity are most important variables. Overall, the combined results indicate that wintertime elevations of PM_{10}-bound PAHs in Zagreb are primarily driven by the interaction between increased combustion emissions during the heating season and meteorological conditions that favour accumulation (low temperatures, reduced solar radiation, low wind speeds, and stable atmospheric conditions associated with temperature inversions), whereas PTEs trends reflect a stronger contribution of traffic-related non-exhaust processes and spatial heterogeneity. This differentiation is important for mitigation, as it implies that reductions in PAHs require measures targeting cold-season combustion sources, while effective control of PTEs additionally depends on traffic management and strategies that address non-exhaust emissions and resuspended road dust.

2.2.1 Seasonal and meteorological effects

The strong seasonal effects observed in PAHs concentrations is consistent with the physicochemical properties of these compounds. Their semi-volatile nature results in enhanced partitioning to PM at low temperatures, leading to increased persistence in winter. In contrast, during warmer months, volatilization and photochemical degradation dominate, reducing atmospheric concentrations (**Naydenova et al., 2022**). This inverse relationship between PAHs and temperature or solar radiation was clearly revealed by the RF and SHAP analyses, which identified these meteorological variables as main predictors of PAHs variability. SHAP plots indicated that temperature exerted a non-linear negative influence on PAH concentrations, a pattern consistent with temperature thresholds reported in other urban studies, where small decreases in cold conditions produced disproportionately large increases in PAH levels (**Michalicová et al., 2024; Shin et al., 2022**). This threshold-like behaviour is consistent with the temperature sensitivity of gas–particle partitioning for semi-volatile organics and with reduced atmospheric processing under low irradiance in winter. However, it should be noted that temperature is also strongly correlated with residential heating activity, meaning that its negative association with PAH concentrations likely reflects a combination of direct physicochemical effects and indirect emission-driven seasonality; disentangling these contributions remains inherently difficult in observational datasets affected by multicollinearity among meteorological and anthropogenic predictors.

For PTEs, meteorological impact was less notable, reflecting their predominantly mechanical emission mechanisms. While wind speed and atmospheric pressure influenced short-term dispersion, their overall temporal variability was modest compared to combustion-related PAHs. Elements such as Fe and Cu exhibited stable annual concentrations, explaining their connection to constant sources like traffic. In contrast, Cd and As showed stronger winter maximum concentrations, reflecting their partial association with combustion and coal-derived aerosols, consistent with findings reported in other studies (**Elminir, 2005; Grigoratos & Martini, 2015; Piscitello et al., 2021; Qi et al., 2016)**. Overall, PAHs were more strongly modulated by thermodynamic and photochemical conditions, whereas PTEs variability reflected a larger contribution from persistent emissions and local dispersion conditions.

2.2.2 Spatial distribution and influence of urban characteristics

Spatial differences among monitoring stations provided additional information into local source contributions. The SIG and ZG-3 stations, located in the southern part of the city near major traffic corridors, consistently recorded higher concentrations of Fe, Cu, Zn, and traffic-associated PAHs. This trend is partly connected to the prevailing north-to-south wind direction in Zagreb, which transports pollutants from the city centre toward the southern districts of Novi Zagreb, where the proximity to the Sava river and associated microclimatic conditions, including frequent fog formation and reduced mixing, promote pollutant accumulation, particularly during winter months (**Bešlić et al., 2007; Sopčić et al., 2024**). In contrast, IMI and ZG-1, situated in residential or mixed-use areas, exhibited relatively lower levels with a stronger winter heating signal. Comparable roadside-background contrasts and near-road gradients (with concentrations decreasing towards background within a few hundred metres from major roads) are widely documented in the literature (**Karner et al., 2010**). For PTEs, Cu and Fe (often with Zn) are well-established tracers of non-exhaust traffic emissions (brake/tyre wear and resuspension), which typically show the strongest enrichment at traffic-influenced sites (**Fussell et al., 2022; Grigoratos & Martini, 2015**).

These spatial contrasts were further reflected in NMF bi-plots, where traffic-associated compounds showed stronger loadings at SIG and ZG-3, while heating-related PAHs dominated at IMI during winter months. Such findings emphasize the importance of local urban morphology and land-use patterns in determining exposure gradients, supporting the argument that city-specific emission inventories should integrate land-use and traffic microdata rather than relying on generalized assumptions. The multi-site design of this research was thus critical in capturing urban variability, which is often overlooked in national monitoring networks.

2.2.3 Application of ML for identifying environmental drivers

Traditional linear approaches such as PCA and regression models are powerful for exploring variability and latent trends in pollution but are often limited when relationships between predictors and pollutant concentrations are non-linear. To address this, RF regression was applied in this research as a flexible, non-parametric method capable of capturing complex dependencies between predictors and pollutant concentrations (**Pan et al., 2024; Rubal & Kumar, 2018**). Air pollution time series typically exhibit strong non-linearities, interactions between emissions and meteorology, and partial collinearity among predictors (e.g., temperature, heating proxies and co-pollutants). In this context, tree-based ensemble models such as RF are widely used for meteorological normalisation and trend attribution because they flexibly capture non-linear responses without strong parametric assumptions. SHAP further supports transparent interpretation by quantifying the direction and magnitude of predictor contributions in a consistent additive framework (**Grange et al., 2018; Mallet, 2021**). RF models were developed separately for PAHs and PTEs using an expanded set of predictors, including meteorological parameters, gas consumption as a proxy for residential heating, NO_2 as an indicator of combustion processes, and temporal or spatial identifiers representing measurement stations and seasons.

RF models achieved higher predictive accuracy than the linear regression models applied in the earlier publication, particularly for PAHs, where R^2 reached up to 0.68, compared to 0.42 for PTEs. This difference highlights the inherently non-linear behavior of semi-volatile organic compounds, whose concentrations depend simultaneously on emission intensity, meteorological transformation, and atmospheric processes. For PTEs, where sources are often more mechanical and temporally stable, the predictive advantage of RF was smaller but still evident, particularly under variable dispersion conditions. The SHAP analysis, applied as a post-hoc interpretation to RF outputs, provided quantitative analysis into variable importance and direction (Figure 9).

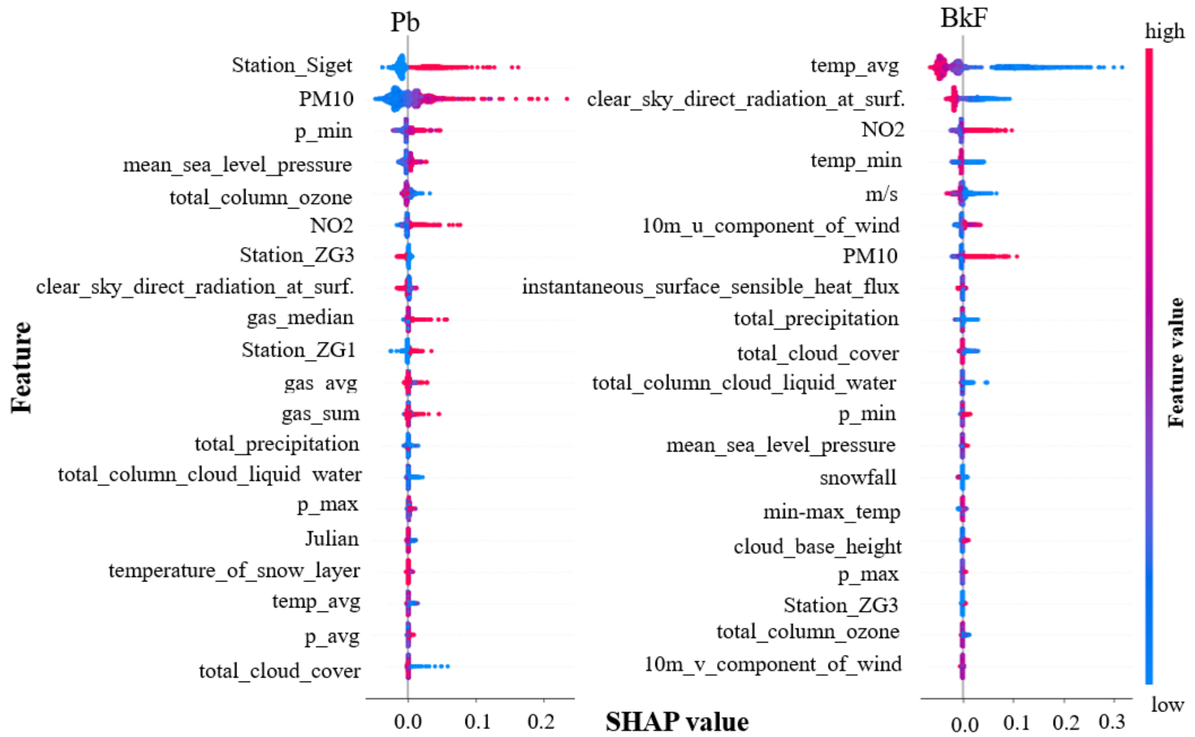


Figure 9. SHAP Feature Importance for best (BkF) and worst (Pb) models. The x-axis shows the SHAP value, representing the impact of each feature on the model output. Positive SHAP values indicate an increase in pollutant concentration, while negative values indicate a decrease.

SHAP results indicated that lower temperatures and lower solar radiation were associated with higher PAH concentrations, while gas consumption showed a positive contribution during the heating period. For PTEs, predictor importance was more strongly linked to PM_{10}, NO_2, and station-related effects, consistent with the importance of particle load, co-pollutant conditions, and local site context. It should be noted that predictor importance as quantified by SHAP reflects statistical associations within the model and does not constitute direct causal attribution or quantitative source apportionment. The identified predictors serve as proxies for underlying emission and transformation processes rather than as directly measured source contributions. Similar RF–SHAP workflows have been applied in air pollution studies to interpret meteorological controls and complex predictor interactions, supporting the relevance of explainable ML for environmental driver attribution (**Rybarczyk & Zalakeviciute, 2018**).

Higher wind speeds were generally associated with lower concentrations, consistent with more effective dispersion, although the strength and shape of this association varied by pollutant. Overall, combining receptor-style decomposition (NMF) with explainable ML (RF–SHAP) provided an interpretable framework for linking pollutant variability to measurable predictors.

This integration also helped distinguish broadly seasonal drivers from site-dependent influences.

2.3 Air–soil coherence and accumulation processes

Urban soils act as long-term receptors of atmospheric deposition and can preserve pollution that are not evident from air measurements alone **(Cui et al., 2024; P. Liu et al., 2023**). In this dissertation, the soil component enabled an evaluation of whether PM_{10}-bound PAHs and PTEs measured in Zagreb show deposition-related coherence with concentrations in nearby soils, and over which time scales such relationships are most detectable. Two depth intervals (0–5 cm and 5–10 cm) provided additional evidence on whether contaminants are primarily retained near the surface, as expected for deposition-driven accumulation.
The interpretation of air–soil relationships require consideration of both the anthropogenic and the geogenic background concentrations inherent to the local soil parent material. Elements such as Fe and Mn are major constituents of crustal minerals and their soil concentrations in Zagreb were broadly consistent with values previously reported for Zagreb soils (**Halamic & Miko, 2009; Romic & Romic, 2003**). However, the strong positive air–soil correlations observed for these elements indicate that atmospheric deposition from traffic-related resuspension provides a measurable anthropogenic contribution added to the lithogenic baseline.
Air–soil relationship was most consistent for particle-bound PTEs (e.g., Pb and Cu), supporting deposition-driven transfer for these elements, whereas PAH coherence was more compound- and time-window dependent, consistent with their semi-volatile behaviour and post-depositional processes. Accordingly, air–soil relationships for several PAHs strengthened when air concentrations were aggregated over longer windows (3–6 months), reflecting seasonal-scale integration rather than day-to-day variability.

To explore the correlation between atmospheric and soil pollution, Spearman correlation analysis was applied to paired datasets of air and soil concentrations for corresponding pollutants. Statistically significant positive correlations were identified for Pb ($R = 0.94$, $p = 0.005$) and Cu ($R = 0.71$) at 0–5 cm depth, while PAH associations, including heavier compounds such as BaP and BghiP, became statistically significant only when atmospheric concentrations were integrated over longer periods (3–6 months). These associations indicate that areas experiencing elevated atmospheric concentrations of these pollutants also tend to exhibit higher soil concentrations, consistent with deposition-driven transfer. The strength of correlations was higher for traffic-related pollutants (Pb, Cu, Zn, BaP, BkF), suggesting that near-surface soil layers act as an integrated archive of urban emission activity. Interestingly, low- to mid-molecular-weight PAHs such as Flu and Pyr showed weaker air–soil relationships, likely due to their higher volatility and lower deposition efficiency. These compounds are more susceptible to photochemical degradation and volatilization, leading to shorter atmospheric lifetimes and less accumulation in soil. Comparable findings have been recorded in other European urban soil studies, which indicated that atmospheric PAHs with fewer aromatic rings are more prone to re-emission and degradation, resulting in limited soil retention **(G. Li et al., 2018; Škrbić et al., 2017**). Conversely, high-ring PAHs and traffic-associated PTEs behave as persistent, low-mobility pollutants, whose deposition contributes to long-term soil pollution even after emission reductions. These results highlight that soil does not function as a uniform archive of atmospheric pollution, but rather as a selective, time-integrated receptor whose response varies across different pollutant. For PTEs, particularly Pb and Cu, the strong and consistent air–soil correlations reflect their chemical persistence, low mobility in soil, and predominantly particle-bound atmospheric transport, all of which favour efficient deposition and long-term retention in surface soils. For PAHs, however, the air–soil relationship is more complex: their semi-volatile nature means that post-depositional processes, including re-volatilisation under higher temperatures, photochemical and microbial degradation, partitioning to soil organic matter, and vertical redistribution through the soil profile, can modify or obscure the original atmospheric signal. This explains why air–soil correlations for PAHs were generally weaker, often negative at short time windows, and became significant only when atmospheric concentrations were integrated over longer periods, reflecting the time needed for cumulative deposition to exceed post-depositional losses. Consequently, the degree to which soil concentrations reflect atmospheric inputs is pollutant-specific and cannot be generalised across different measured compounds.

The correlation structure observed in Zagreb's dataset also revealed the temporal differences between emission events and soil accumulation. While air concentrations fluctuate on daily or seasonal scales, soil concentrations integrate these variations over years, smoothing out short-term dynamics. This makes soil a valuable environmental indicator for assessing historical and cumulative pollution exposure, complementing the temporally resolved information obtained from air quality monitoring. The observed accumulation trends can be explained by several connected processes governing pollutant transfer from air to soil. Dry deposition of particles is particularly important for PAHs and PTEs attached to PM_{10} particles, where gravitational settling and impaction dominate. In contrast, wet deposition through rainfall contributes to the transfer of both particle-bound and dissolved species. Once deposited, pollutant retention in soil depends on chemical and physical factors such as organic carbon content, pH, and texture. Soils rich in organic matter tend to adsorb and stabilize hydrophobic PAHs, while acidic conditions can enhance PTEs mobility. The predominance of clay-rich and moderately acidic soils in Zagreb promotes partial immobilization of PTEs, reducing leaching but enhancing long-term persistence (**Romic & Romic, 2003**).

2.3.1 Soil pollution profiles and spatial differences

Soil concentrations of Σ_{11}PAHs and the analysed PTEs showed spatial heterogeneity across the Zagreb urban area. Higher concentrations were observed at locations characterised by denser urban fabric and stronger traffic influence, while comparatively lower levels occurred at more open residential settings with reduced direct traffic exposure, consistent with the well-documented micro-scale dependence of urban soil pollution (**Różański et al., 2017**). Across sites, concentrations were generally higher in the 0–5 cm layer than in the 5–10 cm layer, indicating that recent and ongoing atmospheric inputs are primarily retained in the upper soil horizon, with limited downward redistribution on short time scales (**Alloway, 2013; Wilcke et al., 2005**).

Among PAHs, high-molecular-weight compounds (e.g., BaP, BghiP, IP) tended to show the clearest spatial contrasts and depth persistence. This behaviour is consistent with their low vapour pressures and strong affinity for PM and soil organic carbon, which favours deposition and stabilisation in surface soils and slows post-depositional losses through volatilisation and degradation **(Guerreiro et al., 2016; Wang et al., 2018; Wilcke et al., 2005**). In contrast, lighter PAHs typically exhibit weaker depth persistence due to greater volatility and higher susceptibility to degradation and re-emission after deposition, which can blur spatial gradients in soils (**Cachada et al., 2018**). A notable exception was the CEN station, where Σ11PAH concentrations at 5–10 cm consistently equalled or exceeded those at 0–5 cm, with the most extreme case in spring 2020 (524 vs. 106 µg/g). This inverted depth profile suggests that elevated PAH levels at CEN reflect local subsurface contamination in the old city centre rather than recent atmospheric deposition, possibly amplified by physical disturbance during post-earthquake clean-up activities in 2020.

For PTEs, Pb, Cu and Zn generally showed the most pronounced enrichment in urban soils, consistent with their association with traffic-related sources and particle-bound deposition. Pb often reflects legacy accumulation from historical leaded fuel emissions; the strong air–soil correlations observed in this study further suggest that ongoing atmospheric deposition continues to supplement these historically accumulated soil pools. Cu and Zn are widely used indicators of non-exhaust traffic emissions such as brake and tyre wear and road dust resuspension (**Grigoratos & Martini, 2015**). By contrast, Fe and Mn are major lithogenic elements, although a measurable anthropogenic contribution from atmospheric deposition cannot be excluded; their spatial gradients in urban soils are therefore generally weaker unless local resuspension or industrial point sources create measurable enrichment (**Alloway, 2013; Reimann, 1998; Wedepohl, 1995**). Beyond atmospheric deposition, urban soils are also subject to direct local inputs that can complicate source interpretation. These include fuel and motor oil leakage from vehicles, spillage of chemical substances, improper waste disposal, and the use of metals in construction and infrastructure works. Such point-source contributions operate independently of atmospheric pathways and may create localised enrichment patterns that are difficult to disentangle from deposition-driven accumulation, adding further complexity to the interpretation of soil pollution profiles in heterogeneous urban environments. The depth distribution of PTEs across the two sampled intervals further reflects their geochemical behaviour: elements with strong affinity for soil organic matter and fine mineral fractions, such as Pb and Cu, tend to accumulate in the uppermost horizon where organic matter content is highest, whereas elements with greater mobility under local pH and redox conditions may show less pronounced depth gradients (**Kabata-Pendias, 2010**). Overall, the observed soil profiles support a mixed urban pollution signature in which traffic proximity and local urban structure act as primary determinants of hotspot formation.

2.4 Comparison with European urban studies

When placed in the context of European urban air pollution research, the patterns observed in Zagreb are consistent with those reported for medium-sized Central and Southern European cities affected by a combination of residential heating and traffic emissions (*https://www.eea.europa.eu/en, accessed on 15 March 2026*). Across the region, wintertime peaks in particle-bound PAHs are commonly attributed to intensified residential combustion, compounded by stable atmospheric conditions such as temperature inversions that suppress vertical mixing and reduce pollutant dispersion. Studies from Ljubljana and Budapest similarly identify heating-related combustion and traffic as the dominant contributors to urban PAH mixtures, with higher cold-season concentrations and a stronger contribution of high-molecular-weight PAHs at traffic-influenced locations (**Drventić et al., 2023; Szaniszló & Ungváry, 2001**). In this respect, Zagreb follows a comparable source structure, with clear seasonality and source-resolving profiles that separate a more traffic-influenced PAHs from a cold-season combustion related one.

For PTEs in PM_{10}, the Zagreb results also align with evidence from other European cities where Fe, Cu, and Zn are repeatedly linked to traffic-related non-exhaust emissions and resuspension, while elements such as Cd, Pb, and As often reflect mixed influences including residual industrial contributions and combustion-related sources. Concentrations of PTEs in PM_{10}, including Pb, As, Cd, and Ni, remain below the EU target values for As, Cd and Ni and the limit value for Pb (Pb: 0.5 µg/ m^3; As: 6 ng/m^3; Cd: 5 ng/m^3; Ni: 20 ng/m^3) (Directive 2008/50/EC) throughout the study period, and are broadly comparable to levels reported at urban background stations in other Central and Southern European cities. Such results have been reported for cities including Barcelona and Milan and are broadly consistent with the station-dependent variability observed across Zagreb's monitoring network (**Amato et al., 2009; Bernardoni et al., 2011**). A methodological strength of the Zagreb framework is the use of gas consumption as an explicit heating proxy, which offers a direct, quantitative connection between heating demand and pollution dynamics and supports more specific interpretation of cold-season drivers than approaches relying only on calendar-based season definitions. Soil PTEs concentrations measured by portable XRF in this study are broadly consistent with values reported for urban soils in other Central European cities, with Pb, Cu and Zn showing the most pronounced anthropogenic enrichment. The XRF method applied here has been shown to provide reliable quantification for these elements, with strong correlations with ICP-MS and ICP-OES reference methods reported for Pb, Cu and Zn ($R^2 > 0.90$), while accuracy for Cd and Ni at lower concentration ranges may be more limited **(Jenkins et al., 2025; Romzaykina et al., 2024**). The observed enrichment of Pb, Cu and Zn in Zagreb's urban soils is consistent with patterns reported in urban geochemical research across European cities, where these elements are identified as markers of anthropogenic overprint on the natural geochemical baseline (**Ajmone-Marsan & Biasioli, 2010; Biasioli et al., 2007**).

A further contribution of this dissertation, compared with much of the European literature, is the integration of multi-station atmospheric observations with targeted soil sampling. Many European studies investigate air and soil pollution separately, whereas the Zagreb dataset allows evaluation of whether patterns observed in PM_{10} are reflected in longer-term accumulation in nearby urban soils. This inter-media and geochemical perspective support interpretation of urban pollution as a complex system, where short-term atmospheric variability and long-term soil deposition can be assessed together, strengthening the evidence base for both air quality management and long-term pollution assessment.

2.5 Environmental and health implications

The coexistence of elevated airborne concentrations and soil accumulation of carcinogenic PAHs such as BaP, as well as potentially toxic elements like Pb and Cd, revealed the potential for multi-pathway human exposure in urban environments. PAHs and PTEs in ambient air represent a significant public health concern in urban environments, as prolonged inhalation of PM-bound pollutants has been associated with respiratory and cardiovascular diseases, as well as genotoxic and carcinogenic effects **(Dominici et al., 2003; Kampa & Castanas, 2008; Vilcassim & Thurston, 2023**), as shown in Figure 10. While air inhalation represents an acute exposure route, soil and dust ingestion or dermal contact contribute to chronic, low-dose exposure, especially in children and populations living near busy roads. The persistence of these substances in soil also implies that reductions in atmospheric emissions may not immediately translate into lower human exposure, as resuspension and soil–air exchange can maintain background levels for years. Similar conclusions were drawn in other studies, which highlighted that soils act not only as sinks but also as secondary emission sources under certain meteorological conditions (**Cachada et al., 2018; X. Zhang et al., 2023**).

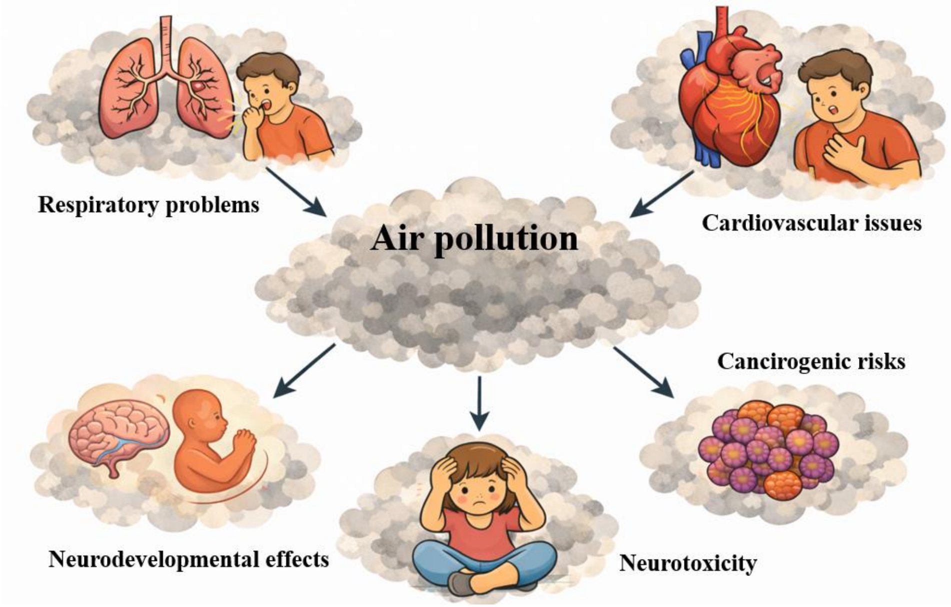


Figure 10. Conceptual overview of urban exposure pathways and major health outcomes associated with air pollution. The schematic is adapted from general mechanisms and health endpoints described in **(Kampa & Castanas, 2008)**, created using ChatGPT, version 5.2 and Adobe Illustrator, version 29.5.1.

Therefore, understanding the sources of both atmospheric and soil pollution is essential for long-term environmental management. The results from Zagreb indicate that traffic-related and heating-derived emissions not only impact short-term air quality but also have lasting effects in urban soils, with implications for environmental health risk assessment and urban planning. The spatio-temporal datasets, source profiles, and air–soil associations generated in this dissertation provide a foundation for quantitative health risk and environmental risk assessment, including inhalation risk from PM-bound PAHs and PTEs, as well as ingestion and dermal exposure pathways through contaminated urban soils.

2.6 Limitations and uncertainties

While this study provides valuable information about the spatio-temporal behaviour of PAHs and PTEs in PM and soil, as well as the impact of meteorological and anthropogenic factors, several limitations and uncertainties must be acknowledged. First, the resolution of the dataset for first paper, primarily based on monthly averages constrains the analysis of short-term pollution spikes and rapid meteorological changes, which could be important for understanding acute exposure risks. Higher temporal resolution (daily or hourly) could uncover different dynamics and improve the sensitivity of source apportionment what is used in second paper. Secondly, although gas consumption was used as a proxy for residential heating, this variable alone cannot fully capture the complexity and diversity of heating practices in urban areas. The lack of disaggregated data on the specific number of households using wood, coal, or other biomass fuels limits the accuracy of associating observed pollutant concentrations with specific emission sources (**Denier Van Der Gon et al., 2015; Du et al., 2017**). The assumption that gas consumption aligns with general heating demand is reasonable but introduces uncertainty, especially in neighbourhoods with mixed energy usage.

Another limitation lies in the reliance on data from a single or a limited number of monitoring stations, particularly when making broader inferences about urban-scale pollution behaviour. While IMI and other Zagreb air quality monitoring stations offer a valuable long-term dataset, the variability across microenvironments, influenced by traffic density, topography, and local industrial activities cannot be fully accounted for in this setup. The lack of vertical sampling and comparisons of indoor and ambient air further constrains the ability to track pollutant migration across environmental compartments. In terms of analytical methodology, while PCA and NMF offer powerful tools for dimensionality reduction and pattern recognition, they also carry certain assumptions and weaknesses. PCA assumes linear relationships and can obscure non-linear interactions among variables. NMF, though better suited for non-negative environmental data, is sensitive to the number of components selected and may not always give stable solutions. Interpretability of NMF components requires a degree of subjectivity, as factor profiles do not directly map to clearly defined emission sources without additional tracer information.

The regression models, while statistically valid, explain only part of the variability in NMF-derived components. This suggests the presence of additional, unmeasured variables influencing pollutant behaviour, such as industrial activity, localized combustion, atmospheric chemistry, or long-range transport. Moreover, multicollinearity, while addressed using PCA, still poses challenges for causal interpretation, especially in observational studies without experimental control.

Finally, uncertainties arise from measurement errors, detection limits, and sample handling procedures, particularly for trace-level analytes like PAHs and some PTEs which are present in low levels. Although some differences may arise if different analytical techniques are applied (for example, PTEs were quantified using different analytical techniques across compartments: ICP-MS for PM_{10} samples and XRF for soil samples).

Recognising these limitations does not undermine the study's conclusions, but it highlights the need for cautious interpretation and continued methodological refinement. Future research would benefit from higher temporal resolution, expanded spatial coverage, more detailed emission indicators, and further development of harmonised analytical workflows to strengthen comparability between atmospheric and soil datasets.

2.7 Future research direction

Building upon the insights from this study, several ideas for future research are identified to enhance the understanding of PAHs and PTEs in the environment, as well as to refine the methodologies used in air and soil pollution assessment. Among the immediate methodological priorities, the development and validation of an ICP-MS based analytical workflow for soil PTEs quantification is planned, which will enable direct cross-validation with the XRF measurements applied in this dissertation and strengthen the quantitative comparability of air and soil datasets in future studies.

Second, expanding the temporal scope of monitoring beyond the four-year period could offer deeper insights into long-term trends, enabling assessment of whether air quality management policies and changes in energy or transport infrastructure have had measurable effects.

Third, incorporating finer spatial resolution by adding more monitoring locations within the city and its surrounding regions would enable a better understanding of spatial heterogeneity and the influence of localized emission sources, such as specific industrial facilities, residential heating practices, or traffic patterns. This could also include high-resolution mapping using satellite data or mobile monitoring units. In particular, the systematic deployment of traffic counters near monitoring stations and along major roads would be highly valuable, as continuous and spatially resolved traffic flow data are not available for this area. Such data would allow more direct quantification of traffic-related emission contributions and improve the specificity of source attribution models.

Fourth, in-depth analysis of additional environmental matrices such as vegetation, sediments, and indoor dust could improve source identification and exposure assessment. Since soil was identified as an important receptor of airborne pollutants, further investigation into soil properties influencing pollutant retention and mobility would be valuable, including experimental studies under controlled conditions.

Fifth, the geochemical component of the soil analysis represents a particularly valuable results for future development. Future research should quantify the geogenic background concentrations of PTEs in Zagreb's urban soils by reference to local parent material composition and apply enrichment factor calculations using lithogenic reference elements (such as Fe or Al) to more rigorously distinguish natural from anthropogenic contributions. Geostatistical mapping of PTEs and PAHs across a denser sampling grid would enable the construction of urban geochemical baseline maps, which are essential for long-term contamination assessment and land-use planning. Expanding soil sampling to include deeper horizons and sediment cores from nearby water bodies would further support interpretation of long-term geochemical accumulation and vertical redistribution processes. The soil component of this dissertation, as the first systematic urban air-soil dataset of this kind in Zagreb, provides the foundation for such geochemical investigations and underscores the relevance of this research to the broader field of environmental and urban geochemistry.

Sixth, the development and application of advanced ML models, including deep learning architectures or hybrid approaches, could enhance the predictive accuracy and robustness of source apportionment analyses. Future studies may also integrate causal inference techniques to move beyond correlation and toward a more mechanistic understanding of pollution dynamics.

Seventh, combining chemical data with human exposure and health outcome data would offer a more comprehensive picture of environmental risks. This could be achieved through interdisciplinary collaboration involving environmental science, public health, and epidemiology.

Lastly, given the limitations of gas consumption as a proxy for heating-related emissions, future work should seek to incorporate more direct indicators of heating types and usage, as well as real-time traffic data, which would improve the accuracy of source attribution models.

Altogether, these research directions can contribute to a more comprehensive and actionable understanding of environmental pollution and human exposure and inform more effective mitigation strategies.

3 CONCLUSION

This dissertation provides a comprehensive investigation into the spatio-temporal patterns and environmental drivers of PAHs and PTEs in PM (PM_{10}) and urban soil, with a particular focus on the Zagreb area. By combining traditional statistical methods with modern ML approaches, such as NMF, PCA, and RF, this research offers new insights into pollutant behavior, source identification, and seasonal dynamics. The soil component of this work further contributes to the fields of environmental and urban geochemistry by examining the geochemical distribution of PTEs and PAHs in urban soil profiles, distinguishing anthropogenic enrichment from geogenic background contributions, and linking atmospheric deposition to long-term geochemical accumulation in surface soils.
The results indicate that air pollution in Zagreb is strongly shaped by seasonal factors, especially during the colder months when increased residential heating significantly elevates PAH concentrations. Compounds such as BaP, Flu, and Pyr show clear seasonal trends, closely linked to temperature and gas consumption patterns. Solar radiation further modulates PAH behavior through degradation pathways, particularly under high UV exposure. In contrast, PM_{10} and NO_2 serve as markers of broader anthropogenic activity and correlate with PTEs like As, Cd, and Pb, indicating sources such as traffic emissions, industrial combustion, and regional atmospheric transport.

Source identification through NMF revealed at least three potential pollution sources: (1) a traffic-related source characterised by higher-molecular-weight PAHs (BaA, Chry, BkF) and non-exhaust PTEs (Fe, Cu, Mn), (2) a residential heating profile associated with lower-molecular-weight PAHs (Flu, Pyr) and (3) a mixed industrial/heating influence with Cd, As, Pb, and Zn. These signatures were further supported through regression models and SHAP-based ML interpretations, showing the explanatory power of meteorological parameters, gas usage, and NO_2 levels in predicting pollutant concentrations. While PAHs exhibited more stable and predictable seasonal behavior, PTEs displayed more complexity due to their episodic or industrial origins.

Additionally, the spatial dimension of the study highlighted localized pollution hotspots, with higher concentrations of traffic- and industry-related pollutants at sites like SIG and ZG-3, while lower overall concentrations were observed at IMI and ZG-1, reflecting reduced industrial and traffic influence at these sites. Urban soils acted as long-term geochemical receptors of atmospheric pollution, with surface enrichment of Pb, Cu and Zn above geogenic background levels consistent with sustained anthropogenic deposition, while Fe and Mn concentrations remained closer to lithogenic baseline values, reflecting their predominantly crustal origin. This research is the first of its kind in Croatia to integrate high-resolution environmental data, meteorological variables, traffic data, energy usage statistics, and ML techniques to investigate air and soil pollution, and represents a novel contribution to environmental and urban geochemistry in the Croatian and broader Central European context. The findings underscore the necessity of seasonally adaptive air quality management, especially in relation to heating emissions in winter and traffic emissions year-round. Policymakers and urban planners can use these results to implement targeted mitigation strategies, such as cleaner heating initiatives, traffic regulation, and improved emission monitoring in identified hotspot areas.

Returning to the hypotheses formulated at the beginning of this research, the following conclusions can be drawn.

Hypothesis (i), that seasonal heating patterns drive PAHs and PTEs variability in air, was largely supported by the results. PAH concentrations showed clear seasonal trends linked to temperature, solar radiation, and gas consumption as a heating proxy. For PTEs, however, the seasonal signal was less pronounced, and their variability was more strongly associated with traffic-related and site-specific factors, indicating that hypothesis (i) applies more strongly to PAHs than to PTEs.

Hypothesis (ii), that traffic emissions represent a main source of PAHs and PTEs in both air and soil, was supported but not uniformly across all pollutants. NMF and SHAP analyses identified a distinct traffic-related component characterised by higher-molecular-weight PAHs (BaA, Chry, BkF) and non-exhaust PTEs (Fe, Cu, Mn), with the highest concentrations observed at traffic-influenced stations. However, the contribution of traffic relative to heating varied by compound and season, and the RF models indicated that traffic-related predictors were more important for PTEs than for PAHs, for which meteorological and heating-related variables dominated. In soil, traffic influence was supported by the enrichment of Cu and Zn at traffic-influenced locations, although spatial patterns for PAHs in soil did not consistently mirror the atmospheric traffic signal, with the highest soil PAH concentrations observed at the urban centre.

Hypothesis (iii), that urban soils accumulate PAHs and PTEs from the atmosphere, was partially supported. Strong and consistent air–soil correlations were observed for Pb and Cu, supporting deposition-driven accumulation for these persistent, particle-bound elements. For PAHs, however, air–soil relationships were weaker, compound-specific, and emerged only at longer temporal integration windows, reflecting the modifying influence of post-depositional processes such as volatilisation, degradation, and sorption to soil organic matter. Consequently, soil functions as a selective, time-integrated receptor rather than a uniform archive of atmospheric pollution, with the clearest deposition signals limited to persistent PTEs and high-molecular-weight PAHs. Seasonal variability in soil concentrations was limited, which is expected given that soil integrates inputs over years to decades and responds more slowly than the atmosphere to short-term emission changes.

Looking ahead, future research should aim to expand this work by incorporating additional monitoring sites, higher-resolution temporal data, and more detailed information on fuel types, traffic and emission sources. Developing urban geochemical baseline maps for Zagreb and expanding soil sampling to include deeper horizons and sediment cores would further strengthen the geochemical interpretation of long-term pollutant accumulation. Integrating satellite-based remote sensing and longer-term soil monitoring could also enhance the understanding of pollutant deposition and accumulation processes. Ultimately, this work contributes to a deeper understanding of urban environmental health and urban geochemistry and supports efforts toward cleaner, healthier urban living environments in Zagreb and beyond.

4 LITERATURE

Agudelo-Castañeda, D. M., & Teixeira, E. C. (2014). Seasonal changes, identification and source apportionment of PAH in PM 1.0. *Atmospheric Environment*, *96*, 186–200. https://doi.org/10.1016/j.atmosenv.2014.07.030

Ajmone-Marsan, F., & Biasioli, M. (2010). Trace Elements in Soils of Urban Areas. *Water, Air, & Soil Pollution*, *213*(1–4), 121–143. https://doi.org/10.1007/s11270-010-0372-6

Alahmad, B., Khraishah, H., Althalji, K., Borchert, W., Al-Mulla, F., & Koutrakis, P. (2023). Connections Between Air Pollution, Climate Change, and Cardiovascular Health. *Canadian Journal of Cardiology*, *39*(9), 1182–1190. https://doi.org/10.1016/j.cjca.2023.03.025

Albayrak, L., Türksoy, V. A., Khalilov, R., & Eftekhari, A. (2023). Investigation of heavy metal exposure and trace element levels in acute exacerbatıon of COPD. *Journal of King Saud University - Science*, *35*(1), 102422. https://doi.org/10.1016/j.jksus.2022.102422

Alloway, B. J. (Ed.). (2013). *Heavy Metals in Soils: Trace Metals and Metalloids in Soils and their Bioavailability* (Vol. 22). Springer Netherlands. https://doi.org/10.1007/978-94-007-4470-7

Alolayan, M. A., Almutairi, A., Aladwani, S. M., & Alkhamees, S. (2023). Investigating major sources of air pollution and improving spatiotemporal forecast accuracy using supervised machine learning and a proxy. *Journal of Engineering Research*, *11*(3), 87–93. https://doi.org/10.1016/j.jer.2023.100126

Amato, F., Pandolfi, M., Viana, M., Querol, X., Alastuey, A., & Moreno, T. (2009). Spatial and chemical patterns of PM10 in road dust deposited in urban environment. *Atmospheric Environment*, *43*(9), 1650–1659. https://doi.org/10.1016/j.atmosenv.2008.12.009

Ambade, B., Kumar, A., & Latif, M. (2021). *Emission sources, Characteristics and risk assessment of particulate bound Polycyclic Aromatic Hydrocarbons (PAHs) from traffic sites*. https://doi.org/10.21203/rs.3.rs-328364/v1

Anderson, J. O., Thundiyil, J. G., & Stolbach, A. (2012). Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health. *Journal of Medical Toxicology*, *8*(2), 166–175. https://doi.org/10.1007/s13181-011-0203-1

Anttila, P., Tuovinen, J.-P., & Niemi, J. V. (2011). Primary NO2 emissions and their role in the development of NO2 concentrations in a traffic environment. *Atmospheric Environment*, *45*(4), 986–992. https://doi.org/10.1016/j.atmosenv.2010.10.050

Arruti, A., Fernández-Olmo, I., & Irabien, Á. (2010). Evaluation of the contribution of local sources to trace metals levels in urban PM2.5 and PM10 in the Cantabria region (Northern Spain). *Journal of Environmental Monitoring*, *12*(7), 1451. https://doi.org/10.1039/b926740a

Ayoobi, A. W., Ahmadi, H., Inceoglu, M., & Pekkan, E. (2022). Seasonal impacts of buildings' energy consumption on the variation and spatial distribution of air pollutant over Kabul City: Application of Sentinel—5P TROPOMI products. *Air Quality, Atmosphere & Health*, *15*(1), 73–83. https://doi.org/10.1007/s11869-021-01085-9

Balcioğlu, E. B., Çevik, F. E., & Aksu, A. (2021). Source Determination and Seasonal Distribution of Polycyclic Aromatic Hydrocarbons (PAHs) in Urban Soil of the Megacity Istanbul. *Polycyclic Aromatic Compounds*, *41*(3), 626–634. https://doi.org/10.1080/10406638.2019.1610465

Bernardoni, V., Vecchi, R., Valli, G., Piazzalunga, A., & Fermo, P. (2011). PM10 source apportionment in Milan (Italy) using time-resolved data. *Science of The Total Environment*, *409*(22), 4788–4795. https://doi.org/10.1016/j.scitotenv.2011.07.048

Bešlić, I., Šega, K., Čačković, M., Bencetić-Klaić, Z., & Vučetić, V. (2007). *Influence of weather types on concentrations of metallic components in airborne PM10 in Zagreb*. 93–107.

Bhatnagar, A. (2022). Cardiovascular Effects of Particulate Air Pollution. *Annual Review of Medicine*, *73*(1), 393–406. https://doi.org/10.1146/annurev-med-042220-011549

Biasioli, M., Grčman, H., Kralj, T., Madrid, F., Díaz-Barrientos, E., & Ajmone-Marsan, F. (2007). Potentially Toxic Elements Contamination in Urban Soils: A Comparison of Three European Cities. *Journal of Environmental Quality*, *36*(1), 70–79. https://doi.org/10.2134/jeq2006.0254

Bizualem, B., Tefera, N., Angassa, K., & Feyisa, G. L. (2023). Spatial and Temporal Analysis of Particulate Matter and Gaseous Pollutants at Six Heavily Used Traffic Junctions in Megenagna, Addis Ababa, Ethiopia. *Aerosol Science and Engineering*, *7*(1), 118–130. https://doi.org/10.1007/s41810-022-00167-0

Błaszczyk, E., Rogula-Kozłowska, W., Klejnowski, K., Fulara, I., & Mielżyńska-Švach, D. (2017). Polycyclic aromatic hydrocarbons bound to outdoor and indoor airborne particles (PM2.5) and their mutagenicity and carcinogenicity in Silesian kindergartens, Poland. *Air Quality, Atmosphere & Health*, *10*(3), 389–400. https://doi.org/10.1007/s11869-016-0457-5

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Anal. Methods*, *6*(9), 2812–2831. https://doi.org/10.1039/C3AY41907J

Bukowska, B., Mokra, K., & Michałowicz, J. (2022). Benzo[a]pyrene—Environmental Occurrence, Human Exposure, and Mechanisms of Toxicity. *International Journal of Molecular Sciences*, *23*(11), 6348. https://doi.org/10.3390/ijms23116348

Cachada, A., Rocha-Santos, T., & Duarte, A. C. (2018). Soil and Pollution. In *Soil Pollution* (pp. 1–28). Elsevier. https://doi.org/10.1016/B978-0-12-849873-6.00001-7

Carslaw, D. (2005). Evidence of an increasing NO/NO emissions ratio from road traffic emissions. *Atmospheric Environment*, *39*(26), 4793–4802. https://doi.org/10.1016/j.atmosenv.2005.06.023

Castel, R., Bertoldo, R., Lebarillier, S., Noack, Y., Orsière, T., & Malleret, L. (2023). Toward an interdisciplinary approach to assess the adverse health effects of dust-containing polycyclic aromatic hydrocarbons (PAHs) and metal(loid)s on preschool children. *Environmental Pollution*, *336*, 122372. https://doi.org/10.1016/j.envpol.2023.122372

Castronuovo, G., Favia, G., Telesca, V., & Vammacigno, A. (2023). Analyzing the Interactions between Environmental Parameters and Cardiovascular Diseases Using Random Forest and SHAP Algorithms. *Reviews in Cardiovascular Medicine*, *24*(11), 330. https://doi.org/10.31083/j.rcm2411330

Chen, L.-C., Maciejczyk, P., & D. Thurston, G. (2022). Metals and air pollution. In *Handbook on the Toxicology of Metals* (pp. 137–182). Elsevier. https://doi.org/10.1016/B978-0-12-823292-7.00004-8

Cigula, M., Fugaš, F., & Valić, F. (1972). Određivanje 1,12-benzperilena, 1,2-benzpirena i kancerogena 3,4-benzpirena u atmosferi metodom ugrijanog prstena. *Arhiv Za Higijenu Rada*.

Cirillo, T., Montuori, P., Mainardi, P., Russo, I., Triassi, M., & Amodio-Cocchieri, R. (2006). Multipathway Polycyclic Aromatic Hydrocarbon and Pyrene Exposure Among Children Living in Campania (Italy). *Journal of Environmental Science and Health, Part A*, *41*(10), 2089–2107. https://doi.org/10.1080/10934520600867854

Cofield, N., Banks, M. K., & Schwab, A. P. (2007). Evaluation of hydrophobicity in PAH-contaminated soils during phytoremediation. *Environmental Pollution*, *145*(1), 60–67. https://doi.org/10.1016/j.envpol.2006.03.040

Cui, H., Zhao, Y., Hu, K., Xia, R., Zhou, J., & Zhou, J. (2024). Impacts of atmospheric deposition on the heavy metal mobilization and bioavailability in soils amended by lime. *Science of The Total Environment*, *914*, 170082. https://doi.org/10.1016/j.scitotenv.2024.170082

Denier Van Der Gon, H. A. C., Bergström, R., Fountoukis, C., Johansson, C., Pandis, S. N., Simpson, D., & Visschedijk, A. J. H. (2015). Particulate emissions from residential wood combustion in Europe – revised estimates and an evaluation. *Atmospheric Chemistry and Physics*, *15*(11), 6503–6519. https://doi.org/10.5194/acp-15-6503-2015

Dominici, F., Sheppard, L., & Clyde, M. (2003). Health Effects of Air Pollution: A Statistical Review. *International Statistical Review*, *71*(2), 243–276. https://doi.org/10.1111/j.1751-5823.2003.tb00195.x

Drventić, I., Glumac, M., Carev, I., & Kroflič, A. (2023). Seasonality of Polyaromatic Hydrocarbons (PAHs) and Their Derivatives in PM2.5 from Ljubljana, Combustion Aerosol Source Apportionment, and Cytotoxicity of Selected Nitrated Polyaromatic Hydrocarbons (NPAHs). *Toxics*, *11*(6), 518. https://doi.org/10.3390/toxics11060518

Du, W., Shen, G., Chen, Y., Zhu, X., Zhuo, S., Zhong, Q., Qi, M., Xue, C., Liu, G., Zeng, E., Xing, B., & Tao, S. (2017). Comparison of air pollutant emissions and household air quality in rural homes using improved wood and coal stoves. *Atmospheric Environment*, *166*, 215–223. https://doi.org/10.1016/j.atmosenv.2017.07.029

Duffus, J. H. (2002). "Heavy metals" a meaningless term? (IUPAC Technical Report). *Pure and Applied Chemistry*, *74*(5), 793–807. https://doi.org/10.1351/pac200274050793

Duffus, J. H. (2003). Toxicology of Metals—Science Confused by Poor Use of Terminology. *Archives of Environmental Health: An International Journal*, *58*(5), 263–266. https://doi.org/10.3200/AEOH.58.5.263-266

Dvorská, A., Lammel, G., & Klánová, J. (2011). Use of diagnostic ratios for studying source apportionment and reactivity of ambient polycyclic aromatic hydrocarbons over Central Europe. *Atmospheric Environment*, *45*(2), 420–427. https://doi.org/10.1016/j.atmosenv.2010.09.063

Echendu, A. J., Okafor, H. F., & Iyiola, O. (2022). Air Pollution, Climate Change and Ecosystem Health in the Niger Delta. *Social Sciences*, *11*(11), 525. https://doi.org/10.3390/socsci11110525

Elminir, H. K. (2005). Dependence of urban air pollutants on meteorology. *Science of The Total Environment*, *350*(1–3), 225–237. https://doi.org/10.1016/j.scitotenv.2005.01.043

Essamlali, I., Nhaila, H., & El Khaili, M. (2024). Supervised Machine Learning Approaches for Predicting Key Pollutants and for the Sustainable Enhancement of Urban Air Quality: A Systematic Review. *Sustainability*, *16*(3), 976. https://doi.org/10.3390/su16030976

European Commission. Joint Research Centre. (2019). *European guide on air pollution source apportionment with receptor models: Revised version 2019.* Publications Office. https://data.europa.eu/doi/10.2760/439106

Finardi, S., Radice, P., Cecinato, A., Gariazzo, C., Gherardi, M., & Romagnoli, P. (2017). Seasonal variation of PAHs concentration and source attribution through diagnostic ratios analysis. *Urban Climate*, *22*, 19–34. https://doi.org/10.1016/j.uclim.2015.12.001

Fussell, J. C., Franklin, M., Green, D. C., Gustafsson, M., Harrison, R. M., Hicks, W., Kelly, F. J., Kishta, F., Miller, M. R., Mudway, I. S., Oroumiyeh, F., Selley, L., Wang, M., & Zhu, Y. (2022). A Review of Road Traffic-Derived Non-Exhaust Particles: Emissions, Physicochemical Characteristics, Health Risks, and Mitigation Measures. *Environmental Science & Technology*, *56*(11), 6813–6835. https://doi.org/10.1021/acs.est.2c01072

Gerić, M., Pehnec, G., Matković, K., Rinkovec, J., Jakovljević, I., Godec, R., Žužul, S., Bešlić, I., Cvitković, A., Delić, L., Wild, P., Guseva Canu, I., Hopf, N. B., & Gajski, G. (2024). Air Pollution and Primary DNA Damage among Zagreb (Croatia) Residents: A Cross-Sectional Study. *Journal of Xenobiotics*, *14*(1), 368–379. https://doi.org/10.3390/jox14010023

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., & Hueglin, C. (2018). Random forest meteorological normalisation models for Swiss PM_{10} trend analysis. *Atmospheric Chemistry and Physics*, *18*(9), 6223–6239. https://doi.org/10.5194/acp-18-6223-2018

Grigoratos, T., & Martini, G. (2015). Brake wear particle emissions: A review. *Environmental Science and Pollution Research*, *22*(4), 2491–2504. https://doi.org/10.1007/s11356-014-3696-8

Guerreiro, C. B. B., Horálek, J., De Leeuw, F., & Couvidat, F. (2016). Benzo(a)pyrene in Europe: Ambient air concentrations, population exposure and health effects. *Environmental Pollution*, *214*, 657–667. https://doi.org/10.1016/j.envpol.2016.04.081

Gunawardena, J., Egodawatta, P., Ayoko, G. A., & Goonetilleke, A. (2012). Role of traffic in atmospheric accumulation of heavy metals and polycyclic aromatic hydrocarbons. *Atmospheric Environment*, *54*, 502–510. https://doi.org/10.1016/j.atmosenv.2012.02.058

Guo, F., Tang, M., Wang, X., Yu, Z., Wei, F., Zhang, X., Jin, M., Wang, J., Xu, D., Chen, Z., & Chen, K. (2022). Characteristics, sources, and health risks of trace metals in PM2.5. *Atmospheric Environment*, *289*, 119314. https://doi.org/10.1016/j.atmosenv.2022.119314

Halamic, J., & Miko, S. (Eds.). (2009). *Geokemijski atlas Republike Hrvatske: = Geochemical atlas of the Republic of Croatia*. HGI Hrvatski geoloski institut.

Han, J., Liu, H., Xiong, H., & Yang, J. (2023). Semi-Supervised Air Quality Forecasting via Self-Supervised Hierarchical Graph Neural Network. *IEEE Transactions on Knowledge and Data Engineering*, *35*(5), 5230–5243. https://doi.org/10.1109/TKDE.2022.3149815

Hopke, P. K. (2003). Recent developments in receptor modeling. *Journal of Chemometrics*, *17*(5), 255–265. https://doi.org/10.1002/cem.796

Jain, S., Kaur, N., Verma, S., Kavita, Hosen, A. S. M. S., & Sehgal, S. S. (2022). Use of Machine Learning in Air Pollution Research: A Bibliographic Perspective. *Electronics*, *11*(21), 3621. https://doi.org/10.3390/electronics11213621

Jakovljević, I., Mešić, I., & Pehnec, G. (2022). Soil pollution of the Labin city area with polycyclic aromatic hydrocarbons derived from Raša coal mining and associated industries. *Rudarsko-Geološko-Naftni Zbornik*, *37*(1), 139–150. https://doi.org/10.17794/rgn.2022.1.12

Jakovljević, I., Pehnec, G., Šišović, A., Vađić, V., Davila, S., & Godec, R. (2016). Concentrations of PAHs and other gaseous pollutants in the atmosphere of a rural area. *Journal of Environmental Science and Health, Part A*, *51*(9), 707–713. https://doi.org/10.1080/10934529.2016.1170431

Jakovljević, I., Pehnec, G., Vađić, V., Čačković, M., Tomašić, V., & Jelinić, J. D. (2018). Polycyclic aromatic hydrocarbons in PM10, PM2.5 and PM1 particle fractions in an urban area. *Air Quality, Atmosphere & Health*, *11*(7), 843–854. https://doi.org/10.1007/s11869-018-0603-3

Jakovljević, I., Pehnec, G., Vadjić, V., Šišović, A., Davila, S., & Bešlić, I. (2015). Carcinogenic activity of polycyclic aromatic hydrocarbons bounded on particle fraction. *Environmental Science and Pollution Research*, *22*(20), 15931–15940. https://doi.org/10.1007/s11356-015-4777-z

Jakovljević, I., Smoljo, I., Sever Štrukil, Z., & Pehnec, G. (2023). Carcinogenic Activity and Risk Assessment of PAHs in Ambient Air: PM10 Particle Fraction and Bulk Deposition. *Toxics*, *11*(3), 228. https://doi.org/10.3390/toxics11030228

Jakovljević, I., Štrukil, Z. S., Pehnec, G., Horvat, T., Sanković, M., Šumanovac, A., Davila, S., Račić, N., & Gajski, G. (2025). Ambient air pollution and carcinogenic activity at three different urban locations. *Ecotoxicology and Environmental Safety*, *289*, 117704. https://doi.org/10.1016/j.ecoenv.2025.117704

Jedrychowski, W. A., Perera, F. P., Maugeri, U., Majewska, R., Mroz, E., Flak, E., Camann, D., Sowa, A., & Jacek, R. (2015). Long term effects of prenatal and postnatal airborne PAH exposures on ventilatory lung function of non-asthmatic preadolescent children. Prospective birth cohort study in Krakow. *Science of The Total Environment*, *502*, 502–509. https://doi.org/10.1016/j.scitotenv.2014.09.051

Jenkins, E. M., Galbraith, J., & Paltseva, A. A. (2025). Portable X-ray fluorescence as a tool for urban soil contamination analysis: Accuracy, precision, and practicality. *SOIL*, *11*(2), 565–582. https://doi.org/10.5194/soil-11-565-2025

Kabata-Pendias, A. (2010). *Trace Elements in Soils and Plants* (0 ed.). CRC Press. https://doi.org/10.1201/b10158

Kamińska, J. A. (2018). The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: A case study in Wrocław. *Journal of Environmental Management*, *217*, 164–174. https://doi.org/10.1016/j.jenvman.2018.03.094

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, *151*(2), 362–367. https://doi.org/10.1016/j.envpol.2007.06.012

Karner, A. A., Eisinger, D. S., & Niemeier, D. A. (2010). Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. *Environmental Science & Technology*, *44*(14), 5334–5344. https://doi.org/10.1021/es100008x

Konduracka, E., & Rostoff, P. (2022). Links between chronic exposure to outdoor air pollution and cardiovascular diseases: A review. *Environmental Chemistry Letters*, *20*(5), 2971–2988. https://doi.org/10.1007/s10311-022-01450-9

Kowalska, J. B., Mazurek, R., Gąsiorek, M., & Zaleski, T. (2018). Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination–A review. *Environmental Geochemistry and Health*, *40*(6), 2395–2420. https://doi.org/10.1007/s10653-018-0106-z

Lee, B.-K. (2010). Sources, Distribution and Toxicity of Polyaromatic Hydrocarbons (PAHs) in Particulate Matter. In V. Villanyi (Ed.), *Air Pollution*. Sciyo. https://doi.org/10.5772/10045

Lekinwala, N. L., & Bhushan, M. (2022). Generalised non-negative matrix factorisation for air pollution source apportionment. *Science of The Total Environment*, *839*, 156294. https://doi.org/10.1016/j.scitotenv.2022.156294

Li, D., Zheng, J., Yang, M., Meng, Y., Yu, X., Zhou, H., Tong, L., Wang, K., Li, Y.-F., Wang, X., & Xiao, H. (2023). Atmospheric wet deposition of trace metal elements: Monitoring and modelling. *Science of The Total Environment*, *893*, 164880. https://doi.org/10.1016/j.scitotenv.2023.164880

Li, G., Sun, G. -X., Ren, Y., Luo, X. -S., & Zhu, Y. -G. (2018). Urban soil and human health: A review. *European Journal of Soil Science*, *69*(1), 196–215. https://doi.org/10.1111/ejss.12518

Liao, Q., Zhu, M., Wu, L., Pan, X., Tang, X., & Wang, Z. (2020). Deep Learning for Air Quality Forecasts: A Review. *Current Pollution Reports*, *6*(4), 399–409. https://doi.org/10.1007/s40726-020-00159-z

Liu, J., Kang, H., Tao, W., Li, H., He, D., Ma, L., Tang, H., Wu, S., Yang, K., & Li, X. (2023). A spatial distribution – Principal component analysis (SD-PCA) model to assess pollution of heavy metals in soil. *Science of The Total Environment*, *859*, 160112. https://doi.org/10.1016/j.scitotenv.2022.160112

Liu, P., Wu, Q., Hu, W., Tian, K., Huang, B., & Zhao, Y. (2023). Effects of atmospheric deposition on heavy metals accumulation in agricultural soils: Evidence from field monitoring and Pb isotope analysis. *Environmental Pollution*, *330*, 121740. https://doi.org/10.1016/j.envpol.2023.121740

Liu, X., Schnelle-Kreis, J., Schloter-Hai, B., Ma, L., Tai, P., Cao, X., Yu, C., Adam, T., & Zimmermann, R. (2019). Analysis of PAHs Associated with PM10 and PM2.5 from Different Districts in Nanjing. *Aerosol and Air Quality Research*, *19*(10), 2294–2307. https://doi.org/10.4209/aaqr.2019.06.0301

Maharaj Kumari, K., & Lakhani, A. (2018). PAHs in Gas and Particulate Phases: Measurement and Control. In T. Gupta, A. K. Agarwal, R. A. Agarwal, & N. K. Labhsetwar (Eds.), *Environmental Contaminants* (pp. 43–75). Springer Singapore. https://doi.org/10.1007/978-981-10-7332-8_3

Mallet, M. D. (2021). Meteorological normalisation of PM10 using machine learning reveals distinct increases of nearby source emissions in the Australian mining town of Moranbah. *Atmospheric Pollution Research*, *12*(1), 23–35. https://doi.org/10.1016/j.apr.2020.08.001

Matković, K., Jurič, A., Jakovljević, I., Kazensky, L., Milić, M., Kašuba, V., Davila, S., Pehnec, G., Brčić Karačonji, I., Cvitković, A., Wild, P., Guseva Canu, I., Hopf, N. B., Gajski, G., & Gerić, M. (2025). Evaluating air pollution and BTEX exposure effects on DNA damage: A human biomonitoring study in Zagreb, Croatia. *Atmospheric Environment*, *343*, 121004. https://doi.org/10.1016/j.atmosenv.2024.121004

Mehta, S., Shin, H., Burnett, R., North, T., & Cohen, A. J. (2013). Ambient particulate air pollution and acute lower respiratory infections: A systematic review and implications for estimating the global burden of disease. *Air Quality, Atmosphere, & Health*, *6*(1), 69–83. https://doi.org/10.1007/s11869-011-0146-3

Michalicová, R., Hegrová, J., Svoboda, J., & Ličbinský, R. (2024). Seasonal and spatial variations of arsenic and its species in particulate matter in an urban environment of Brno, Czech Republic. *Environmental Science and Pollution Research*, *31*(43), 55251–55262. https://doi.org/10.1007/s11356-024-34645-4

Morillo, E., Romero, A. S., Madrid, L., Villaverde, J., & Maqueda, C. (2007). Characterization and Sources of PAHs and Potentially Toxic Metals in Urban Environments of Sevilla (Southern Spain). *Water, Air, and Soil Pollution*, *187*(1–4), 41–51. https://doi.org/10.1007/s11270-007-9495-9

Mostert, M. M. R., Ayoko, G. A., & Kokot, S. (2010). Application of chemometrics to analysis of soil pollutants. *TrAC Trends in Analytical Chemistry*, *29*(5), 430–445. https://doi.org/10.1016/j.trac.2010.02.009

Naydenova, St., Veli, A., Mustafa, Z., Hudai, S., Hristova, E., & Gonsalvesh-Musakova, L. (2022). Atmospheric levels, distribution, sources, correlation with meteorological parameters and other pollutants and health risk of PAHs bound in $PM_{2.5}$ and PM_{10} in Burgas, Bulgaria – a case study. *Journal of Environmental Science and Health, Part A*, *57*(4), 306–317. https://doi.org/10.1080/10934529.2022.2060669

Ozaki, N., Nitta, K., & Fukushima, T. (2006). Dispersion and dry and wet deposition of PAHs in an atmospheric environment. *Water Science and Technology*, *53*(2), 215–224. https://doi.org/10.2166/wst.2006.055

Pacyna, J. M., & Pacyna, E. G. (2001). An assessment of global and regional emissions of trace metals to the atmosphere from anthropogenic sources worldwide. *Environmental Reviews*, *9*(4), 269–298. https://doi.org/10.1139/a01-012

Pan, M., Xia, B., Huang, W., Ren, Y., & Wang, S. (2024). PM2.5 Concentration Prediction Model Based on Random Forest and SHAP. *International Journal of Pattern Recognition and Artificial Intelligence*, *38*(05), 2452012. https://doi.org/10.1142/S0218001424520128

Pandey, P., Patel, D. K., Khan, A. H., Barman, S. C., Murthy, R. C., & Kisku, G. C. (2013). Temporal distribution of fine particulates ($PM_{2.5}$, PM_{10}), potentially toxic metals, PAHs and Metal-bound carcinogenic risk in the population of Lucknow City, India. *Journal of Environmental Science and Health, Part A*, *48*(7), 730–745. https://doi.org/10.1080/10934529.2013.744613

Pandey, S. K., Kim, K.-H., & Brown, R. J. C. (2011). A review of techniques for the determination of polycyclic aromatic hydrocarbons in air. *TrAC Trends in Analytical Chemistry*, *30*(11), 1716–1739. https://doi.org/10.1016/j.trac.2011.06.017

Pehnec, G., & Jakovljević, I. (2018). Carcinogenic Potency of Airborne Polycyclic Aromatic Hydrocarbons in Relation to the Particle Fraction Size. *International Journal of Environmental Research and Public Health*, *15*(11), 2485. https://doi.org/10.3390/ijerph15112485

Peng, Z., Zhang, B., Wang, D., Niu, X., Sun, J., Xu, H., Cao, J., & Shen, Z. (2024). Application of machine learning in atmospheric pollution research: A state-of-art review. *Science of The Total Environment*, *910*, 168588. https://doi.org/10.1016/j.scitotenv.2023.168588

Piscitello, A., Bianco, C., Casasso, A., & Sethi, R. (2021). Non-exhaust traffic emissions: Sources, characterization, and mitigation measures. *Science of The Total Environment*, *766*, 144440. https://doi.org/10.1016/j.scitotenv.2020.144440

Polichetti, G., Cocco, S., Spinali, A., Trimarco, V., & Nunziata, A. (2009). Effects of particulate matter (PM10, PM2.5 and PM1) on the cardiovascular system. *Toxicology*, *261*(1–2), 1–8. https://doi.org/10.1016/j.tox.2009.04.035

Popoola, L. T., Adebanjo, S. A., & Adeoye, B. K. (2018). Assessment of atmospheric particulate matter and heavy metals: A critical review. *International Journal of Environmental Science and Technology*, *15*(5), 935–948. https://doi.org/10.1007/s13762-017-1454-4

Poster, D. L., Schantz, M. M., Sander, L. C., & Wise, S. A. (2006). Analysis of polycyclic aromatic hydrocarbons (PAHs) in environmental samples: A critical review of gas chromatographic (GC) methods. *Analytical and Bioanalytical Chemistry*, *386*(4), 859–881. https://doi.org/10.1007/s00216-006-0771-0

Potter, N. A., Meltzer, G. Y., Avenbuan, O. N., Raja, A., & Zelikoff, J. T. (2021). Particulate Matter and Associated Metals: A Link with Neurotoxicity and Mental Health. *Atmosphere*, *12*(4), 425. https://doi.org/10.3390/atmos12040425

Qi, L., Chen, M., Ge, X., Zhang, Y., & Guo, B. (2016). Seasonal Variations and Sources of 17 Aerosol Metal Elements in Suburban Nanjing, China. *Atmosphere*, *7*(12), 153. https://doi.org/10.3390/atmos7120153

Raaschou-Nielsen, O., Andersen, Z. J., Hvidberg, M., Jensen, S. S., Ketzel, M., Sørensen, M., Hansen, J., Loft, S., Overvad, K., & Tjønneland, A. (2011). Air pollution from traffic and cancer incidence: A Danish cohort study. *Environmental Health*, *10*(1), 67. https://doi.org/10.1186/1476-069X-10-67

Račić, N., Pehnec, G., Jakovljević, I., Štrukil, Z. S., Mureddu, F., Forsmann, M., & Lovrić, M. (2025). Machine learning analysis of drivers of differences in PAH content between PM1 and PM10 in Zagreb, Croatia. *Atmospheric Pollution Research*, *16*(7), 102541. https://doi.org/10.1016/j.apr.2025.102541

Račić, N., Ružičić, S., Terzić, T., Pehnec, G., Jakovljević, I., Štrukil, Z. S., Žužul, S., Rinkovec, J., & Lovrić, M. (2024). Analyzing the relationship between gas consumption and airborne pollutants: Case study of Zagreb, Croatia. *Air Quality, Atmosphere & Health*. https://doi.org/10.1007/s11869-024-01655-7

Ravindra, K., Sokhi, R., & Vangrieken, R. (2008). Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation. *Atmospheric Environment*, *42*(13), 2895–2921. https://doi.org/10.1016/j.atmosenv.2007.12.010

Reimann, C. (with Caritat, P. de). (1998). *Chemical Elements in the Environment: Factsheets for the Geochemist and Environmental Scientist*. Springer Berlin / Heidelberg.

Ribeiro, J., Silva, T., Filho, J. G. M., & Flores, D. (2012). Polycyclic aromatic hydrocarbons (PAHs) in burning and non-burning coal waste piles. *Journal of Hazardous Materials*, *199–200*, 105–110. https://doi.org/10.1016/j.jhazmat.2011.10.076

Rinkovec, J., Pehnec, G., Godec, R., Davila, S., & Bešlić, I. (2018). Spatial and temporal distribution of platinum, palladium and rhodium in Zagreb air. *Science of The Total Environment*, *636*, 456–463. https://doi.org/10.1016/j.scitotenv.2018.04.295

Romic, M., & Romic, D. (2003). Heavy metals distribution in agricultural topsoils in urban area. *Environmental Geology*, *43*(7), 795–805. https://doi.org/10.1007/s00254-002-0694-9

Romzaykina, O. N., Slukovskaya, M. V., Paltseva, A. A., Losev, A. I., Korneykova, M. V., & Vasenev, V. I. (2024). Rapid assessment of soil contamination by potentially toxic metals in the green spaces of Moscow megalopolis using the portable X-ray analyzer. *Journal of Soils and Sediments*. https://doi.org/10.1007/s11368-024-03758-1

Różański, S., Jaworska, H., Matuszczak, K., Nowak, J., & Hardy, A. (2017). Impact of highway traffic and the acoustic screen on the content and spatial distribution of heavy metals in soils. *Environmental Science and Pollution Research*, *24*(14), 12778–12786. https://doi.org/10.1007/s11356-017-8910-z

Rubal, & Kumar, D. (2018). Evolving Differential evolution method with random forest for prediction of Air Pollution. *Procedia Computer Science*, *132*, 824–833. https://doi.org/10.1016/j.procs.2018.05.094

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review. *Applied Sciences*, *8*(12), 2570. https://doi.org/10.3390/app8122570

Schraufnagel, D. E. (2020). The health effects of ultrafine particles. *Experimental & Molecular Medicine*, *52*(3), 311–317. https://doi.org/10.1038/s12276-020-0403-3

Shin, S. M., Lee, J. Y., Shin, H. J., & Kim, Y. P. (2022). Seasonal variation and source apportionment of Oxygenated Polycyclic Aromatic Hydrocarbons (OPAHs) and Polycyclic Aromatic Hydrocarbons (PAHs) in PM2.5 in Seoul, Korea. *Atmospheric Environment*, *272*, 118937. https://doi.org/10.1016/j.atmosenv.2022.118937

Šišović, A., Pehnec, G., Jakovljević, I., Šilović Hujić, M., Vađić, V., & Bešlić, I. (2012). Polycyclic Aromatic Hydrocarbons at Different Crossroads in Zagreb, Croatia. *Bulletin of Environmental Contamination and Toxicology*, *88*(3), 438–442. https://doi.org/10.1007/s00128-011-0516-4

Šišović, A., Škrbec, A., Vadjić, V., Kalinić, N., & Hršak, J. (2002). PAH Levels and Profiles in the Suspended Particulate Matter in Zagreb through Four Seasons. *Environmental*

Monitoring and Assessment, *74*(3), 217–224. https://doi.org/10.1023/A:1014228321207

Šišović, A., Vadjić, Z., Šega, K., Bešlić, I., & Vadjić, V. (2005). Comparison between PAH Mass Concentrations Measured in PM10 and Pm2.5 Particle Fractions. *Bulletin of Environmental Contamination and Toxicology*, *75*(1), 121–126. https://doi.org/10.1007/s00128-005-0727-7

Siudek, P. (2022). Compositional and seasonal differences of gas and particle phase polycyclic aromatic hydrocarbons (PAHs) over the southern Baltic Sea coast. *Scientific Reports*, *12*(1), 21005. https://doi.org/10.1038/s41598-022-25666-5

Škrbić, B. D., Đurišić-Mladenović, N., Tadić, Đ. J., & Cvejanov, J. Đ. (2017). Polycyclic aromatic hydrocarbons in urban soil of Novi Sad, Serbia: Occurrence and cancer risk assessment. *Environmental Science and Pollution Research*, *24*(19), 16148–16159. https://doi.org/10.1007/s11356-017-9194-z

Soffianian, A., Madani, E., & Arabi, M. (2014). Risk assessment of heavy metal soil pollution through principal components analysis and false color composition in Hamadan Province, Iran. *Environmental Systems Research*, *3*(1), 3. https://doi.org/10.1186/2193-2697-3-3

Sopčić, S., Pehnec, G., & Bešlić, I. (2024). Specific biomass burning tracers in air pollution in Zagreb, Croatia. *Atmospheric Pollution Research*, *15*(8), 102176. https://doi.org/10.1016/j.apr.2024.102176

Sørensen, M., Autrup, H., Møller, P., Hertel, O., Jensen, S. S., Vinzents, P., Knudsen, L. E., & Loft, S. (2003). Linking exposure to environmental pollutants with biological effects. *Mutation Research/Reviews in Mutation Research*, *544*(2–3), 255–271. https://doi.org/10.1016/j.mrrev.2003.06.010

Souza, J. B., Reisen, V. A., Franco, G. C., Ispány, M., Bondon, P., & Santos, J. M. (2018). Generalized Additive Models with Principal Component Analysis: An Application to Time Series of Respiratory Disease and Air Pollution Data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *67*(2), 453–480. https://doi.org/10.1111/rssc.12239

Subramaniam, S., Raju, N., Ganesan, A., Rajavel, N., Chenniappan, M., Prakash, C., Pramanik, A., Basak, A. K., & Dixit, S. (2022). Artificial Intelligence Technologies for Forecasting Air Pollution and Human Health: A Narrative Review. *Sustainability*, *14*(16), 9951. https://doi.org/10.3390/su14169951

Szaniszló, J., & Ungváry, G. (2001). POLYCYCLIC AROMATIC HYDROCARBON EXPOSURE AND BURDEN OF OUTDOOR WORKERS IN BUDAPEST. *Journal of Toxicology and Environmental Health, Part A*, *62*(5), 297–306. https://doi.org/10.1080/152873901300017995

Truong, M. T., Nguyen, L. S. P., Hien, T. T., Pham, T. D. H., & Do, T. T. L. (2022). Source Apportionment and Risk Estimation of Heavy Metals in PM10 at a Southern Vietnam Megacity. *Aerosol and Air Quality Research*, *22*(8), 220094. https://doi.org/10.4209/aaqr.220094

Vadjić, V., Žužul, S., & Pehnec, G. (2010). Zinc Levels in Suspended Particulate Matter in Zagreb Air. *Bulletin of Environmental Contamination and Toxicology*, *85*(6), 628–631. https://doi.org/10.1007/s00128-010-0142-6

Vardar, N., Esen, F., & Tasdemir, Y. (2008). Seasonal concentrations and partitioning of PAHs in a suburban site of Bursa, Turkey. *Environmental Pollution*, *155*(2), 298–307. https://doi.org/10.1016/j.envpol.2007.11.026

Vega García, M., & Aznarte, J. L. (2020). Shapley additive explanations for NO2 forecasting. *Ecological Informatics*, *56*, 101039. https://doi.org/10.1016/j.ecoinf.2019.101039

Vilcassim, R., & Thurston, G. D. (2023). Gaps and future directions in research on health effects of air pollution. *eBioMedicine*, *93*, 104668. https://doi.org/10.1016/j.ebiom.2023.104668

Wan Mahiyuddin, W. R., Ismail, R., Mohammad Sham, N., Ahmad, N. I., & Nik Hassan, N. M. N. (2023). Cardiovascular and Respiratory Health Effects of Fine Particulate Matters (PM2.5): A Review on Time Series Studies. *Atmosphere*, *14*(5), 856. https://doi.org/10.3390/atmos14050856

Wang, X., Zong, Z., Tian, C., Chen, Y., Luo, C., Tang, J., Li, J., & Zhang, G. (2018). Assessing on toxic potency of PM2.5-bound polycyclic aromatic hydrocarbons at a national atmospheric background site in North China. *Science of The Total Environment*, *612*, 330–338. https://doi.org/10.1016/j.scitotenv.2017.08.208

Wedepohl, H. K. (1995). The composition of the continental crust. *Geochimica et Cosmochimica Acta*, *59*(7), 1217–1232. https://doi.org/10.1016/0016-7037(95)00038-2

WHO. (2021). *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. WHO European Centre for Environment and Health.

Wilcke, W. (2000). SYNOPSIS Polycyclic Aromatic Hydrocarbons (PAHs) in Soil—A Review. *Journal of Plant Nutrition and Soil Science*, *163*(3), 229–248. https://doi.org/10.1002/1522-2624(200006)163:3%3C229::AID-JPLN229%3E3.0.CO;2-6

Wilcke, W., Krauss, M., Safronov, G., Fokin, A. D., & Kaupenjohann, M. (2005). Polycyclic Aromatic Hydrocarbons (PAHs) in Soils of the Moscow Region—Concentrations, Temporal Trends, and Small-Scale Distribution. *Journal of Environmental Quality*, *34*(5), 1581–1590. https://doi.org/10.2134/jeq2005.0005

Yang, J., Fu, Q., Guo, X., Chu, B., Yao, Y., Teng, Y., & Wang, Y. (2015). Concentrations and Seasonal Variation of Ambient PM2.5 and Associated Metals at a Typical Residential Area in Beijing, China. *Bulletin of Environmental Contamination and Toxicology*, *94*(2), 232–239. https://doi.org/10.1007/s00128-014-1443-y

Zhang, L., Chen, C., & Murlis, J. (2001). Study on Winter Air Pollution Control in Lanzhou, China. *Water, Air, and Soil Pollution*, *127*(1–4), 351–372. https://doi.org/10.1023/A:1005257722202

Zhang, X., Qi, A., Wang, P., Huang, Q., Zhao, T., Yan, C., Yang, L., & Wang, W. (2023). Spatial Distribution, Sources, Air–Soil Exchange, and Health Risks of Parent PAHs and Derivative-Alkylated PAHs in Different Functional Areas of an Oilfield Area in the Yellow River Delta, North China. *Toxics*, *11*(6), 540. https://doi.org/10.3390/toxics11060540

Zhang, Y., & Tao, S. (2008). Seasonal variation of polycyclic aromatic hydrocarbons (PAHs) emissions in China. *Environmental Pollution*, *156*(3), 657–663. https://doi.org/10.1016/j.envpol.2008.06.017

Žibert, J., Cedilnik, J., & Pražnikar, J. (2016). Particulate matter (PM10) patterns in Europe: An exploratory data analysis using non-negative matrix factorization. *Atmospheric Environment*, *132*, 217–228. https://doi.org/10.1016/j.atmosenv.2016.03.005

5 APPENDIX

5.1 ORIGINAL SCIENTIFIC PUBLICATIONS

This thesis is based on the works described in the following articles.

- **Publication I**: Račić, N.; Ružičić, S.; Terzić, T.; Pehnec, G.; Jakovljević, I.; Sever Štrukil, Z.; Žužul, S.; Rinkovec, J.; Lovrić, M. (2024) Analyzing the relationship between gas consumption and airborne pollutants: case study of Zagreb, Croatia. Air Quality Atmosphere and Health, 18, 507-519. https://doi.org/10.1007/s11869-024-01655-7

- **Publication II**: Račić, N., Ružičić, S., Petrić, V., Terzić, T., Antunović, M., Škaro, I., Pehnec, G., Bešlić, I., Jakovljević, I., Sever Štrukil, Z., Rinkovec, J., Žužul, S., Lovrić, M. (2026) Assessment of contributors to airborne PAHs and heavy metals in PM10 using temporal, spatial, traffic and heating data in explainable machine learning models. Atmospheric Environment: X, 29, https://doi.org/10.1016/j.aeaoa.2026.100413

- **Publication III**: Račić, N.; Ružičić, S.; Pehnec, G.; Jakovljević, I.; Sever Štrukil, Z.; Rinkovec, J.; Žužul, S.; Smoljo, I.; Zgorelec, Ž.; Lovrić, M. (2025) Linking Atmospheric and Soil Contamination: A Comparative Study of PAHs and Metals in PM10 and Surface Soil near Urban Monitoring Stations. Toxics 2025, 13, 866. https://doi.org/10.3390/toxics13100866

5.2 Publication I

- *Račić, N.; Ružičić, S.; Terzić, T.; Pehnec, G.; Jakovljević, I.; Sever Štrukil, Z.; Žužul, S.; Rinkovec, J.; Lovrić, M. (2024) Analyzing the relationship between gas consumption and airborne pollutants: case study of Zagreb, Croatia. Air Quality Atmosphere and Health, 18, 507-519.* [*https://doi.org/10.1007/s11869-024-01655-7*](https://doi.org/10.1007/s11869-024-01655-7)

Air Quality, Atmosphere & Health
https://doi.org/10.1007/s11869-024-01655-7

Analyzing the relationship between gas consumption and airborne pollutants: case study of Zagreb, Croatia

Nikolina Račić[1] · Stanko Ružičić[2] · Teo Terzić[3] · Gordana Pehnec[1] · Ivana Jakovljević[1] · Zdravka Sever Štrukil[1] · Silva Žužul[1] · Jasmina Rinkovec[1] · Mario Lovrić[4,5]

Received: 10 June 2024 / Accepted: 4 October 2024

Abstract

Polycyclic aromatic hydrocarbons (PAHs) and metals in particulate matter significantly contribute to the health risks associated with air pollution. Hence, their measurements and source apportionment are relevant. This paper comprehensively analyzes the relationship between gas consumption and PAHs and metals in the PM_{10} fraction of particulate matter. The study investigates the potential associations using statistical techniques and quantifies the relationship between gas consumption patterns, meteorological conditions, and the measured concentrations of PAHs and metals in the atmosphere. The statistical methods comprise correlation analysis, Non-Negative Matrix Factorization (NMF), and linear regression. NMF analysis was employed to understand relationships among variables and potential sources of pollutants. NMF results revealed seasonal influences and different sources of pollutants in the studied area. PAHs with four aromatic rings have been grouped separately from 5- and 6-ring PAHs, suggesting two distinct sources of pollution – heating and traffic emissions. Metals such as As, Pb, Zn, and Cd are grouped, indicating mixed anthropogenic sources. The separation of Mn, Fe, and Cu in a distinguished group signifies their distinct origin, probably non-combustion traffic emissions (vehicle parts wearing).

Keywords Air pollution · Gas consumption · Linear regression · Metals · NMF · PAHs

Introduction

Air pollution is an enduring and pressing global concern with profound implications for human health. It can cause or contribute to increased mortality or severe illness or pose a present or potential hazard to human health (Kampa and Castanas 2008; Poulsen et al. 2023; Sørensen et al. 2022). Among the various pollutants that contribute to this issue, particulate matter (PM) is one of the primary and concerning components of health (Anderson et al. 2012; Goshua et al. 2022; Pritchett et al. 2022), consisting of solid particles or liquid droplets suspended in the air from many natural and human-induced sources (Karagulian et al. 2015). The composition of PM is diverse, as it can absorb and transport a variety of pollutants. However, its main components include metals, organic compounds, materials of biological origin, ions, reactive gases, etc. (Kampa and Castanas 2008; Pöschl 2005). It is a well-established fact that pollutant concentrations in the air of central and south-eastern Europe tend to increase during the winter season due to heightened emissions associated with heating sources, reaching levels that could pose a risk to human health (Cao et al. 2009; Cvetković et al. 2015; Huang and Wang 2014; Pehnec et al. 2020; Siudek and Frankowski 2018).This observation underscores the significance of investigating the correlation between gas consumption as a heating surrogate variable, meteorological parameters (mainly temperature), and pollutant levels (Akyüz and Çabuk 2009). Employing statistical techniques, including source apportionment, facilitates

Mario Lovrić
mario.lovric@inantro.hr

1 Institute for Medical Research and Occupational Health, Ksaverska cesta 2, Zagreb, Croatia

2 Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Pierottijeva 6, Zagreb, Croatia

3 Ascalia d.o.o, Trate 16, Čakovec 40000, Croatia

4 Centre for Bioanthropology, Institute for Anthropological Research, Gajeva 32, Zagreb HR-10000, Croatia

5 The Lisbon Council, Rue de la Loi 155, Brussels 1040, Belgium

Published online: 19 October 2024

exploring these relationships (Bi et al., 2007; Thurston et al., 2011; Zhang et al. 2020). Recognizing the important impact of household heating on air quality, this study specifically focuses on the relationship between gas consumption, as a marker of overall heating intensity, and the chemical composition of PM_{10}. While exact data on the number of homes using gas versus wood or other heating methods is unavailable, heating gas consumption serves as a practical indicator of overall heating demands. This approach assumes that gas consumption patterns closely align with periods of higher, overall energy consumption, which in turn reflects increased pollution levels. By analyzing heating gas consumption, we can better understand how seasonal heating activity, especially during colder months, influences PM_{10} levels. This is particularly relevant in areas like the IMROH monitoring station included in this study, where alternative heating sources, such as wood, are prevalent. Even without precise data on heating methods, focusing on heating gas consumption allows us to explore the broader impact of heating practices on air quality and the composition of PM_{10} in urban areas, providing valuable insights into the environmental and health implications of particulate matter pollution.

Within the context of airborne PM, a specific subgroup enriched with polycyclic aromatic hydrocarbons (PAHs) emerges as a focal point of scientific interest and environmental concern (Mallah et al. 2022; Rogula-Kozłowska et al. 2013; Singh et al. 2023). PAHs are organic compounds composed of multiple condensed aromatic rings, often originating from incomplete combustion processes (Krzyszczak and Czech 2021; Ren et al. 2017). These compounds have garnered considerable attention due to their potential for adverse health effects and environmental impact (Abdel-Shafy and Mansour 2016; Jakovljević et al. 2023; Kim et al. 2013; Mahasakpan et al. 2023). They are often encountered in various environmental matrices, including soil, water, sediment, and even food, reflecting their widespread distribution in nature (Barathi et al. 2023; X.-T. Wang et al. 2013). PAHs can find their way into the environment through many pathways, mainly arising from anthropogenic activities such as the combustion of fossil fuels, wood or organic materials, industrial processes, or agricultural activities. (Dat and Chang 2017; Liang et al. 2006; Singh et al. 2022). Natural sources, including wildfires and volcanic eruptions, also contribute to introducing PAHs into the environment (Stout et al. 2001). Understanding the complicated pathways through which PAHs enter and circulate within the environment is vital for comprehending their impact on ecosystems and public health. PAHs have been identified as potent carcinogens and mutagens, capable of initiating genetic mutations and promoting cancer development in individuals subjected to prolonged or high-level exposures (Dieme et al. 2012; Jakovljević et al. 2020; Kumar et al. 2016). Furthermore, PAHs are known to induce oxidative stress and inflammation in the human body, processes that are associated with a spectrum of health problems, including respiratory and cardiovascular diseases (Agudelo-Castañeda et al. 2017; Cachon et al. 2014; Landkocz et al. 2017; Leclercq et al. 2016; Shahriyari et al. 2022). Understanding the pathways through which PAHs can impact human health is paramount, as it underscores the urgency of mitigating PAH emissions and reducing exposure levels.

Metals within PM constitute a noteworthy concern, given their significant presence and potential health implications. These metals encompass a range of elements and compounds arising from various natural and anthropogenic sources. Their appearance in PM particles results from processes such as industrial emissions, combustion of fossil fuels, and the resuspension of dust and soil particles (Chen and Lippmann 2009; Wu et al. 2022). Exposure to PM-containing metals can profoundly and adversely affect human health. Some metals, including lead, cadmium, and nickel, are recognized neurotoxins capable of impairing cognitive development, particularly in children (L.-C. Chen et al. 2022). Additionally, certain metals, such as arsenic and chromium, possess carcinogenic properties and have been linked to an elevated risk of cancer among exposed populations (Pandey et al. 2013). These metals can also incite inflammation and oxidative stress upon inhalation, leading to respiratory symptoms and exacerbating conditions such as asthma and chronic obstructive pulmonary disease (COPD) (MacNee and Donaldson 2003; Zetlen et al. 2023). Furthermore, the translocation of metals from the lungs into the bloodstream raises concerns about potential cardiovascular effects, including an increased susceptibility to heart attacks and strokes (Guo et al. 2022). The mechanisms through which metals induce these health effects involve generating reactive oxygen species, disrupting enzyme function, and interfering with cellular processes (Valavanidis et al. 2008).

This paper discusses air pollution and public health protection by systematically investigating the following aims: understand the sources of polycyclic aromatic hydrocarbons (PAHs) and metals within PM_{10} fraction of particulate matter, quantify the relationships between PAHs and metals in PM_{10} and gas consumption patterns, analyze the impact of meteorological conditions on the concentrations of PAHs and metals in the atmosphere, and understand interactions among gas consumption patterns, meteorological conditions, and the presence of PAHs and metals in PM_{10} fraction of the particulate matter. To our knowledge, this study is the first that combines air quality data with meteorological and gas consumption data in the researched area. This investigation use an extensive dataset to understand the air pollution dynamics, with implications for environmental and public health outcomes, and therefore contribute to effective air quality management and public health protection by providing information about the sources and relationships of pollutants in the atmosphere.

Methods

Study area description

The study area for this research is the *Ksaverska cesta* measuring station (altitude 116 m a.s.l. 45°50’7” N, 15°58’42” E), located at the Institute for Medical Research and Occupational Health (IMROH), Zagreb, Croatia (Fig. 1). The measuring station represents an urban residential area for pollution assessment. It is one of the six air pollution monitoring stations funded by the City of Zagreb and one of the oldest. Over the years, the IMROH has collected data on PAHs and metals at this station, systematically and continuously recording valuable information about air pollution daily. The station and surrounding area was previously described by Jakovljević et al. (2020). The Institute also acquired gas consumption data, providing essential information regarding energy consumption patterns. Combining air quality data from the *Ksaverska cesta* measuring station and meteorological and gas consumption data forms the foundation of this comprehensive investigation into the interplay between energy consumption, meteorological conditions, and the concentrations of PAHs and metals in the atmosphere.

From January 1st, 2017, to December 31st, 2020, 24-hour samples of particulate matter fraction PM_{10} (PM with an equivalent aerodynamic diameter < 10 µm) were collected during the four years. Their content was determined, including metals arsenic (As), cadmium (Cd), lead (Pb), manganese (Mn), iron (Fe), copper (Cu), and zinc (Zn), following PAHs: benzo(a)pyrene (BaP), benzo(a)anthracene (BaA), benzo(ghi)perylene (BghiP), dibenzo(ah)anthracene (DahA), chrysene (Chry), benzo(k)fluoranthene (BkF), benzo(b)fluoranthene (BbF), benzo(j)fluoranthene (BjF), fluoranthene (Flu), and pyrene (Pyr). Simultaneously, nitrogen dioxide concentrations (NO_2) were also measured.

PM_{10} samples were collected on quartz filters with low-volume samplers (Sven Leckel, ~55 m^3/day). For the metals analysis, the entire quartz filter was used. For the PAH analysis, a punch was taken, and the concentration was recalculated to account for the entire sample using a conversion factor. PAHs were extracted from filters in an ultrasonic bath using a solvent mixture of cyclohexane and toluene, centrifugated, evaporated to dryness. and redissolved in acetonitrile. The analysis used an Agilent Infinity 1260

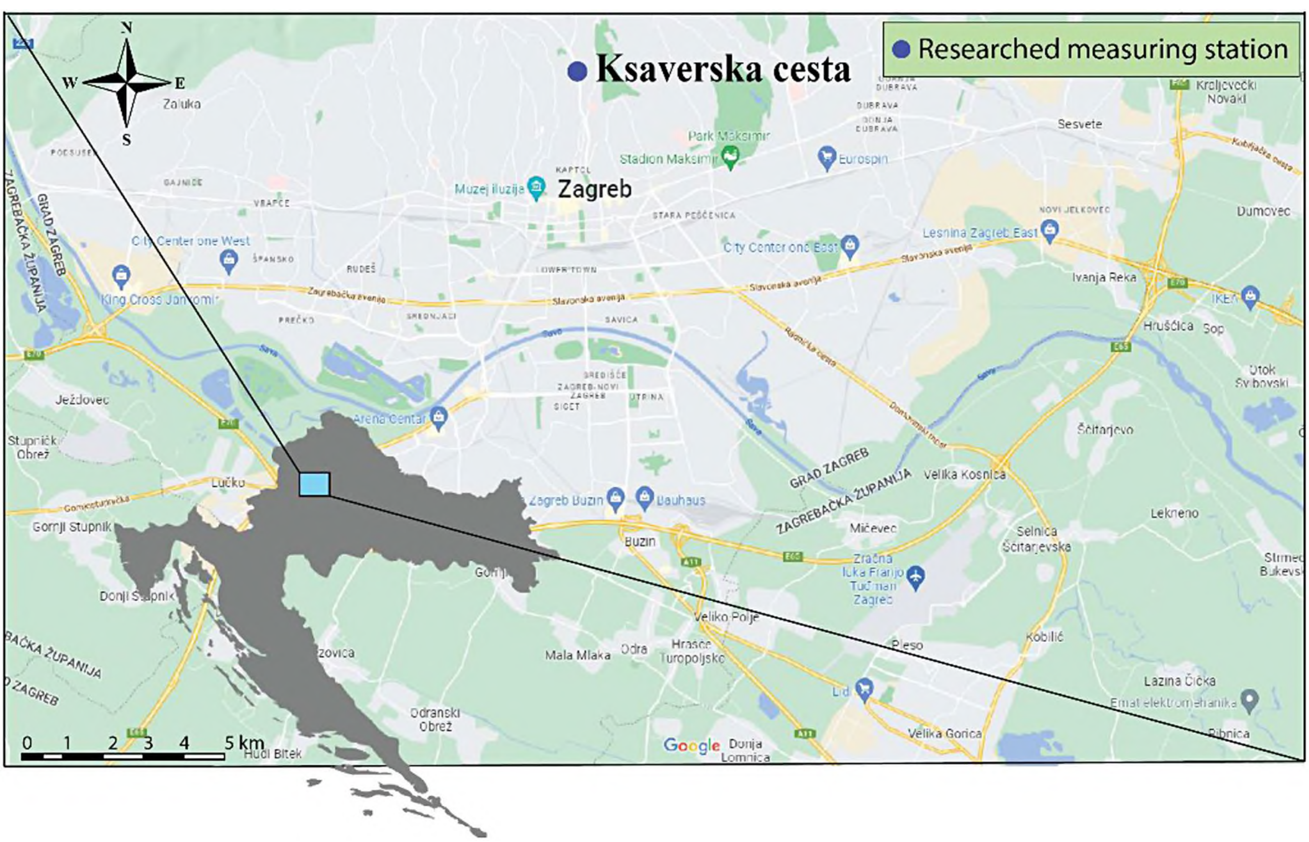


Fig. 1 Study area with air quality measuring station (*Ksaverska cesta*), altitude 116 m a.s.l., latitude 45°50’7” N, longitude 15°58’42” E

HPLC system with a fluorescence detector (Jakovljević et al. 2020). The method detection (DL) and quantification limits (QL) for HPLC were calculated as concentration equivalents to three and ten times the signal-to-noise ratio, respectively. The DL ranged from 0.0005 ng/m^3 for Chry to 0.03 ng/m^3 for BjF, while the DL for BaP, as a carcinogen PAHs for humans, was 0.001 ng/m^3. The QL ranged from 0.002 ng/m^3 for BaA to 0.1 ng/m^3 for BjF, while the QL for BaP, as a carcinogen PAHs for humans, was 0.003 ng/m^3. The detailed procedure was already described by Pehnec et al. (2016) and Jakovljević et al.(2020). For metal analysis, samples were microwave digested in nitric acid and analyzed using the inductively coupled plasma mass spectrometer (ICP-MS) Agilent Technologies, model 7500cx (Beslic et al. 2020). In addition to the comprehensive data on metals and PAH concentrations within the PM_{10} fraction, this study includes records of monthly gas consumption, average monthly temperature, pressure, precipitation, and wind speed over the same four-year period from 2017 to the end of 2020. The gas consumption data were obtained from the IMROH, which also operates the measurement station at Ksaverska cesta. This information is crucial for a better understanding of pollution concentration variations associated with heating seasonality and, more broadly, the elevated levels of pollutants during the winter period. The gas consumption data were used as a predictor variable, indicating increased emissions on specific winter days with higher heating demand. Since inhabitants around the measurement station use wood for heating, accurate data on this proportion are not available. In the area under research, the majority of vehicles typically operate on gasoline or diesel. This research aimed to analyze overall emission patterns and their impact on air quality, rather than the specific fuel types of individual vehicles.

The meteorological data were obtained from the Croatian Meteorological and Hydrological Service (DHMZ) from a reference measuring station located within the research area, and this data was also used to understand seasonal oscillations in individual pollutants better.

For subsequent statistical analysis, the data was prepared by calculating monthly means for each variable under consideration.

Statistical analysis

In this study, Python (version 3.10) with the libraries scikit-learn (Pedregosa et al. 2011), statsmodels (Seabold and Perktold 2010) and seaborn (Waskom et al. 2017), served as the primary tool for the statistical analysis. Monthly means of the measured analytes were taken to facilitate an analysis since the gas consumption data are in a monthly frequency. The resulting dataset comprised 48 data points, representing each month during the four-year measurement period. A correlation matrix was constructed and analyzed to explore the interrelationships between variables. The correlation matrix provided insights into which variables exhibited significant correlations. This initial examination served as a foundation for further exploratory statistical techniques. Non-Negative Matrix Factorization (NMF) was employed to dissect the dataset further, providing valuable information about potential source contributions and temporal variations. This technique decomposed the data into non-negative components, enabling the identification of factors influencing the concentrations of PAHs and metals over the studied period and location. NMF is a dimensionality reduction technique that factorizes a non-negative data matrix into two non-negative matrices, often called the basis matrix and the coefficient matrix (Boutsidis and Gallopoulos 2008; Marinetti et al. 2006). Mathematically, NMF can be represented as follows: given a non-negative data matrix X (with dimensions m x n), NMF aims to find two non-negative matrices W (basis matrix) and H (coefficient matrix) such that $X \approx WH$ (Pauca et al. 2006). Moreover, the analysis was extended to linear regression modelling to deepen the understanding of the factors influencing pollutant concentrations. This step allowed the exploration of relationships within specific pollutant groups identified through NMF. Each NMF component was used as a dependent variable in separate regression models, with average air temperature, pressure, wind speed, precipitation, NO_2, and gas consumption as independent variables (Deka et al. 2016). In the linear regression model, Principal Component Analysis (PCA) addressed the significant collinearity observed among average temperature, NO_2, and gas consumption, as determined by Spearman's correlation heatmap (Fig. 2). The application of PCA led to the derivation of principal components (PC1 and PC2), which were used in the regression model instead of the original collinear variables. Additionally, the model incorporated average monthly pressure (p_avg), precipitation, and wind speed as explanatory variables. This approach effectively reduces multicollinearity issues, improving the model's reliability and interpretability. Principal Component Analysis (PCA) was used as a dimensionality reduction technique, enabling a more transparent comprehension of how variables related to one another. PCA is a linear transformation technique that reduces the dimensionality of a dataset while preserving as much of its variance as possible (Bro and Smilde 2014). It identifies a set of orthogonal axes, known as principal components (PCs), which are linear combinations of the original variables. The mathematical representation of PCA provides a way to transform data into a new coordinate system that captures the most important variations in the data, making it easier to analyze while retaining much of the original information. Variables that displayed strong correlations or

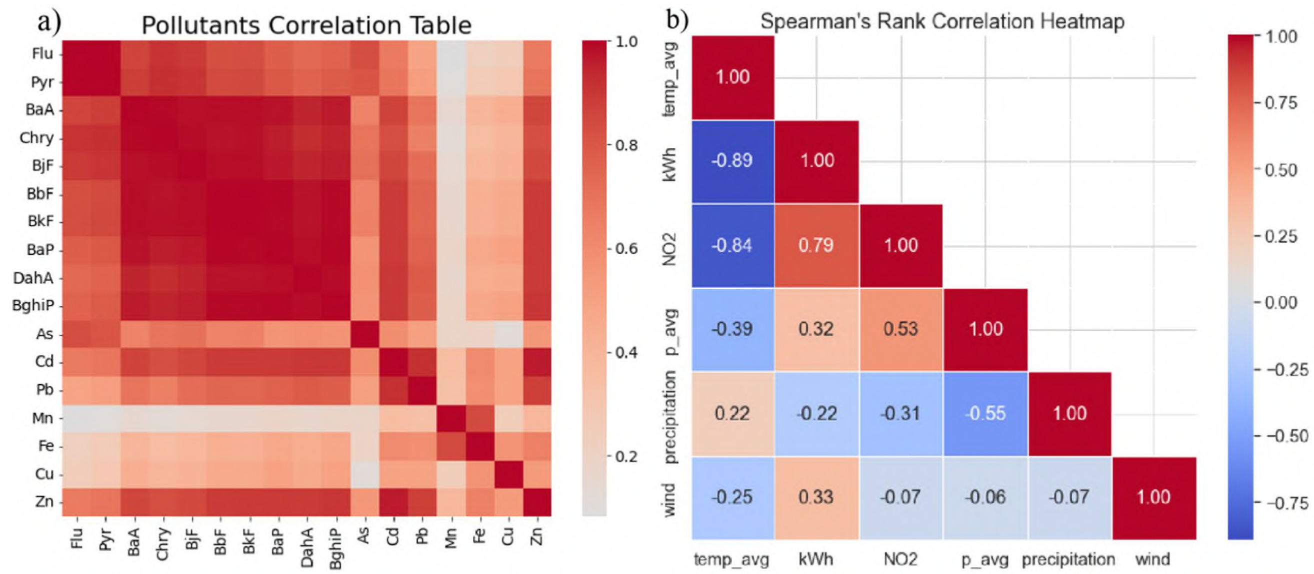


Fig. 2 Heatmap illustrating pairwise correlations among dataset variables, with color intensity representing the strength and direction of associations for (**a**) air pollutants and (**b**) independent variables (meteorological parameters: temp_avg – average temperature, p_avg – average pressure, precipitation and wind speed), gas consumption, and NO_2)

similar patterns of variation were found to cluster together in the PCA results (Pandey et al. 2014; Souza et al. 2018; Statheropoulos et al. 1998).

Results and discussion

Non-Negative Matrix Factorization (NMF)

The correlation matrix was calculated to summarize relationships among pollutant variables, helping to identify highly correlated pollutants and understand their behavior and common origins. Furthermore, heatmaps were made to visually represent the complex variable correlations within the dataset (Fig. 2). The intercorrelation of pollutants was determined based on Pearson's correlations (Fig. 2a), while Spearman's correlation was made between NO_2, meteorological parameters, and gas consumption (Fig. 2b). There are several instances of strong positive correlations, colored by darker red,, which may indicate common sources or similar environmental behaviors. Conversely, strong negative correlations, indicated by dark blue, imply an inverse relationship, hinting at distinct sources or divergent environmental processes for the variables. Figure 2a shows strong positive correlations among several pollutants, particularly between groups of PAHs such as BaA, Chry, BjF, BbF, BkF, DahA, BghiP, and BaP, indicating they might occur together. Zn also shows strong correlations within this group. These suggest that PAHs, along with Zn, likely share common emission sources, such as traffic and industrial activities. Zn in PM_{10} is commonly linked with industrial processes, tire wear, and brake wear emissions, contributing to their simultaneous presence in the atmosphere (Ravindra et al. 2008). Flu and Pyr show slightly less positive correlation compared to the other PAHs. This is likely due to their specific interactions with environmental factors such as salt particles from road salting and varying emission sources (Harrison et al. 1996; Ravindra et al. 2008). Additionally, a notable cluster of strong correlations among heavy metals such as As, Cd, Pb, and Mn suggests these pollutants are likely to be found together in the same environments. Common sources include industrial emissions, combustion of fossil fuels, and the use of contaminated materials in various industrial processes (Harrison and Yin 2000; Pope and Dockery 2006). Other elements, such as Fe and Cu, show moderate correlations with these clusters.

NMF was used to understand the potential sources and the mutual relationships deeper, revealing temporal variability within the dataset. All pollutant variables were used in NMF analysis: BaP, BaA, BghiP, DahA, Chry, BkF, BbF, BjF, Flu, Pyr, As, Cd, Pb, Mn, Fe, Cu, and Zn. Notably, despite most PAHs exhibiting a high degree of positive intercorrelation and suggesting similar behavior, all were included to ensure a comprehensive assessment of the pollution sources.

The following sections are the outcomes of NMF, where the original data matrix was transformed into two new matrices, one representing the different sources of pollution and the other the contribution of each basis component to the observed data at each time point. This has given vectors that we refer to as NMF1 and NMF2 (Fig. 3).

The time series graph (Fig. 3a) shows the temporal profiles of three NMF factors, inspired by (Šimić et al. 2020),

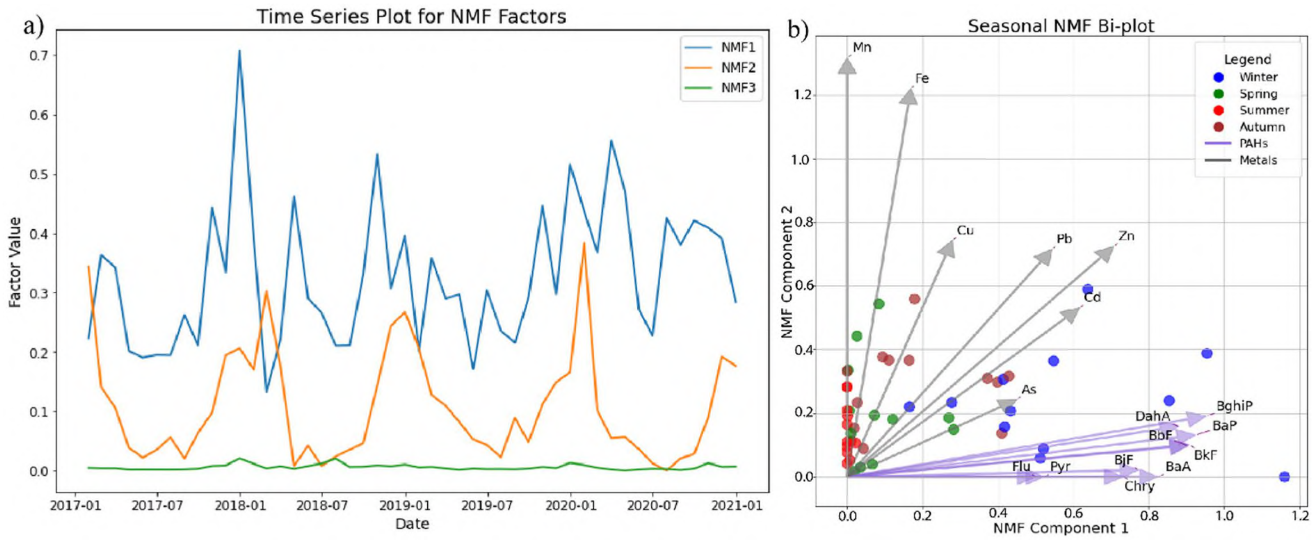


Fig. 3 (**a**) Time series graph for components NMF1, NMF2 and NMF3, (**b**) bi-plot for NMF components

from the beginning of 2017 to the end of 2020. Each factor will likely represent a different combination or source of pollutants captured over the monthly averaged data. NMF1 shows significant variability over the observed period, with several peaks suggesting periodic increases in this factor value. NMF2 shows a less pronounced set of peaks than NMF1 but is represented with clear, high seasonality. Although NMF1 shows higher peaks in the colder part of the year (winter, autumn) compared to NMF2, it also shows some additional, smaller peaks in other seasons, suggesting the influence of other factors. There is some overlap between NMF1 and NMF2, with NMF2 also showing a peak in January each year. NMF2, identified as motor vehicle emissions, typically results in relatively stable PAH emissions throughout the year. However, the observed January peak could be attributed to several factors. Increased vehicle activity during colder months, including higher instances of idling and extended engine warm-ups, may contribute to elevated emissions. Additionally, winter meteorological conditions, such as temperature inversions, can exacerbate pollutant concentrations by trapping them near the ground. It is also important to consider that traffic-related emissions encompass both combustion sources, such as gasoline, and non-combustion sources like tire wear, brake wear, and road dust resuspension, which can also significantly impact PM_{10} levels and exhibit seasonal variations. The third NMF factor is relatively stable and flat, with minimal variation over time. This indicates that the variables of sources represented by NMF3 are consistent, with no significant changes or seasonal effects.

The NMF bi-plot was used to represent the source information within our dataset visually (Fig. 3b). In this bi-plot, each variable is depicted as a vector, originating from the graph's origin. The angles between vectors convey the degree of association or dissociation between variables, illuminating the source dynamics and interrelationships. The length and direction of the vectors indicate the strength and nature of each pollutant's influence on the components, with longer vectors, such as those for Mn and Fe, suggesting a dominant influence on NMF2. This bi-plot identifies potential common sources for grouped pollutants and underscores the seasonal variance in pollutant levels. Variables that clustered closely in the bi-plot shared strong associations or exhibited analogous seasonal variations, while those with divergent orientations indicated relatively independent influences. NMF1 could include certain PAHs like Pyr, Flu, BjF, Chry, and BaA, which may indicate more widespread sources.

Metals in ambient air may originate from various sources, such as industrial emissions, combustion of fossil fuels, traffic emissions mining, or agriculture (Chen and Lippmann 2009; Soleimani et al. 2018). Mn, Fe, and Cu with long vectors pointing in a similar direction (Fig. 3b) could be indicative of traffic emissions not related to combustion processes (which are also the source of PAHs), but those related to the wear of tires, brakes or other parts of the vehicles, or resuspension of road dust (Fu et al. 2023; Sternbeck et al. 2002; Wang et al. 2021). Mn, Fe, and Cu vectors suggest a strong influence on the data structure captured by NMF2. Other metals (Pb, Zn, Cd, and As) formed a separated cluster with shorter vectors and positions between axes of the NMF1 and NMF2 components and could be indicative of traffic emissions as well as high-temperature combustion (Furusjö et al. 2007; Sternbeck et al. 2002).

Other pollutants that cluster closer to the origin with shorter vectors, such as certain PAHs (Pyr, Flu, BjF, Chry, BaA), may represent more diffuse or varied sources, such as residential heating (NMF1), which is also confirmed by the seasonality of NMF1. Especially Flu and Pyr are more related to the combustion of wood, grass, or coal (Jakovljević et al. 2015). Notably, coal has not been used in Croatia for decades, and heating in the study area mainly relies on gas or wood. However, neighboring countries (Bosnia and Herzegovina, Serbia, North Macedonia) still widely use coal (Pehnec et al. 2020), contributing to cross-border air pollution, heavy industry emissions, and possibly sources of metal pollutants. In the Western Balkans, biomass meets 42% of residential heating needs, leading to significant air pollution, particularly in winter, with PM_{10} levels exceeding limit values (European Community 2008) for 120–180 days annually (World Bank 2019). Despite the availability of natural gas in parts of Croatia, Bosnia and Herzegovina, and Serbia, the lower cost of biomass drives its use, contributing to higher particulate matter emissions, especially during colder months when outdated stoves and wet wood are commonly used (HEAL, 2022). This pattern is mirrored in other European countries like Greece and the Czech Republic, where economic factors have driven an increased reliance on biomass, resulting in higher $PM_{2.5}$ levels (Saffari et al. 2013). While we don't have exact percentages of the population in Zagreb regarding the specific heating sources used in households, it is evident that a significant number of households in areas near air quality monitoring stations still rely on wood for heating. According to the latest census, out of 300.329 households in Zagreb, 274.347 are connected to the gas grid for heating or cooking, indicating that a considerable portion of the population may be using alternatives like biomass or wood, particularly in suburban or rural parts of the city (City of Zagreb 2023). The presence of many houses that rely on wood for heating near the monitoring stations can contribute to higher particulate matter concentrations in those areas. These factors highlight the complexity of accurately quantifying the contribution of different heating sources to air pollution in Zagreb, further underscoring the need for more detailed data on energy use and its impact on air quality (Sopčić et al. 2024).

Other PAHs (DahA, BghiP, BaP, BbF, and BkF) have vector directions similar to those previously described PAHs. Still, pretty much the same length vectors can put them in a separate cluster with potentially a bit different pollution source, probably from vehicle exhaust emissions (combustion of gasoline in cars usually results in emissions of 5- and 6-ring PAHs) (Jakovljević et al. 2020). When comparing the findings of the present study with previous studies (Liu et al. 2021; Wang et al. 2018; Xu et al. 2006), there are parallels in the identification of pollution sources. While the current study suggests Mn, Fe, and Cu, as well as BghiP, DahA, BaP, BkF, and BbF, may be indicative of traffic emissions, Liu et al. (2021) also recognize a distinct factor associated with vehicle exhaust, marked by high molecular weight PAHs. Additionally, the clustering of NMF components in this study may resonate with Liu et al.'s factor related to coal combustion, attributed to moderate molecular weight PAHs. This underscores a commonality in source attribution across different analytical models and studies.

Pollutants that form a cluster, primarily composed of samples from the winter season, could indicate active or more prominent sources during colder weather, such as heating systems or reduced atmospheric dispersion (PAHs and partly As and Cd). Arsenic and cadmium are typical for high-temperature combustion emissions; compared with previous studies, these findings match source determination in other researched areas (Sternbeck et al. 2002).

Principal component analysis (PCA)

PCA was employed to address the significant collinearity observed among the predictors, i.e., to decorrelate pollutant concentration drivers such as average temperature (temp_avg), NO_2 and gas consumption (kWh), as determined by Spearman's correlation heatmap (Fig. 2b). It enabled us to use components that combine the information from various meteorological parameters, including temperature in the downstream linear regression model used to explain contributors to pollutant concentrations. This approach helps in understanding the overall impact of seasonal patterns and their relationship with pollutant levels without the complications of multicollinearity in further analysis. Sources of NO_2 are related to energy production and combustion of fossil fuels. In urban areas, NO_2 is usually a marker of traffic intensity (Šimić et al. 2020), and it is a typical product in vehicle exhausts (Carslaw 2005; Kendrick et al. 2015).

The decorrelation between meteorological data and NO_2 levels in this study is relevant because meteorological factors such as temperature, pressure, wind speed, and precipitation directly influence the dispersion, concentration, and chemical reactions of NO_2 in the atmosphere. For example, higher wind speeds can disperse pollutants, leading to lower NO_2 concentrations, while calm conditions can cause pollutants to accumulate. Additionally, seasonal variations in meteorological conditions, particularly during colder months, can affect NO_2 levels, as increased heating activities and specific weather patterns, like temperature inversions, can trap pollutants near the ground, resulting in higher concentrations of NO_2 (Voiculescu et al. 2020; Wang et al. 2022). Nitrogen dioxide is not only a significant air pollutant but also a primary indicator of traffic-related emissions, making it crucial for understanding pollution sources. In urban environments, NO_2 is predominantly emitted from vehicles, and its presence often correlates with other harmful pollutants, such as

PAHs. However, in this study, NO_2, along with meteorological parameters, was selected as a predictive variable. The analysis shows that average temperature ($p = 4.91 \times 10^{-14}$), energy consumption ($p = 2.68 \times 10^{-11}$), average pressure ($p = 9.08 \times 10^{-5}$), and precipitation ($p = 0.035$) have statistically significant correlations with NO_2 levels, indicating these factors significantly influence NO_2 concentrations. In contrast, wind speed ($p = 0.629$) does not show a significant correlation, suggesting it may not strongly impact NO_2 levels in this dataset. Using PCA, we created two latent vectors, the principal components (PC1 and PC2), used in the regression model instead of the original collinear variables since linear regression assumes independence among the predictors.

A new heatmap was generated for further analysis to assess whether incorporating PC1 and PC2 effectively reduced multicollinearity among the variables (Fig. 4a). In the subsequent linear regression analysis, the variables PC1 and PC2 were used, along with the average pressure, the wind speed, and the precipitation data. PC1 and PC2 captured 96.23% of the variance (90.03% from PC1 and 6.20% from PC2). This suggests that most of the variability in the data is represented by these two components. The time series plot for PCA components (Fig. 4b) shows that PC1 reflects a seasonal cycle, where emissions driven by energy consumption typically rise when it gets colder. PC2 seems to capture specific conditions where NO_2 increases without a corresponding rise in energy use, suggesting influences beyond just temperature (possible traffic). Loadings are calculated to see how much each original variable contributes to the principal components. PC1 is heavily influenced by NO_2 (0.570), kWh (0.577), and temp_avg (-0.585), indicating that higher NO_2 and kWh values are associated with lower temperatures. PC2 is primarily influenced by NO_2 (0.786) and kWh (-0.591), with temp_avg (0.183) playing a minor role, suggesting that NO_2 increases independently of energy use, possibly due to traffic emissions. This interpretation aligns with the observed variance, where PC1 captures the general seasonal trend, and PC2 reflects conditions that specifically increase NO_2 levels without a corresponding rise in energy use. However, Fig. 4a shows that PC2 strongly negatively correlates with wind speed and positively with pressure. This suggests that during winter, high-pressure systems (anticyclones) cause temperature inversions, trapping pollutants near the ground. Low wind speeds further exacerbate the situation, leading to higher concentrations of all pollutants in the air, indicating that such meteorological conditions could also cause pollution peaks. PC3 may be capturing noise or complex behaviors in the data that are not associated with clear trends or seasonal effects. This is the reason why it is not used in further analysis.

Linear regression

Linear regression was used in data analyses to quantify the relationship between pollutants, and the decorrelated contributors, i.e. PCA-transformed meteorological parameters, and gas consumption as independent predictors. Gas consumption (kWh), average temperature (temp_avg), and NO_2 were hence not used in the models directly because of the strong correlation (Fig. 2). Regression analysis aimed to capture any observable trends or patterns in a multivariate space and consequently explain sources. Table 1 provides multiple linear regression models for three NMF factors (target variables), which were used as a dependent variable in separate regression models, with PC1, PC2, p_avg, precipitation, and wind as independent variables.

The regression analysis reveals relationships between the latent environmental factors and each of the NMF components (NMF1, NMF2, and NMF3). For NMF1 (latent

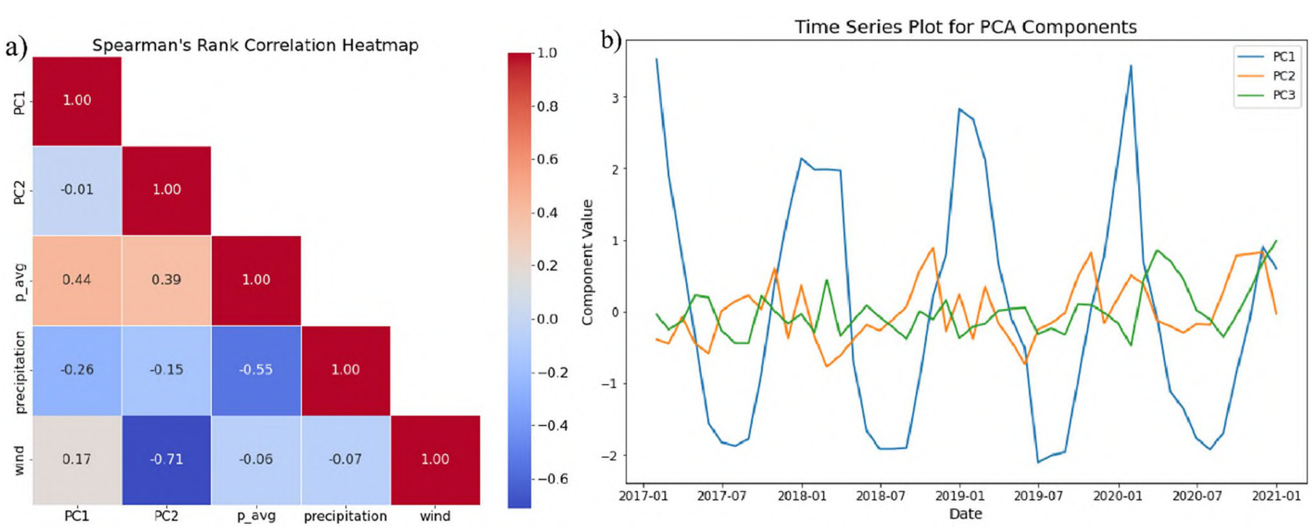


Fig. 4 (**a**) Spearman's rank correlation heatmap for PC1 and PC2 among other variables, (**b**) time series graph for first three PCA components

Table 1 Results of the multiple linear regression model for NMF1, NMF2 and NMF3 as a dependent variables and PC1, PC2, p_avg, wind and precipitation as a predictor

	Endpoint $/to$	NMF1	NMF2	NMF3
	Model Parameter			
Model Statistics	R-squared	0.362	0.817	0.274
	F-statistic	4.771	37.54	3.165
	Prob (F-stat)	0.002	$1.89\ 10^{-14}$	0.016
	No. observations	48	48	48
Coefficients	Constant	0.469	-4.588	0.236
	PC1	0.019	0.048	0.001
	PC2	0.164	-0.033	0.002
	p_avg	-0.001	0.005	-0.0002
	wind	0.036	-0.039	0.0003
	precipitation	-0.004	0.004	-0.003

representation of Pyr, Flu, BjF, Chry, BaA), the model explains 36.2% of the variability (R-squared = 0.362), indicating a relatively good fit. The adjusted R-squared is lower at 0.286, reflecting the model's predictive power after adjusting for the number of predictors. PC2 shows a positive and significant relationship with the dependent variable. However, the other variables, including PC1, do not significantly predict NMF1.

In contrast, the model for NMF2 (latent representation of Mn, Fe, Cu) demonstrates a robust fit, explaining 81.7% of the variability (R-squared = 0.817). The F-statistic is significant, suggesting the model is a good fit for the data. PC1 is a significant predictor with a positive coefficient, while PC2, precipitation, and wind are not statistically significant at the 0.05 level. This could be because the metals in NMF2 are likely from the mechanical wear of vehicles. The high seasonality of PC1 could be linked to the seasonality of traffic density, with higher traffic density in winter. This interpretation aligns with findings by Rinkovec et al. (Rinkovec et al. 2018), where similar seasonal variations of platinum, palladium, and rhodium (originating from automotive catalytic converters) were observed at three locations in Zagreb. Therefore, the significant influence of PC1 on NMF2 may reflect the increased vehicle-related emissions during winter months. In the case of NMF3, the model exhibits a limited explanatory power similar to NMF1 (R-squared = 0.274). PC1 shows a significant positive impact, but other variables, including PC2, are not statistically significant.

The results indicate that while NMF2 is strongly influenced by the seasonal changes represented by PC1, which captures temperature and gas consumption, the variability in NMF1 and NMF3 is less clearly explained by the variables included in the analysis. NMF2, representing metals likely from vehicle wear, aligns with the high traffic density in winter, as reflected in PC1. NMF1 shows a positive relationship with PC2, which can be explained by traffic emissions, as NO_2 is a major component of PC2. However, this connection might be more complex, as NMF1 represents PAHs typically from biomass burning, suggesting that other factors might also be significant. NMF3 shows a relationship to PC1 similar to that of NMF2, indicating that seasonal variations in temperature and gas consumption influence both NMF2 and NMF3, although NMF3's specific sources remain less clear. This suggests that while NMF1 and NMF3 share a similar relationship with PC1, the factors influencing NMF1 are more varied and complex.

The integration of PCA and NMF methodologies combined with linear regression has facilitated a better understanding of pollution sources. Metals such as Mn, Fe, and Cu, identified through NMF2 with longer vectors, seem to connected to traffic emissions but probably not in the context of fuel combustion. Instead, these metals probably originate from other sources such as engine wear, material wear, brake pads, and tire wear (Jiang et al. 2019; Tsai et al. 2007). In contrast, Pb, Zn, Cd, and As, characterized by shorter vectors, suggest a mixed profile of sources, including traffic and heating sources (Argyropoulos et al. 2013; Viana et al. 2006; Zhang et al. 2020). PAHs present a diverse picture: while compounds like Pyr and Flu cluster near the origin, indicating sources such as residential heating, others such as BghiP and BaP are consistent with vehicular emissions, as denoted by their vector lengths and directions. This analysis helps in understanding the interactions of pollutants, revealing source contributions of pollution in the urban environment.

Conclusions

This study, located around the Ksaverska cesta air quality measuring station in Zagreb, provides valuable insights into the relationship between gas consumption, meteorological conditions, and the concentrations of PAHs and metals in the atmosphere. This research combines air quality data with meteorological and gas consumption data, offering a better understanding of the factors influencing pollution dynamics in the specified study region by integrating NMF with regression models. The integration of NMF and regression models provided a conclusion about the behavior of gas consumption and meteorological factors in interaction with air pollutants. Notably, it identified significant seasonal influences, as well as heating and traffic emissions as dominant sources of PAHs and metals in the researched area. The analysis showed that PAHs such as BaP, BghiP, and DahA were strongly associated with traffic emissions, highlighting their presence in vehicle exhaust and their correlation with other high molecular weight compounds. Flu and Pyr, although less correlated, were linked to varied emission sources (probably residential heating). Metals like Mn, Fe, and Cu were probably predominantly associated with vehicle wear and non-combustion sources, indicating their presence from tire wear, brake wear, and road dust resuspension. Pb, Zn, Cd, and As assumed to be associated with both traffic emissions and heating sources, reflecting a mix of combustion processes and industrial activities. The study also revealed significant seasonal variations, with higher pollutant concentrations during colder months due to increased heating activities and traffic density.

In conclusion, this study not only contributes to understanding the dynamics influencing air quality in the Zagreb area but also emphasizes the need for a multidimensional approach to capture the diverse factors affecting pollutant concentrations. The integration of air quality, meteorological, and gas consumption data offers a robust foundation for further research and policy-making, with potential implications for environmental and public health outcomes. Future work may include fine-tuning these analyses with additional variables and measuring stations or alternate statistical methods to improve understanding of this complex ecological system further.

Author contributions Study conception and design: Nikolina Račić, Mario Lovrić, Stanko Ružičić. Data analysis: Nikolina Račić, Mario Lovrić, Teo Terzić. Laboratory analysis were performed by Ivana Jakovljević, Zdravka Sever Štrukil, Jamina Rinkovec and Silva Žužul. The first draft of the manuscript was written by Nikolina Račić. Manuscript revision: Gordana Pehnec, Mario Lovrić, Stanko Ružičić. All authors approved the submitted version.

Funding This study was performed using the facilities and equipment funded within the European Regional Development Fund project KK.01.1.1.02.0007 "Research and Education Centre of Environmental Health and Radiation Protection – Reconstruction and Expansion of the Institute for Medical Research and Occupational Health", and funded by the European Union – NextGenerationEU (Program Contract of 8 December 2023, Class: 643-02/23 – 01/00016, Reg. no. 533-03-23-0006)-EnvironPollutHealth. ML is supported by Horizon Europe (EDIAQI project #101057497) and European Union—NextGenerationEU, grant number IA-INT-2024-BioAntroPoP granted to the Institute for Anthropological Research, Zagreb, Croatia.

Data availability Supporting data may be obtained from the website: http://iszz.azo.hr/iskzl/, either through the query on the website or by the written request.

Declarations

Ethics approval and consent to participate Since this study did not collect personal information, no ethical approval was required.

Competing interests The authors declare that they have no competing interests.

References

Abdel-Shafy HI, Mansour MSM (2016) A review on polycyclic aromatic hydrocarbons: source, environmental impact, effect on human health and remediation. Egyptian J Petroleum 25(1):107–123. https://doi.org/10.1016/j.ejpe.2015.03.011

Agudelo-Castañeda DM, Teixeira EC, Schneider IL, Lara SR, Silva LFO (2017) Exposure to polycyclic aromatic hydrocarbons in atmospheric PM1.0 of urban environments: carcinogenic and mutagenic respiratory health risk by age groups. Environ Pollut 224:158–170. https://doi.org/10.1016/j.envpol.2017.01.075

Akyüz M, Çabuk H (2009) Meteorological variations of PM2.5/PM10 concentrations and particle-associated polycyclic aromatic hydrocarbons in the atmospheric environment of Zonguldak, Turkey. J Hazard Mater 170(1):13–21. https://doi.org/10.1016/j.jhazmat.2009.05.029

Anderson JO, Thundiyil JG, Stolbach A (2012) Clearing the Air: a review of the effects of Particulate Matter Air Pollution on Human Health. J Med Toxicol 8(2):166–175. https://doi.org/10.1007/s13181-011-0203-1

Argyropoulos G, Grigoratos T, Voutsinas M, Samara C (2013) Concentrations and source apportionment of PM10 and associated elemental and ionic species in a lignite-burning power generation area of southern Greece. Environ Sci Pollut Res 20(10):7214–7230. https://doi.org/10.1007/s11356-013-1721-y

Barathi S, Rathinasamy JG, Sabapathi G, Aruljothi N, Lee KN, J., Kandasamy S (2023) Recent trends in polycyclic aromatic hydrocarbons pollution distribution and counteracting bio-remediation strategies. Chemosphere 337:139396. https://doi.org/10.1016/j.chemosphere.2023.139396

Beslic I, Burger J, Cadoni F, Centioli D, Kranjc I, Van Den Bril B, Rinkovec J, Sega K, Zang T, Zuzul S, Gladtke D (2020) Determination of as, cd, Ni and Pb in PM10 – comparison of different sample work-up and analysis methods/Bestimmung Von as, cd, ni und pb in PM10 – vergleich verschiedener Probenaufbereitungs-Und Analysenverfahren. Gefahrstoffe 80(06):227–233. https://doi.org/10.37544/0949-8036-2020-06-17

Bi X, Feng Y, Wu J, Wang Y, Zhu T (2007) Source apportionment of PM10 in six cities of northern China. Atmos Environ 41(5):903–912. https://doi.org/10.1016/j.atmosenv.2006.09.033

Boutsidis C, Gallopoulos E (2008) SVD based initialization: a head start for nonnegative matrix factorization. Pattern Recogn 41(4):1350–1362. https://doi.org/10.1016/j.patcog.2007.09.010

Bro R, Smilde AK (2014) Principal component analysis. Anal Methods 6(9):2812–2831. https://doi.org/10.1039/C3AY41907J

Cachon BF, Firmin S, Verdin A, Ayi-Fanou L, Billet S, Cazier F, Martin PJ, Aissi F, Courcot D, Sanni A, Shirali P (2014) Proinflammatory effects and oxidative stress within human bronchial epithelial cells exposed to atmospheric particulate matter (PM2.5 and PM > 2.5) collected from Cotonou, Benin. Environ Pollut 185:340–351. https://doi.org/10.1016/j.envpol.2013.10.026

Cao J, Shen Z, Chow JC, Qi G, Watson JG (2009) Seasonal variations and sources of mass and chemical composition for PM10 aerosol in Hangzhou, China. Particuology 7(3):161–168. https://doi.org/10.1016/j.partic.2009.01.009

Carslaw D (2005) Evidence of an increasing NO/NO emissions ratio from road traffic emissions. Atmos Environ 39(26):4793–4802. https://doi.org/10.1016/j.atmosenv.2005.06.023

Chen LC, Lippmann M (2009) Effects of metals within Ambient Air Particulate Matter (PM) on Human Health. Inhalation Toxicol 21(1):1–31. https://doi.org/10.1080/08958370802105405

Chen L-C, Maciejczyk P, Thurston D (2022) G. Metals and air pollution. In *Handbook on the Toxicology of Metals* (pp. 137–182). Elsevier. https://doi.org/10.1016/B978-0-12-823292-7.00004-8

City of Zagreb (2023) *Statistički ljetopis grada Zagreba*

Cvetković A, Jovašević-Stojanović M, Marković D, Ristovski Z (2015) Concentration and source identification of polycyclic aromatic hydrocarbons in the metropolitan area of Belgrade, Serbia. Atmos Environ 112:335–343. https://doi.org/10.1016/j.atmosenv.2015.04.034

Dat N-D, Chang MB (2017) Review on characteristics of PAHs in atmosphere, anthropogenic sources and control technologies. Sci Total Environ 609:682–693. https://doi.org/10.1016/j.scitotenv.2017.07.204

Deka J, Sarma KP, Hoque RR (2016) Source contributions of Polycyclic Aromatic Hydrocarbons in soils around oilfield in the Brahmaputra Valley. Ecotoxicol Environ Saf 133:281–289. https://doi.org/10.1016/j.ecoenv.2016.07.031

Dieme D, Cabral-Ndior M, Garçon G, Verdin A, Billet S, Cazier F, Courcot D, Diouf A, Shirali P (2012) Relationship between physicochemical characterization and toxicity of fine particulate matter (PM2.5) collected in Dakar city (Senegal). Environ Res 113:1–13. https://doi.org/10.1016/j.envres.2011.11.009

European Community (2008) Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe (L152/1; European Community). Official Journal of the European Union 2008

Fu Z, Wu Y, Zhao S, Bai X, Liu S, Zhao H, Hao Y, Tian H (2023) Emissions of multiple metals from vehicular brake linings wear in China, 1980–2020. Sci Total Environ 889:164380. https://doi.org/10.1016/j.scitotenv.2023.164380

Furusjö E, Sternbeck J, Cousins AP (2007) PM10 source characterization at urban and highway roadside locations. Sci Total Environ 387(1–3):206–219. https://doi.org/10.1016/j.scitotenv.2007.07.021

Goshua A, Akdis CA, Nadeau KC (2022) World Health Organization global air quality guideline recommendations. Exec Summary Allergy 77(7):1955–1960. https://doi.org/10.1111/all.15224

Guo L-C, Lv Z, Ma W, Xiao J, Lin H, He G, Li X, Zeng W, Hu J, Zhou Y, Li M, Yu S, Xu Y, Zhang J, Zhang H, Liu T (2022) Contribution of heavy metals in PM2.5 to cardiovascular disease mortality risk, a case study in Guangzhou, China. Chemosphere 297:134102. https://doi.org/10.1016/j.chemosphere.2022.134102

Harrison RM, Yin J (2000) Particulate matter in the atmosphere: which particle properties are important for its effects on health? Sci Total Environ 249(1–3):85–101. https://doi.org/10.1016/S0048-9697(99)00513-6

Harrison RM, Smith DJT, Luhana L (1996) Source Apportionment of Atmospheric Polycyclic Aromatic Hydrocarbons Collected from an Urban Location in Birmingham, U.K. *Environmental Science & Technology*, *30*(3), 825–832. https://doi.org/10.1021/es950252d

HEAL (Health and Environment Alliance) (2022) The Health Perspective of Wood Burning in the Western Balkans Region

Huang L, Wang G (2014) Chemical characteristics and source apportionment of atmospheric particles during heating period in Harbin, China. J Environ Sci 26(12):2475–2483. https://doi.org/10.1016/j.jes.2014.04.008

Jakovljević I, Pehnec G, Vadjić V, Šišović A, Davila S, Bešlić I (2015) Carcinogenic activity of polycyclic aromatic hydrocarbons bounded on particle fraction. Environ Sci Pollut Res 22(20):15931–15940. https://doi.org/10.1007/s11356-015-4777-z

Jakovljević I, Sever Štrukil Z, Godec R, Bešlić I, Davila S, Lovrić M, Pehnec G (2020) Pollution sources and carcinogenic risk of PAHs in PM1 particle fraction in an urban area. Int J Environ Res Public Health 17(24):9587. https://doi.org/10.3390/ijerph17249587

Jakovljević I, Smoljo I, Sever Štrukil Z, Pehnec G (2023) Carcinogenic activity and Risk Assessment of PAHs in Ambient Air: PM10 particle fraction and bulk deposition. Toxics 11(3):228. https://doi.org/10.3390/toxics11030228

Jiang N, Liu X, Wang S, Yu X, Yin S, Duan S, Wang S, Zhang R, Li S (2019) Pollution characterization, source identification, and Health Risks of Atmospheric-Particle-Bound Heavy Metals in PM10 and PM2.5 at multiple sites in an emerging megacity in the Central Region of China. Aerosol Air Qual Res 19(2):247–271. https://doi.org/10.4209/aaqr.2018.07.0275

Kampa M, Castanas E (2008) Human health effects of air pollution. Environ Pollut 151(2):362–367. https://doi.org/10.1016/j.envpol.2007.06.012

Karagulian F, Belis CA, Dora CFC, Prüss-Ustün AM, Bonjour S, Adair-Rohani H, Amann M (2015) Contributions to cities' ambient particulate matter (PM): a systematic review of local source contributions at global level. Atmos Environ 120:475–483. https://doi.org/10.1016/j.atmosenv.2015.08.087

Kendrick CM, Koonce P, George LA (2015) Diurnal and seasonal variations of NO, 2 and PM 2.5 mass as a function of traffic volumes alongside an urban arterial. Atmos Environ 122:133–141. https://doi.org/10.1016/j.atmosenv.2015.09.019

Kim K-H, Jahan SA, Kabir E, Brown RJC (2013) A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. Environ Int 60:71–80. https://doi.org/10.1016/j.envint.2013.07.019

Krzyszczak A, Czech B (2021) Occurrence and toxicity of polycyclic aromatic hydrocarbons derivatives in environmental matrices. Sci Total Environ 788:147738. https://doi.org/10.1016/j.scitotenv.2021.147738

Kumar V, Kothiyal NC, Saruchi, Vikas P, Sharma R (2016) Sources, distribution, and health effect of carcinogenic polycyclic aromatic hydrocarbons (PAHs) – current knowledge and future directions. J Chin Adv Mater Soc 4(4):302–321. https://doi.org/10.1080/22243682.2016.1230475

Landkocz Y, Ledoux F, André V, Cazier F, Genevray P, Dewaele D, Martin PJ, Lepers C, Verdin A, Courcot L, Boushina S, Sichel F, Gualtieri M, Shirali P, Courcot D, Billet S (2017) Fine and ultrafine atmospheric particulate matter at a multi-influenced urban site: physicochemical characterization, mutagenicity and cytotoxicity. Environ Pollut 221:130–140. https://doi.org/10.1016/j.envpol.2016.11.054

Leclercq B, Happillon M, Antherieu S, Hardy EM, Alleman LY, Grova N, Perdrix E, Appenzeller BM, Guidice L, Coddeville J-M, P., Garçon G (2016) Differential responses of healthy and chronic obstructive pulmonary diseased human bronchial epithelial cells repeatedly exposed to air pollution-derived PM4. Environ Pollut 218:1074–1088. https://doi.org/10.1016/j.envpol.2016.08.059

Liang F, Lu M, Birch ME, Keener TC, Liu Z (2006) Determination of polycyclic aromatic sulfur heterocycles in diesel particulate matter and diesel fuel by gas chromatography with atomic emission detection. J Chromatogr A 1114(1):145–153. https://doi.org/10.1016/j.chroma.2006.02.096

Liu H, Li B, Qi H, Ma L, Xu J, Wang M, Ma W, Tian C (2021) Source apportionment and Toxic Potency of Polycyclic Aromatic Hydrocarbons (PAHs) in the air of Harbin, a Cold City in Northern China. Atmosphere 12(3):297. https://doi.org/10.3390/atmos12030297

MacNee W, Donaldson K (2003) Mechanism of lung injury caused by PM10 and ultrafine particles with special reference to COPD. Eur Respir J 21(Supplement 40):47S–51s. https://doi.org/10.1183/09031936.03.00403203

Mahasakpan N, Chaisongkaew P, Inerb M, Nim N, Phairuang W, Tekasakul S, Furuuchi M, Hata M, Kaosol T, Tekasakul P, Dejchanchaiwong R (2023) Fine and ultrafine particle- and gas-polycyclic aromatic hydrocarbons affecting southern Thailand air quality during transboundary haze and potential health effects. J Environ Sci 124:253–267. https://doi.org/10.1016/j.jes.2021.11.005

Mallah MA, Changxing L, Mallah MA, Noreen S, Liu Y, Saeed M, Xi H, Ahmed B, Feng F, Mirjat AA, Wang W, Jabar A, Naveed

M, Li J-H, Zhang Q (2022) Polycyclic aromatic hydrocarbon and its effects on human health: an overeview. Chemosphere 296:133948. https://doi.org/10.1016/j.chemosphere.2022.133948
Marinetti S, Finesso L, Marsilio E (2006) Matrix factorization methods: application to thermal NDT/E. NDT E International 39(8):611–616. https://doi.org/10.1016/j.ndteint.2006.04.008
Pandey P, Patel DK, Khan AH, Barman SC, Murthy RC, Kisku GC (2013) Temporal distribution of fine particulates (PM $_{2.5,}$ PM $_{10}$), potentially toxic metals, PAHs and Metal-bound carcinogenic risk in the population of Lucknow City, India. J Environ Sci Health Part A 48(7):730–745. https://doi.org/10.1080/10934529.2013.744613
Pandey B, Agrawal M, Singh S (2014) Assessment of air pollution around coal mining area: emphasizing on spatial distributions, seasonal variations and heavy metals, using cluster and principal component analysis. Atmospheric Pollution Res 5(1):79–86. https://doi.org/10.5094/APR.2014.010
Pauca VP, Piper J, Plemmons RJ (2006) Nonnegative matrix factorization for spectral data analysis. Linear Algebra Appl 416(1):29–47. https://doi.org/10.1016/j.laa.2005.06.025
Pedregosa F, Michel V, Grisel O, Blondel M, Prettenhofer P, Weiss R, Vanderplas J, Cournapeau D, Varoquaux G, Gramfort A, Thirion B, Dubourg V, Passos A, Brucher M, Duchesnay É, Thirion B, Prettenhofer P (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830. https://doi.org/10.1007/s13398-014-0173-7.2
Pehnec G, Jakovljević I, Šišović A, Bešlić I, Vađić V (2016) Influence of ozone and meteorological parameters on levels of polycyclic aromatic hydrocarbons in the air. Atmos Environ 131:263–268. https://doi.org/10.1016/j.atmosenv.2016.02.009
Pehnec G, Jakovljević I, Godec R, Sever Štrukil Z, Žero S, Huremović J, Džepina K (2020) Carcinogenic organic content of particulate matter at urban locations with different pollution sources. Sci Total Environ 734:139414. https://doi.org/10.1016/j.scitotenv.2020.139414
Pope CA, Dockery DW (2006) Health effects of Fine Particulate Air Pollution: lines that connect. Journal Air Waste Manage Association 56(6):709–742. https://doi.org/10.1080/10473289.2006.10464485
Pöschl U (2005) Atmospheric aerosols: Composition, Transformation, Climate and Health effects. Angew Chem Int Ed 44(46):7520–7540. https://doi.org/10.1002/anie.200501122
Poulsen AH, Sørensen M, Hvidtfeldt UA, Christensen JH, Brandt J, Frohn LM, Ketzel M, Andersen C, Raaschou-Nielsen O (2023) Source-specific air Pollution including Ultrafine particles and risk of myocardial infarction: a Nationwide Cohort Study from Denmark. Environ Health Perspect 131(5):057010. https://doi.org/10.1289/EHP10556
Pritchett N, Spangler EC, Gray GM, Livinski AA, Sampson JN, Dawsey SM, Jones RR (2022) Exposure to Outdoor Particulate Matter Air Pollution and Risk of gastrointestinal cancers in adults: a systematic review and Meta-analysis of epidemiologic evidence. Environ Health Perspect 130(3):036001. https://doi.org/10.1289/EHP9620
Ravindra K, Sokhi R, Vangrieken R (2008) Atmospheric polycyclic aromatic hydrocarbons: source attribution, emission factors and regulation. Atmos Environ 42(13):2895–2921. https://doi.org/10.1016/j.atmosenv.2007.12.010
Ren Y, Zhou B, Tao J, Cao J, Zhang Z, Wu C, Wang J, Li J, Zhang L, Han Y, Liu L, Cao C, Wang G (2017) Composition and size distribution of airborne particulate PAHs and oxygenated PAHs in two Chinese megacities. Atmos Res 183:322–330. https://doi.org/10.1016/j.atmosres.2016.09.015
Rinkovec J, Pehnec G, Godec R, Davila S, Bešlić I (2018) Spatial and temporal distribution of platinum, palladium and rhodium in Zagreb air. Sci Total Environ 636:456–463. https://doi.org/10.1016/j.scitotenv.2018.04.295
Rogula-Kozłowska W, Kozielska B, Klejnowski K (2013) Concentration, origin and health hazard from fine particle-bound PAH at three characteristic sites in Southern Poland. Bull Environ Contam Toxicol 91(3):349–355. https://doi.org/10.1007/s00128-013-1060-1
Saffari A, Daher N, Samara C, Voutsa D, Kouras A, Manoli E, Karagkiozidou O, Vlachokostas C, Moussiopoulos N, Shafer MM, Schauer JJ, Sioutas C (2013) Increased biomass burning due to the Economic Crisis in Greece and its adverse impact on Wintertime Air Quality in Thessaloniki. Environmental Science Technology 47(23):13313–13320. https://doi.org/10.1021/es403847h
Seabold S, Perktold J (2010) Statsmodels: Econometric and Statistical Modeling with Python. *Proc. of the 9th Python in Science Conf, December*, 92–96
Shahriyari HA, Nikmanesh Y, Jalali S, Tahery N, Zhiani Fard A, Hatamzadeh N, Zarea K, Cheraghi M, Mohammadi MJ (2022) Air pollution and human health risks: mechanisms and clinical manifestations of cardiovascular and respiratory diseases. Toxin Reviews 41(2):606–617. https://doi.org/10.1080/15569543.2021.1887261
Šimić I, Lovrić M, Godec R, Kröll M, Bešlić I (2020) Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon. Environ Pollut 263:114587. https://doi.org/10.1016/j.envpol.2020.114587
Singh S, Ashesh A, Linthoingambi Devi N, Yadav IC (2022) A comprehensive review on occurrence, source, effect, and measurement techniques of polycyclic aromatic hydrocarbons in India. Microchem J 183:108005. https://doi.org/10.1016/j.microc.2022.108005
Singh A, Banerjee T, Latif MT, Ramanathan S, Suradi H, Othman M, Murari V (2023) Molecular distribution, sources and potential health risks of fine particulate-bound polycyclic aromatic hydrocarbons during high pollution episodes in a subtropical urban city. Chemosphere 340:139943. https://doi.org/10.1016/j.chemosphere.2023.139943
Siudek P, Frankowski M (2018) The Role of sources and Atmospheric conditions in the Seasonal variability of particulate phase PAHs at the Urban Site in Central Poland. Aerosol Air Qual Res 18(6):1405–1418. https://doi.org/10.4209/aaqr.2018.01.0037
Soleimani M, Amini N, Sadeghian B, Wang D, Fang L (2018) Heavy metals and their source identification in particulate matter (PM2.5) in Isfahan City, Iran. J Environ Sci 72:166–175. https://doi.org/10.1016/j.jes.2018.01.002
Sopčić S, Pehnec G, Bešlić I (2024) Specific biomass burning tracers in air pollution in Zagreb, Croatia. Atmospheric Pollution Res 15(8):102176. https://doi.org/10.1016/j.apr.2024.102176
Sørensen M, Poulsen AH, Hvidtfeldt UA, Brandt J, Frohn LM, Ketzel M, Christensen JH, Im U, Khan J, Münzel T, Raaschou-Nielsen O (2022) Air pollution, road traffic noise and lack of greenness and risk of type 2 diabetes: a multi-exposure prospective study covering Denmark. Environ Int 170:107570. https://doi.org/10.1016/j.envint.2022.107570
Souza JB, Reisen VA, Franco GC, Ispány M, Bondon P, Santos JM (2018) Generalized additive models with principal component analysis: an application to Time Series of Respiratory Disease and Air Pollution Data. J Royal Stat Soc Ser C: Appl Stat 67(2):453–480. https://doi.org/10.1111/rssc.12239
Statheropoulos M, Vassiliadis N, Pappa A (1998) Principal component and canonical correlation analysis for examining air pollution and meteorological data. Atmos Environ 32(6):1087–1095. https://doi.org/10.1016/S1352-2310(97)00377-4
Sternbeck J, Sjödin Å, Andréasson K (2002) Metal emissions from road traffic and the influence of resuspension—results from two

tunnel studies. Atmos Environ 36(30):4735–4744. https://doi.org/10.1016/S1352-2310(02)00561-7
Stout SA, Magar VS, Uhler RM, Ickes J, Abbott J, Brenner R (2001) Characterization of naturally-occurring and anthropogenic PAHs in Urban Sediments-Wycoff/Eagle Harbor Superfund Site. Environ Forensics 2(4):287–300. https://doi.org/10.1080/713848281
Thurston GD, Ito K, Lall R (2011) A source apportionment of U.S. fine particulate matter air pollution. Atmos Environ 45(24):3924–3936. https://doi.org/10.1016/j.atmosenv.2011.04.070
Tsai J, Lin K, Chen C, Ding J, Choa C, Chiang H (2007) Chemical constituents in particulate emissions from an integrated iron and steel facility. J Hazard Mater 147(1–2):111–119. https://doi.org/10.1016/j.jhazmat.2006.12.054
Valavanidis A, Fiotakis K, Vlachogianni T (2008) Airborne Particulate Matter and Human Health: Toxicological Assessment and Importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. J Environ Sci Health Part C 26(4):339–362. https://doi.org/10.1080/10590500802494538
Viana M, Querol X, Alastuey A, Gil JI, Menéndez M (2006) Identification of PM sources by principal component analysis (PCA) coupled with wind direction data. Chemosphere 65(11):2411–2418. https://doi.org/10.1016/j.chemosphere.2006.04.060
Voiculescu M, Constantin D-E, Condurache-Bota S, Călmuc V, Roșu A, Dragomir Bălănică CM (2020) Role of Meteorological parameters in the diurnal and seasonal variation of NO2 in a Romanian Urban Environment. Int J Environ Res Public Health 17(17):6228. https://doi.org/10.3390/ijerph17176228
Wang X-T, Miao Y, Zhang Y, Li Y-C, Wu M-H, Yu G (2013) Polycyclic aromatic hydrocarbons (PAHs) in urban soils of the megacity Shanghai: occurrence, source apportionment and potential human health risk. Sci Total Environ 447:80–89. https://doi.org/10.1016/j.scitotenv.2012.12.086
Wang X, Zong Z, Tian C, Chen Y, Luo C, Tang J, Li J, Zhang G (2018) Assessing on toxic potency of PM2.5-bound polycyclic aromatic hydrocarbons at a national atmospheric background site in North China. Sci Total Environ 612:330–338. https://doi.org/10.1016/j.scitotenv.2017.08.208
Wang JM, Jeong C-H, Hilker N, Healy RM, Sofowote U, Debosz J, Su Y, Munoz A, Evans GJ (2021) Quantifying metal emissions from vehicular traffic using real world emission factors. Environ Pollut 268:115805. https://doi.org/10.1016/j.envpol.2020.115805
Wang J, Alli AS, Clark S, Hughes A, Ezzati M, Beddows A, Vallarino J, Nimo J, Bedford-Moses J, Baah S, Owusu G, Agyemang E, Kelly F, Barratt B, Beevers S, Agyei-Mensah S, Baumgartner J, Brauer M, Arku RE (2022) Nitrogen oxides (NO and NO2) pollution in the Accra metropolis: spatiotemporal patterns and the role of meteorology. Sci Total Environ 803:149931. https://doi.org/10.1016/j.scitotenv.2021.149931
Waskom M, Botvinnik O, O'Kane D, Hobson P, Lukauskas S, Gemperline DC, Augspurger T, Halchenko Y, Cole JB, Warmenhoven J, de Ruiter J, Pye C, Hoyer S, Vanderplas J, Villalba S, Kunter G, Quintero E, Bachant P, Martin M, Qalieh A (2017), September *Mwaskom/Seaborn: V0.8.1 (September 2017)*. https://doi.org/10.5281/zenodo.883859
World Bank (2019) Air Pollution and Human Health: the case of the Western Balkans, report – AQM in Bosnia and Herzegovina. AUS0001227
Wu Y, Li G, An T (2022) Toxic Metals in Particulate Matter and Health Risks in an E-Waste Dismantling Park and its surrounding areas: analysis of three PM size groups. Int J Environ Res Public Health 19(22). https://doi.org/10.3390/ijerph192215383
Xu S, Liu W, Tao S (2006) Emission of Polycyclic Aromatic hydrocarbons in China. Environmental Science Technology 40(3):702–708. https://doi.org/10.1021/es0517062
Zetlen HL, Lee S, Nurhussien A, Sun L, Kang W, Zanobetti C-M, A., Rice MB (2023) Personal air pollution exposure and metals in the nasal epithelial lining fluid of COPD patients. Environ Research: Health 1(2):021002. https://doi.org/10.1088/2752-5309/acbbe5
Zhang W, Liu B, Zhang Y, Li Y, Sun X, Gu Y, Dai C, Li N, Song C, Dai Q, Han Y, Feng Y (2020) A refined source apportionment study of atmospheric PM2.5 during winter heating period in Shijiazhuang, China, using a receptor model coupled with a source-oriented model. Atmos Environ 222:117157. https://doi.org/10.1016/j.atmosenv.2019.117157

5.3 Publication II

- *Račić, N., Ružičić, S., Petrić, V., Terzić, T., Antunović, M., Škaro, I., Pehnec, G., Bešlić, I., Jakovljević, I., Sever Štrukil, Z., Rinkovec, J., Žužul, S., Lovrić, M. (2025) Assessment of contributors to airborne PAHs and heavy metals in PM_{10} using temporal, spatial, traffic and heating data in explainable machine learning models. Atmospheric Environment: X, 29, https://doi.org/10.1016/j.aeaoa.2026.100413*

Contents lists available at ScienceDirect

Atmospheric Environment: X

journal homepage: www.journals.elsevier.com/atmospheric-environment-x

Assessment of contributors to airborne PAHs and heavy metals in PM_{10} using temporal, spatial, traffic and heating data in explainable machine learning models

Nikolina Račić [a,b], Stanko Ružičić [c], Valentino Petrić [d,b], Teo Terzić [d], Mario Antunović [d], Ivan Škaro [e], Gordana Pehnec [a], Ivan Bešlić [a], Ivana Jakovljević [a], Zdravka Sever Štrukil [a], Jasmina Rinkovec [a], Silva Žužul [a], Mario Lovrić [b,f,g,*]

[a] Institute for Medical Research and Occupational Health, Ksaverska cesta 2, 10000, Zagreb, Croatia
[b] The Lisbon Council, Rue de la Loi 155, 1040, Brussels, Belgium
[c] Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva 6, 10000, Zagreb, Croatia
[d] Ascalia d.o.o., Ulica Trate 16, Čakovec, Croatia
[e] Hrvatske autoceste d.o.o., Širolina 4, 10000, Zagreb, Croatia
[f] Institute for Anthropological Research, Gajeva 32, 10000, Zagreb, Croatia
[g] Josip Juraj Strossmayer University of Osijek, Faculty of Food Technology Osijek, Franje Kuhača 20, 31000, Osijek, Croatia

ARTICLE INFO

Keywords:
Gas consumption
Machine learning
Metals
PAHs
Random forest
Traffic data

ABSTRACT

Air pollution in urban areas originates from multiple interacting sources and is strongly influenced by meteorology, yet direct emission data are often incomplete. This study quantifies how meteorological conditions, station location, and proxy indicators of traffic and residential heating affect PM_{10}-bound polycyclic aromatic hydrocarbons (PAHs) and metals in Zagreb, Croatia. Daily concentrations of PM_{10}, selected PAHs, metals and NO_2 from four monitoring stations (2017–2020) were combined with local and ERA5 meteorology, highway traffic counts and gas consumption as emission proxies. Non-negative Matrix Factorization (NMF) was applied separately to PAHs and metals to identify dominant source-related patterns, while Random Forest regression and SHapley Additive Explanations (SHAP) were used to evaluate the influence of temporal, spatial, meteorological, traffic and heating predictors. NMF separated a heating-related PAH component dominated by Pyr and Flu from a traffic-related component characterised by BaA, Chry and BkF, and indicated enrichment of As and Pb at traffic- and industry-affected stations. Random Forest models showed higher predictive skill for PAHs ($R^2 \approx 0.60$–0.68) than for metals ($R^2 \approx 0.24$–0.42). Temperature and solar radiation were the main predictors for PAHs, whereas PM_{10}, NO_2 and station indicators dominated the prediction of metals. These results demonstrate that integrating proxy emission indicators with explainable machine learning provides an efficient framework for characterising sources and supports season- and location-specific air quality management in data-limited urban environments.

1. Introduction

Air pollution represents one of the most pressing environmental challenges of our time, with significant implications for both human health and the environment. It affects billions of people worldwide and contributes to respiratory and cardiovascular diseases, cancer, and adverse developmental outcomes (Li and Wang, 2024; Raaschou-Nielsen et al., 2011; Ramamoorthy et al., 2024; Sørensen et al., 2003; Wei et al., 2021). Beyond direct health effects, air pollution disrupts ecosystems, damages biodiversity, and accelerates climate change, making it a key focus of environmental and public health policies (Echendu et al., 2022; Kaur and Pandey, 2021). Among air pollutants, particulate matter (PM) is of particular concern due to its complex composition, multiple sources, and ability to penetrate deep into the respiratory system (Fazakas et al., 2024; Guo et al., 2022).

A specific concern is the presence of toxic constituents in PM, such as metals and polycyclic aromatic hydrocarbons (PAHs). These components contribute disproportionately to the toxic, mutagenic, and carcinogenic effects of airborne particles and thus have an important role in risk assessment and regulatory strategies. Understanding the factors that

* Corresponding author. Institute for Anthropological Research, Gajeva 32, 10000, Zagreb, Croatia.
E-mail address: mario.lovric@inantro.hr (M. Lovrić).

https://doi.org/10.1016/j.aeaoa.2026.100413
Received 8 October 2025; Received in revised form 12 January 2026; Accepted 14 January 2026
Available online 16 January 2026

control their concentrations and spatial–temporal variability is therefore essential for effective air quality management.

1.1. PAHs and metals in PM and their relevance for urban air quality

PAHs are a class of organic compounds primarily generated by incomplete combustion of organic materials such as fossil fuels, coal, wood, and oil (Dat and Chang, 2017; Kumar et al., 2016; Marchand et al., 2004; Patel et al., 2020; Račić et al., 2024; Ravindra et al., 2008). They are of high toxicological interest because of their carcinogenic and mutagenic properties, mediated through DNA damage, oxidative stress, and altered gene expression (Jakovljević et al., 2015; Kazensky et al., 2024; Kim et al., 2013; Masiol et al., 2012; Stading et al., 2021; Verma et al., 2021). While individual concentration levels of PAHs may not lead to immediate health effects, their mixtures have been observed to create synergistic toxic effects, as demonstrated in various studies (Engraff et al., 2011; Nováková et al., 2020; Olsson et al., 2010). Exposure to PAHs has been associated with a wide range of health outcomes, encompassing respiratory and cardiovascular diseases, cancer, and development abnormalities (Yang et al., 2021).

Previous research in Zagreb has shown statistically significant PAH content in the finer and more health-relevant $PM_{2.5}$ fraction, especially in winter (Šišović et al., 2005), highlighting the importance of combustion-related emissions during the heating season. PAHs are also present in other environmental matrices, including indoor air, water, soil, and food (Balcıoğlu, 2016; Jakovljević et al., 2022; Kljaković-Gašpić et al., 2024; Lovrić et al., 2024; Lovrić et al., 2024; Račić et al., 2025; Sellami et al., 2025; Wilcke, 2000; Yebra-Pimentel et al., 2015), highlighting their relevance for cumulative human exposure. Consequently, minimizing human exposure to PAHs has become a matter of importance in protecting health and reducing associated diseases.

Metals associated with PM constitute another major toxic component. Metal-containing compounds, originating from both natural and anthropogenic activities, can become entrapped in PM particles, facilitating their inhalation and subsequent deposition in the respiratory system (L.-C. Chen et al., 2022; Gilmour et al., 2004; Harrison and Yin, 2000; Manalis et al., 2005; Reshmy et al., 2025; Sielski et al., 2021; Thomaidis et al., 2003; Uzoekwe et al., 2021). Certain metals, such as lead, cadmium, and nickel, are neurotoxic and can impair cognitive development, especially in children, while arsenic and chromium are established carcinogens (Manalis et al., 2005; Popoola et al., 2018; Potter et al., 2021). Metals in PM can trigger inflammation and oxidative stress in the lungs upon inhalation, leading to respiratory symptoms and exacerbating conditions such as asthma and chronic obstructive pulmonary disease (COPD) (Albayrak et al., 2023; MacNee and Donaldson, 2003). Furthermore, their ability to translocate from the lungs into the bloodstream raises concerns about potential cardiovascular effects, including increased risk of heart attacks and strokes (Bhatnagar, 2022). These effects are mediated by multiple mechanisms, including interference with enzymatic systems and disruption of cellular processes (L. C. Chen and Lippmann, 2009; Wu et al., 2022).

Air pollution is not uniformly distributed within urban areas. Instead, pollutant levels and composition vary with proximity to emission sources such as traffic corridors, industrial facilities, and residential heating, and are further shaped by local meteorology and urban morphology. Numerus studies have shown that location itself can be a significant variable in understanding pollution trends, as different areas may experience varying levels of pollutants due to proximity to traffic, industrial activities, residential heating, and other localized sources (Nguyen et al., 2020; Sellami et al., 2025; Shukla et al., 2020; Suel et al., 2022). Examining the spatial distribution of PM-bound PAHs and metals within cities such as Zagreb is therefore crucial for identifying hotspots, designing targeted mitigation measures, and protecting vulnerable populations.

However, detailed emission inventories are often incomplete or unavailable, particularly for specific toxic components such as individual PAHs and trace metals. In such cases, proxy information is frequently used to approximate emission patterns. Temporal variables (e.g., day of week, season) and surrogate indicators have been used successfully as emission proxies in previous work in the region (Lovrić et al., 2021; Šimić et al., 2020). The increasing complexity of temporally and spatially resolved environmental datasets has also motivated the application of machine learning (ML) methods to improve understanding of pollutant dynamics and anthropogenic drivers (Lovrić et al., 2021; Petrić et al., 2024; Poelzl et al., 2025). By integrating various predictors like meteorological data, traffic intensity, and localized heating sources, ML models such as Land Use Regression (LUR) (Azmi et al., 2023) can effectively geographically map pollution levels and identify significant local contributing factors. Previous studies have demonstrated that models like Random Forest regression, neural networks and hybrid ML approaches improve predictive accuracy by accounting for complex, non-linear interactions among variables related to air pollutants (Kecorius et al., 2024; Petrić et al., 2024; Šimić et al., 2020). Despite this progress, there remains a lack of studies that focus specifically on PM_{10}-bound PAHs and metals and simultaneously integrate multiple proxy datasets – traffic, heating, meteorology, and spatial indicators – within explainable ML frameworks. In cities such as Zagreb, where emission inventories are incomplete, there is a particular need for approaches that combine proxy emission information with advanced modelling and interpretation tools to disentangle the contributions of heating, traffic, and regional background to toxic PM components.

The present study addresses this gap by analysing a four-year dataset (2017–2020) of PM_{10}, PAHs, metals, and NO_2 at four monitoring stations in Zagreb, Croatia, together with local and ERA5 meteorology, highway traffic counts, and gas consumption as proxy indicators of traffic and residential heating. Nonnegative Matrix Factorization (NMF) is used to identify dominant source-related patterns in PAHs and metals, while Random Forest regression and SHapley Additive Explanations (SHAP) are applied to quantify the relative importance and direction of effect of meteorological, temporal, traffic, heating, and spatial predictors. The study aims to:

- Characterise the spatial and temporal variability of PM_{10}-bound PAHs and metals across four urban monitoring locations in Zagreb
- Identify and interpret dominant source-related components for PAHs and metals using NMF
- Quantify the contributions of meteorology, traffic and heating proxies, and spatial location to pollutant variability using explainable ML

By integrating proxy-based emission information with interpretable ML models, the study provides a framework for source-oriented assessment of toxic PM components in data-limited urban environments and supports the development of season- and location-specific air quality management strategies.

2. Materials and methods

2.1. Study area description and pollutant assessment

Zagreb, the capital and largest city of Croatia, has a population of about 800,000 inhabitants (City of Zagreb, 2023), experiences moderate traffic levels that may contribute to local air pollution and associated impacts on public health. These challenges underscore the city's ongoing efforts to monitor, manage, and improve air quality for the well-being of its residents.

This study selected four air quality monitoring stations within Zagreb: IMI, ZG-1, ZG-3, and Siget (Fig. 1) based on their character (urban background or traffic) and their position within the city to capture spatial variability, as well as on the availability of pollutant data. At these stations, daily measurements of PM_{10} and its chemical content

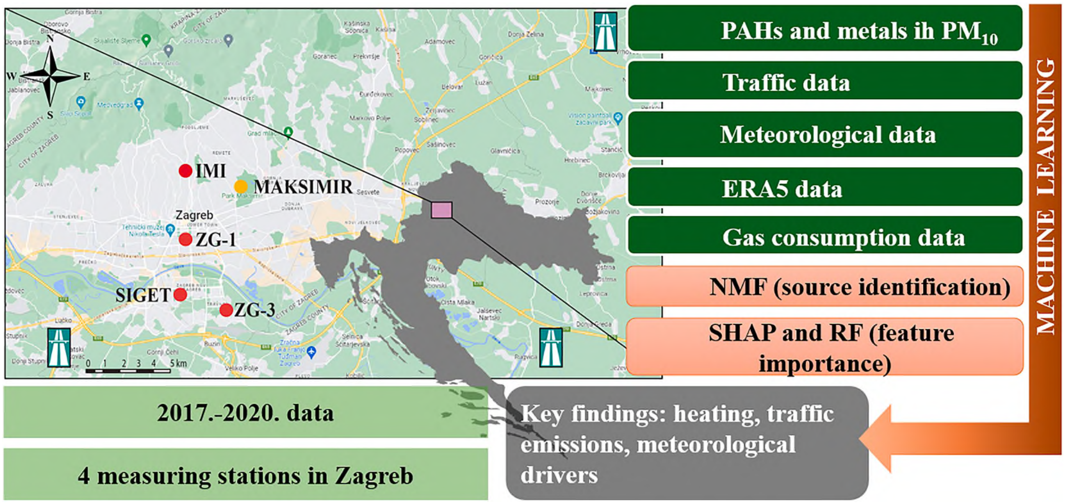


Fig. 1. Monitoring station locations with main highway access and an overview of the data and methods used. The figure was created using Google Maps and Adobe Illustrator. IMI – (altitude 116 m a.s.l., 45°50′7″ N, longitude 15°58′42″ E); Siget - (altitude 116 m a.s.l., latitude 45°46′25″ N, longitude 15°59′04″ E), ZG-1 - (altitude 113 m a.s.l., 45°48′18″ N, longitude 15°58′27″ E), ZG-3 - (altitude 116 m a.s.l., 45°40′46″ N, longitude 16°0′18″ E), Maksimir – meteorological station.

(PAHs, metals) were continuously carried out by standardized reference methods (CEN: EN 12341:2023, EN 14902:2005, EN 14902:2005/AC:2006, EN 15549:2008, CEN/TS 16645:2014). For this study, the period January 1, 2017, through December 31, 2020, was selected. Throughout this study, 24-h PM_{10} samples were collected continuously over four years. These samples were analyzed to determine concentrations of selected metals, including arsenic (As), cadmium (Cd), lead (Pb), manganese (Mn), iron (Fe), copper (Cu), and zinc (Zn), along with specific PAHs: benzo(a)pyrene (BaP), benzo(a)anthracene (BaA), benzo(ghi)perylene (BghiP), dibenzo(ah)anthracene (DahA), chrysene (Chry), benzo(k)fluoranthene (BkF), benzo(b)fluoranthene (BbF), benzo(j)fluoranthene (BjF), fluoranthene (Flu), and pyrene (Pyr). Additionally, nitrogen dioxide (NO_2) concentrations (EN 14211:2024) were monitored in parallel by automatic devices at IMI, ZG-1 and ZG-3 and by passive samplers at Siget (see Table 1). PM_{10} samples were collected on quartz or membrane filters using low-volume samplers (Sven Leckel, ~55 m^3/day). PAHs were extracted from the quartz filters using an ultrasonic bath with a cyclohexane and toluene solvent mixture, followed by centrifugation, evaporation to dryness, and re-dissolution in acetonitrile. The extracted samples were analyzed using high-performance liquid chromatography (HPLC) on an Agilent Infinity 1260 HPLC system equipped with a fluorescence detector, based on established protocols (Jakovljević et al., 2020). For metal analysis, quartz or membrane filter samples were digested using microwave-assisted nitric acid digestion (Ultraclave IV, Milestone Srl) and analyzed by inductively coupled plasma mass spectrometry (Agilent Technologies ICP-MS, model 7500cx) following methods detailed by Beslic et al. (2020). This extensive dataset, capturing daily levels of metals and PAHs in PM_{10}, provides valuable information about the composition and variability of PM in Zagreb. Meteorological information was obtained from the Zagreb - Maksimir weather station. Fig. 1 provides a visual representation of the monitoring locations mentioned and will be further detailed in the following sections. The monitoring stations were previously described in several studies (Jakovljević et al., 2018, 2021; Lovrić et al., 2022; Pehnec et al., 2016; Račić et al., 2025).

Monitoring stations IMI and Siget are part of the local air quality monitoring network, financially supported by the City of Zagreb, specifically, the City Office for Economy, Energetics, and Environment Protection. In contrast, the monitoring stations ZG1 and ZG-3 are part of the State network for continuous air quality monitoring, with funding emanating from The Environmental Protection and Energy Efficiency Fund of Croatia. Monitoring station IMI is classified as urban background and is situated within the northern residential part of Zagreb. It is positioned approximately 20 m away from a street characterized by moderate traffic density. This location is primarily characterized by single-family houses where heating predominantly relies on gas, wood or crude oil sources. Monitoring station Siget is located approximately 4 m above ground level and roughly 30 m distant from the nearest roadway. Situated within a densely populated urban sector, this monitoring station is exposed to significant traffic density. Moreover, it is subject to the prevailing northern winds originating from the city center and emissions from residential heating systems, which are fueled by gas, oil, or wood. Monitoring station ZG-1 is strategically positioned near the city center, situated alongside a heavily trafficked road. It is situated approximately 10 m from the intersection of two major and highly congested roads. This area is characterized by a high population density

Table 1
Summary statistics of the analyzed parameters for four monitoring stations in Zagreb, Croatia. The table shows the mean concentrations of the pollutants in ng/m^3 (metals, PAHs) and in μg/m^3 (PM_{10}, NO_2) with standard deviations.

variables	IMI	Siget	ZG-1	ZG-3	variables	IMI	Siget	ZG-1	ZG-3
Flu	0.51 (1.06)	0.83 (2.13)	0.74 (1.7)	0.98 (2.3)	**IP**	1.13 (1.61)	2.05 (3.27)	1.31 (2.23)	1.89 (3.06)
BaA	0.58 (1.1)	1.31 (3.07)	0.79 (1.91)	1.17 (2.63)	**Cd**	0.13 (0.13)	0.20 (0.19)	0.06 (0.08)	0.11 (0.22)
BaP	1.08 (1.77)	2.24 (4.23)	1.32 (2.64)	1.9 (3.52)	**Mn**	4.12 (3.48)	7.06 (4.29)	2.04 (2.87)	2.21 (3.42)
BbF	1.37 (2.07)	2.58 (4.32)	1.64 (2.83)	2.3 (3.68)	**Cu**	8.91 (9.82)	17.27 (13.17)	5.55 (9.0)	4.32 (8.83)
BghiP	1.14 (1.63)	2.23 (3.5)	1.33 (2.2)	1.85 (2.87)	**Fe**	254.48 (222.83)	536.51 (349.29)	169.23 (254.97)	133.72 (203.9)
BjF	0.84 (1.36)	1.53 (2.71)	1.24 (2.03)	1.36 (2.32)	**Zn**	14.4 (11.13)	28.84 (21.52)	7.10 (9.43)	9.37 (14.56)
BkF	0.54 (0.82)	1.01 (1.69)	0.63 (1.1)	0.89 (1.44)	**Pb**	4.77 (5.72)	7.88 (8.50)	2.04 (3.08)	4.08 (8.34)
DahA	0.16 (0.26)	0.3 (0.52)	0.17 (0.32)	0.24 (0.39)	**As**	0.34 (0.35)	0.41 (0.37)	0.12 (0.16)	0.20 (0.35)
Chry	0.99 (1.77)	2.11 (4.38)	1.34 (2.79)	1.87 (3.68)	**NO_2**	17 (11)	44 (17)	39 (15)	26 (11)
Pyr	0.51 (1.03)	0.90 (2.53)	0.72 (1.84)	0.91 (2.31)	**PM_{10}**	24 (17)	34 (23)	28 (21)	30 (26)

and many buildings mostly heated by natural gas. Monitoring station ZG-3 is located within a densely populated urban area, predominantly comprised of residential buildings and main roads characterized by heavy traffic flow. It is situated about 20 m away from a main road that connects numerous neighborhoods to the center of Zagreb.

2.2. Meteorological, traffic and heating data sources

In this study, time series data from two sources were used for meteorological information: local meteorological data and the ERA5 reanalysis dataset. This approach has also been utilized in our previous work by Petrić et al. (2024). Local meteorological data provided site-specific measurements, while ERA5 offered high-resolution, gridded global reanalysis data to supplement and cross-validate the local observations. The ERA5 data consists of meteorological parameters that were pre-calculated into daily values and offers high-resolution daily meteorological indicators that are suitable for environmental research, with parameters such as daily mean temperature, precipitation, solar radiation, humidity, wind speed and others (described in the Supplement). The ERA5 reanalysis data used in this study are provided on a regular latitude-longitude grid with a horizontal resolution of $0.25^\circ \times 0.25^\circ$, corresponding to approximately 31 km at the equator. The data represent hourly atmospheric variables at single levels near the surface and are part of ECMWF's fifth-generation climate reanalysis dataset (Copernicus Climate Change Service, 2018).

For the purposes of this research, traffic data were obtained from the "Hrvatske autoceste d.o.o." (http://www.hac.hr/hr). These data encompass information regarding the number of vehicles, categorized by driving directions and days of the week. Analysis was conducted on three main directions of traffic exits, covering the northeastern part of Sveta Helena (45°55′13.1"N 16°16′25.0"E), the southeastern part of Rugvica (45°45′58.5"N 16°13′52.6"E), and the vicinity of the Buzin and Lučko (45°44′55.6"N 15°53′03.2"E) interchanges (Fig. 1). These data provide valuable information about traffic patterns and were used as a proxy for in-city traffic to derive temporal traffic patterns and to characterise a pollution source surrounding the city. Daily heating gas consumption data for public buildings were obtained from Agency for legal transactions and real estate mediation (http://apn.hr/) as well as from Information system for energy management (http://isge.hr/login.xhtml) based on formal request. Incorporating the analysis of heating, particularly through daily gas consumption data aggregated for the given locations, adds a significant dimension to our understanding of the factors influencing air quality and is another proxy for emissions. According to previous findings, there is a significant correlation between heating gas consumption and pollutant levels on a monthly basis, indicating that heating has an important role in the dynamics of air pollution (Račić et al., 2024) which supports the use of these variables as emission proxies based on their temporal patterns.

2.3. Machine learning and statistics

The dataset for this study comprises daily records of local meteorological, satellite-derived, temporal, traffic data, heating data and air pollution data from four air quality monitoring stations. Predictor variables included local meteorological measurements (e.g., daily mean, minimum and maximum temperature, relative humidity, precipitation, wind speed and surface pressure), ERA5-derived indicators (e.g., clear-sky direct solar radiation at the surface, total column ozone, 10 m wind components), temporal variables (e.g., Julian day, day of week, month), traffic proxies derived from highway vehicle counts, heating proxies based on daily gas consumption (sum, mean and median), and station identifiers encoded as binary dummy variables (IMI, Siget, ZG-1, ZG-3). A complete overview of predictors is provided in the Supplementary Materials.

To ensure a substantial dataset for ML analysis, missing data points were imputed using a backfill strategy, where missing values were filled with data from the following day. The analysis was conducted using the Python programming language (www.python.org, v3.10), following the methods described in previous works (Lovrić et al., 2021; Šimić et al., 2020). It is assumed that pollutant concentrations can be effectively modelled using the given datasets of independent variables and their proxies. Data collected during the COVID-19 lockdown (March 15 to May 11, 2020) was excluded from the analysis due to the unique changes in traffic and industrial activities (Lovrić et al., 2021), which significantly altered pollution distributions. This exclusion ensures that the results accurately reflect typical urban conditions and provide reliable conclusions about spatial variability under regular activity levels.

Target pollutant concentration variables (metals and PAHs) exhibited skewed distributions. To improve the predictive performance of ML models (RF), the target variables were log-transformed. However, for Nonnegative Matrix Factorization (NMF) analysis, pollutant concentration data were normalized to a [0,1] range using *MinMaxScaler*, ensuring non-negative inputs suitable for NMF decomposition. In this study NMF was applied separately to PAHs and metals data to source-related components and compositional patterns (e.g., traffic, industrial activities, and residential heating), rather than to perform a full quantitative receptor-model source apportionment. For each pollutant group and station, a matrix of normalized daily concentrations was constructed and factorised into two non-negative components (rank $k = 2$), based on preliminary tests and interpretability of the resulting factors. NMF was implemented using the scikit-learn library with default regularisation settings. Components were interpreted to uncover relationships between pollutants and underlying source contributions.

The primary statistical approach employed in this study is inspired by Land Use Regression (LUR), a widely used method for modeling air pollution (Azmi et al., 2023; Ma et al., 2024; L. Zhang et al., 2021). Instead of classical LUR, this study employs location proxy-based ML models, with station information included as a variable in the models. The models aimed to capture spatial and temporal variability in pollutant concentrations across monitoring stations. The RF algorithm is an ensemble method consisting of many decision trees that utilize bootstrap aggregation (bagging) and feature randomness during training. It is particularly suited for modeling complex, non-linear relationships in environmental data (Kamińska, 2018; Lovrić et al., 2021, 2022; Račić et al., 2025; Rubal & Kumar, 2018; Šimić et al., 2020). Separate RF regression models were fitted for each pollutant (PAHs and metals), resulting in 18 models in total. For each model, the combined dataset was randomly split into a training subset (80 % of daily observations) and a testing subset (20 %) to evaluate out-of-sample performance. Hyperparameters such as the number of trees, maximum tree depth, minimum number of samples required to split an internal node and minimum number of samples required at a leaf node were optimised using a grid search with 5-fold cross-validation on the training set. The final configuration was selected based on the highest mean cross-validated coefficient of determination (R^2), and predictive performance was then quantified on the independent test set using R^2. The model outputs were evaluated using R^2 scores to assess the goodness of fit, and feature importance scores were analyzed to identify key drivers of pollutant variability.

In this analysis, RF models were used to: validate the predictive power of variables identified in the models and to investigate feature importances to compare with NMF and ML results. In addition, SHAP analysis was employed to enhance the interpretability of feature importance in predicting pollutant concentrations. SHAP values explain how individual predictor variables influence the model's output, providing understanding of their contribution to pollution variability. Tree-based SHAP values were computed for each RF model, providing local (sample-specific) contributions of each predictor to the predicted log-concentration. SHAP summary plots were generated to visualize the distribution, sign and magnitude of SHAP values for key predictors, distinguishing between positive and negative contributions to pollutant levels. This approach allowed for the identification of main meteorological and spatial drivers, confirming the robustness of the RF feature importance rankings. To further quantify and compare the significance of different predictor variables across pollutants, feature importance rankings were computed for each model, and the average ranking of each variable was calculated separately for PAHs and metals. Additionally, Kendall's Tau correlation analysis was performed on feature importance rankings to evaluate relationships between predictor variables, identifying shared drivers of PAH and metal concentrations.

3. Results

3.1. General pollution level

Prior to modelling, an explorative analysis of the pollution data was conducted. For this purpose, there is general statistics presented in Table 1. Table 1 provides an overview of the mean and standard deviation of pollutant concentrations across the four monitoring stations (IMI, Siget, ZG-1, and ZG-3). The highest average concentrations of PAHs, such as BaP, BbF, and Chry, were observed at Siget and ZG-3, while IMI and ZG-1 also showed notable levels. Among metals, Cu, Zn, and Fe had the highest concentrations at Siget and ZG-3, while Mn and Pb showed elevated values at IMI and ZG-1. The highest recorded NO_2 and PM_{10} concentrations were at Siget and ZG-3, with lower levels at IMI and ZG-1. This is also visible in Fig. 2, where the concentrations of all PAHs were significantly higher at the ZG-3 and Siget monitoring stations. Overall, the four stations share a broadly similar PAH composition dominated by 4–5-ring compounds (e.g., BaP, BbF, BkF, Chry), but Siget and ZG-3 exhibit higher absolute levels across most PAHs compared with stations IMI and ZG-1. For metals, Siget and ZG-3 are characterised by relatively high Cu, Zn and Fe, whereas IMI and ZG-1 show a larger contribution of Mn and Pb.

Based on the station-mean concentrations, the relative PAH composition (expressed as the percentage contribution to ΣPAH) is broadly similar across all four sites. Individual 4–5-ring PAHs such as BaP, BbF, BghiP, Chry and IP each contribute approximately 11–16 % of ΣPAH at all stations, together accounting for about 60–70 % of the total PAH mass (Table S1 in Supplementary materials). Flu and Pyr contribute a further ~5–7 % each, with slightly lower shares at Siget and ZG-3 compared with IMI and ZG-1. In contrast, the metal composition shows much stronger spatial heterogeneity. At IMI, ZG-1 and ZG-3, Fe dominates ΣMetal (≈87–91 %), with Zn contributing a further ≈4–6 % and all other metals individually accounting for ≤3 %. At Siget, Fe contributes only ≈44 % of ΣMetal, whereas As and Cd account for about 34 % and 17 %, respectively.

NMF was applied to both PAHs and metals as a factor-analytic tool to extract source-related components and compositional patterns, rather than as a full quantitative source apportionment model. The NMF biplots for PAHs (Fig. 3) provide a visualization of the spatial and temporal dynamics of PAHs across Zagreb at the four monitoring stations. Each biplot shows results through two main components, which reflect pollution sources or processes.

Across all stations, Pyr and Flu are dominant contributors to Component 2, with their longer arrow lengths indicating a strong influence on the observed pollution. While all stations exhibit similar overall patterns, the magnitude and direction of the arrows vary slightly, revealing site-specific variations. For instance, the IMI and Siget stations, situated in residential areas, exhibit a stronger influence of Chry and BjF, while ZG1 and ZG3, located closer to traffic-intensive urban areas, show a more pronounced contribution of BaA and BkF. Chry and BaA consistently appear between Flu and Pyr on the one side and remaining PAHs on the other, across all stations.

The NMF biplots for metals (Fig. 4) shows that across all stations, arsenic was the most influential metal to the Component 2, with long arrow lengths. Lead and cadmium contributed to both components, with their directional variations across stations indicating site-specific differences in source contributions while zinc, manganese, and iron are strongly aligned with Component 1 at all stations. Metal composition varied at different stations, with As and Pb contributing more at Siget and ZG3, while Zn, Mn, and Fe show a more balanced distribution at IMI and ZG1, and the tight clustering of samples near the origin in all biplots indicates relatively low variability in most observations.

3.2. Using machine learning to understand contributions

The following section interprets the results from RF developed to understand influence of previously described predictive variables on pollutant concentrations (metals and PAHs) across the monitoring stations. In total, 18 RF regression models were developed, one for each pollutant variable. Fig. 5 illustrates RF feature importance plot for cadmium (other pollutants in Supplementary materials), showing the dominant predictors influencing their concentration. The most significant variable is PM_{10}, which strongly correlates with the presence of metals, solar radiation (clear_sky_direct_solar_radiation_at_surface) also has impact, as well as the station-specific factors, particularly for ZG3 and Siget. Additionally, meteorological variables such as minimum temperature (temp_min), wind speed (m/s), and NO_2 further contribute to Cd variability.

This analysis was extended to all pollutants for better understanding of feature importance across PAHs and metals. The cumulative results are summarized in Table 2, which presents the top five most influential predictors for each pollutant alongside their corresponding R^2 scores. The overall trend indicates that temperature, PM, and NO_2 consistently rank among the most important predictors.

The RF models demonstrated varying predictive performance across pollutants, with R^2 values ranging from 0.245 (Pb) to 0.685 (BkF), indicating differences in how well pollutant concentrations can be

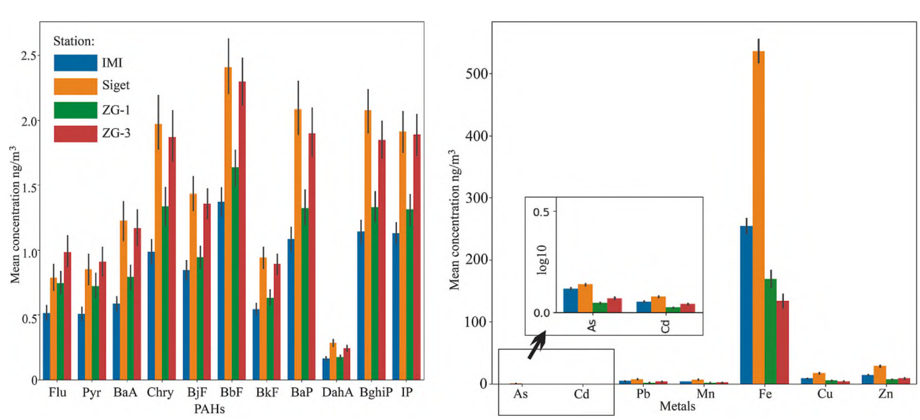


Fig. 2. Average concentrations of PAHs (in µg/m^3) and metals (in ng/m^3) at different monitoring stations.

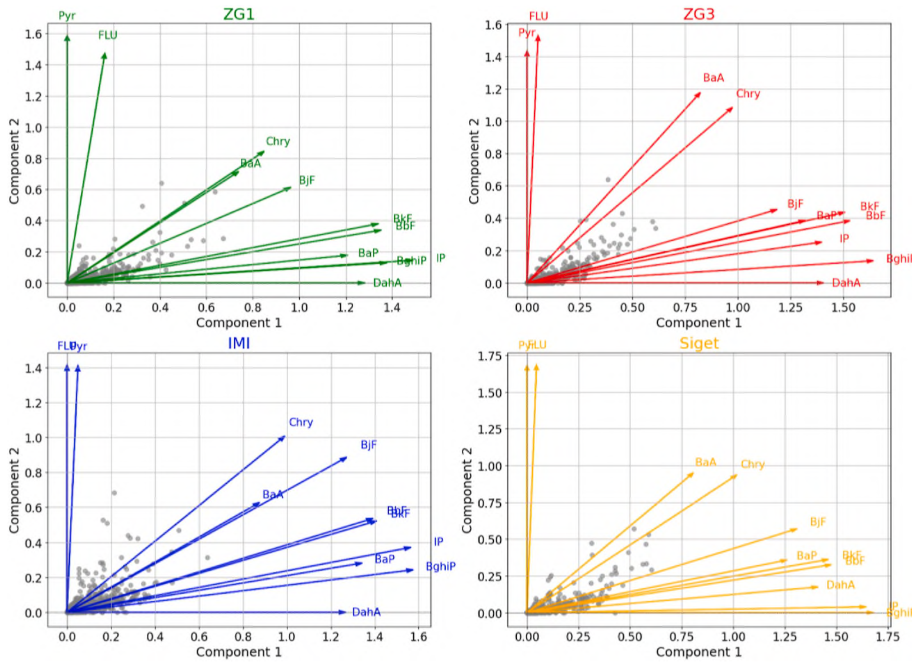


Fig. 3. Non-negative Matrix Factorization (NMF) biplots showing the spatial patterns of PAHs across the four monitoring stations (ZG-1, ZG-3, IMI, and Siget) in Zagreb.

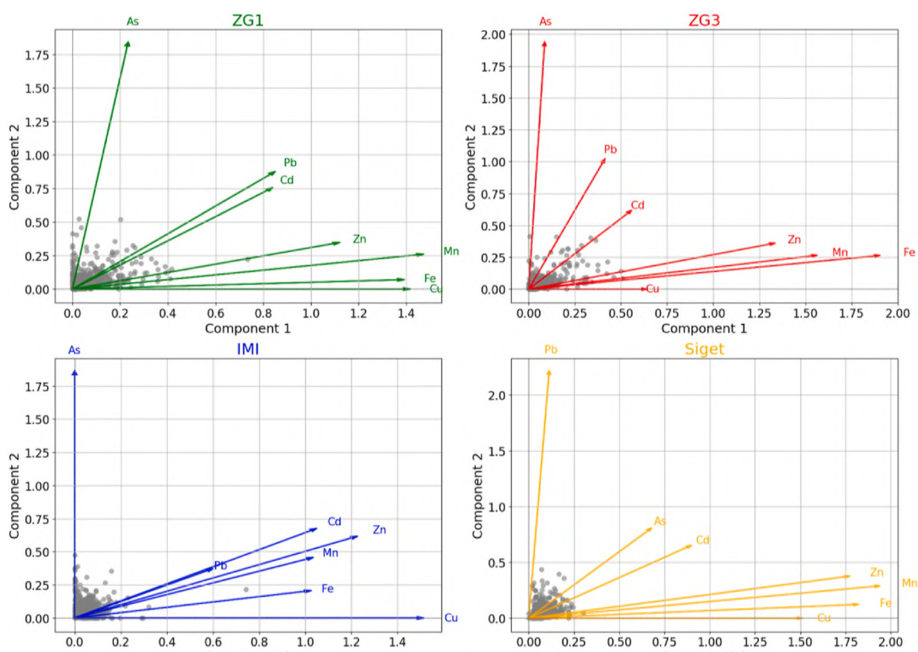


Fig. 4. Non-negative Matrix Factorization (NMF) biplots showing the spatial patterns of metals across the four monitoring stations (ZG-1, ZG-3, IMI, and Siget) in Zagreb.

explained by the selected predictor variables. In general, PAHs exhibited higher R^2 values, with BkF (0.685), BbF (0.679), and Chry (0.673) being among the best-predicted compounds. Among metals, Cd (0.419), Cu (0.419), and As (0.405) showed moderate predictive power, while Pb (0.245) and Mn (0.249) had the lowest R^2 values.

The feature importance ranking analysis allowed for a systematic evaluation of predictor variables across all pollutants using RF (Table S2 in Supplementary materials). The goal was to determine which variables have the highest explanatory power for pollutant concentrations and whether location-based models, meteorological variables, ERA5, heating, and traffic proxies provide meaningful informations. The results indicate that temperature-related variables are the most dominant predictors for PAHs, particularly BaP, BkF, and Chry, while PM_{10} and NO_2 are the strongest predictors for metals. NO_2, station_ZG3, station_Siget are more prominent in explaining Cu, Fe, and Mn variability, while heating proxies (Gas_Avg, Gas_Sum) contribute moderately to PAH predictions. The most significant ERA5 variables are temp_avg, clear_sky_direct_solar_radiation_at_surface, and 10m_u_component_of_wind.

To further explore the relationships between predictor variables across different pollutants, Kendall's Tau correlation matrices were

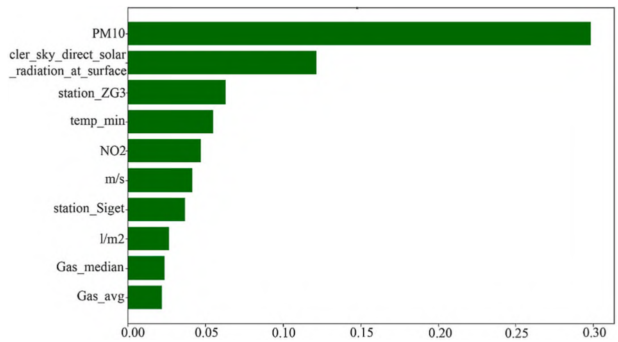


Fig. 5. Random Forrest feature importance plot for Cd ($R^2 = 0.419$). Feature importance shows the contribution of meteorological (e.g., temperature, solar radiation, wind speed), spatial (station indicators), and pollution-related variables (e.g., PM_{10}, NO_2) to the variability of cadmium concentrations.

Table 2
Main 5 feature importance (bold indicates negative impact) and R^2 for every pollutant.

Pollutant	Top Features	R^2 Score
Flu	**temp_min, temp_avg**, PM10, m/s, NO2	0.628
Pyr	**temp_min, temp_avg**, PM10, NO2, m/s	0.638
BaA	**temp_min, temp_avg**, clear_sky_direct_solar_radiation_at_surface, PM10, m/s	0.634
Chry	**temp_avg, temp_min**, NO2, PM10, clear_sky_direct_solar_radiation_at_surface	0.673
BjF	**temp_avg**, NO2, **temp_min**, clear_sky_direct_solar_radiation_at_surface, 10m_u_component_of_wind	0.654
BbF	**temp_avg**, clear_sky_direct_solar_radiation_at_surface, PM10, NO2, m/s	0.679
BkF	**temp_sred**, clear_sky_direct_solar_radiation_at_surface, NO2, PM10, **temp_min**	0.685
BaP	**temp_avg**, clear_sky_direct_solar_radiation_at_surface, m/s, **temp_min**, 10m_u_component_of_wind	0.641
DahA	**temp_avg**, clear_sky_direct_solar_radiation_at_surface, **temp_min**, m/s, PM10	0.626
BghiP	**temp_avg**, clear_sky_direct_solar_radiation_at_surface, 10m_u_component_of_wind, m/s, NO2	0.660
IP	**temp_avg**, clear_sky_direct_solar_radiation_at_surface, PM10, m/s, NO2	0.666
As	PM10, 10m_u_component_of_wind, Julian, Gas_Sum, Gas_Avg	0.405
Cd	PM10, clear_sky_direct_solar_radiation_at_surface, station_ZG3, **temp_min**, NO2	0.419
Pb	PM10, station_Siget, NO2, p_min, **mean_sea_level_pressure**	0.245
Cu	NO2, station_ZG3, Julian, station_Siget, PM10	0.419
Fe	station_Siget, PM10, NO2, total_column_ozone, m/s	0.277
Mn	station_Siget, PM10, NO2, 10m_u_component_of_wind, total_column_ozone	0.249
Pb	PM10, station_Siget, NO2, p_min, **mean_sea_level_pressure**	0.245
Zn	PM10, station_Siget, clear_sky_direct_solar_radiation_at_surface, **temp_min**, NO2	0.318

calculated based on the feature importance rankings from the RF models. These matrices (Fig. 6) shows the statistical associations between main predictors, revealing common drivers of PAH and metal concentrations. Strong correlations were observed among BaP, BbF, and BkF, as well as metals like Cd, Pb, and Zn. Unlike Pearson's correlation (which captures linear relationships), Kendall's Tau is more appropriate for feature importance rankings since it measures the consistency of order rather than absolute values.

The SHapley Additive Explanations (SHAP) analysis was conducted to complement the feature importance results obtained from the RF models, providing information about the influence of individual predictors on pollutant concentrations (Fig. 7).

The SHAP analysis provided a detailed interpretation of how meteorological, location-based, and emission-related variables influence pollutant levels by assessing both the magnitude and direction of their impact. Fig. 7 presents the SHAP feature importance plot for the best and worst-performing model (BkF, $R^2 = 0.685$, Pb, $R^2 = 0.245$), illustrating how high and low values of each feature affect predicted BkF and Pb concentrations. Additionally, solar radiation (clear_sky_direct_solar_radiation_at_surface) appeared as a significant contributor, where higher radiation levels correlated with lower BkF concentrations. Wind speed (10m_u_component_of_wind, m/s) showed a dispersive effect, where stronger winds led to lower pollutant concentrations. In this study, PM_{10} showed a stronger association with certain PAHs, such as BkF, BbF while others like BjF, BaP, DahA, and BghiP exhibited weaker relationships.

Pressure-related variables, including mean_sea_level_pressure, p_max, and p_min, showed significant associations with pollutant concentrations, while precipitation-related variables such as snowfall and total precipitation, along with total_column_ozone, exhibited lower feature importance scores.

4. Discussion

NMF revealed Pyr and Flu are main contributors to Component 2. These PAHs are likely associated with high-temperature combustion processes, such as fossil fuel and biomass burning from residential heating, which are prevalent sources in urban environments (Y. Chen et al., 2004; Jakovljević et al., 2018; Wang et al., 2024). This trend is further confirmed by the ML results, where SHAP analysis revealed temperature-related variables, PM_{10}, and NO_2 as the main predictors of PAH concentrations, indicating a strong influence of seasonal heating and traffic-related emissions. The negative SHAP values for solar radiation further support the impact of photochemical degradation in reducing PAH levels under high UV exposure. Conversely, Component 1 is heavily influenced by BaA, Chry, BjF, and BkF, which are commonly linked to traffic-related emissions and industrial activities, suggesting their prominence in localized sources (Dat and Chang, 2017; Kumar et al., 2016). The magnitude and direction of the arrows vary slightly, explaining site-specific variations in source contributions, environmental dispersion, and potential meteorological influences. For instance, the IMI and Siget stations, situated in residential areas, but at different part of the city, exhibit a stronger influence of Chry and BjF, indicative of emissions from mixed residential heating and industrial processes. Meanwhile, ZG-1 and ZG-3, located closer to main traffic-intensive areas, show a more pronounced contribution of BaA and BkF, which are well-known markers of vehicular emissions (Fang et al., 2024; Safo-Adu et al., 2025; W. Zhang et al., 2008).

The NMF biplots for metals shows that arsenic has the highest loading on the principal components, suggesting its potential association with episodic or spatially industrial emissions or resuspended soil contamination (Michalicová et al., 2024). Lead and cadmium contribute to both components, reflecting their origins in mixed sources such as traffic emissions (e.g., brake and tire wear) and industrial activities, with their directional variations across stations indicating site-specific differences in source contributions (Patil et al., 2024). Zinc, manganese, and iron are strongly aligned with Component 1 at all stations, indicating their association with stable, widespread sources such as natural dust, road resuspension, or industrial processes (Khare and Baruah, 2010; Pant and Harrison, 2013; Rajšić et al., 2008). Station-specific variations are evident, with Siget and ZG-3 showing slightly stronger contributions from As and Pb, while IMI and ZG-1 exhibit more balanced influences from metals like Zn, Mn, and Fe. These differences likely reflect the characteristics of the station locations: Siget and ZG-3 are positioned in the southern part of the city, closer to industrial zones, the municipal landfill and the airport, which may lead to localized enrichment with heavy metals like As and Pb. In contrast, IMI and ZG-1 are situated in more residential or mixed-use areas, where metal concentrations are more influenced by diffuse urban background sources rather than point-source emissions. The analysis of PAHs and metals across

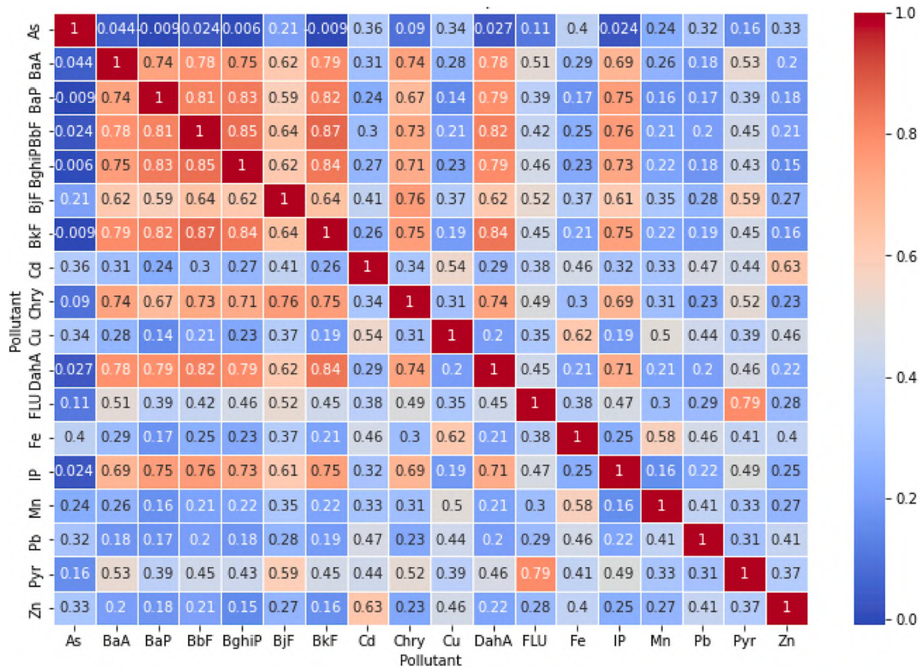


Fig. 6. Kendall's Tau correlation heatmap for pollutants feature importances.

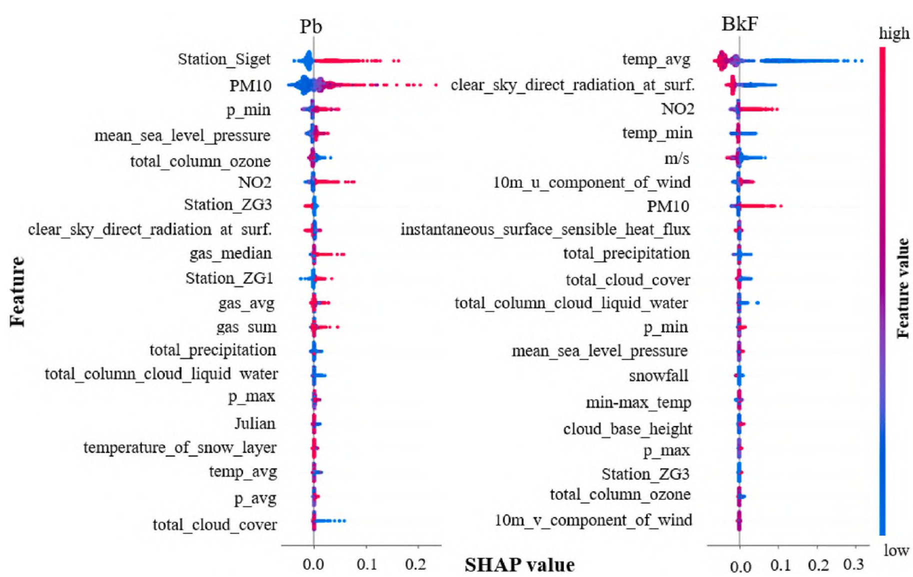


Fig. 7. SHAP Feature Importance for best (BkF) and worst (Pb) models. The x-axis shows the SHAP value, representing the impact of each feature on the model output. Positive SHAP values indicate an increase in pollutant concentration, while negative values indicate a decrease. Each point represents a sample; color gradient (blue to red) reflects low to high feature values. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Zagreb's monitoring stations identified heating and traffic as the primary pollution sources (Račić et al., 2024). PAHs such as BaA, Chry, and BkF, along with metals like Pb and Cd, are strongly associated with traffic emissions, originating from vehicle exhaust, brake wear, and tire abrasion (Ambade et al., 2021; Idani et al., 2020; Naydenova et al., 2022). Simultaneously, PAHs like Pyr and Flu, as well as As, are indicative of emissions from residential heating, particularly from the combustion of fossil fuels and biomass during colder months. The relationship of heating and traffic emissions indicate their combined impact on air quality in Zagreb, particularly during winter when both sources are most active (Pehnec et al., 2016, Šišović et al., 2002).

The RF and SHAP analyses show that, for metals, PM_{10} is the dominant predictor, which is consistent with its role as the main carrier of trace metals in the atmosphere. Among the meteorological variables, temperature, wind speed and solar radiation have the main influence on the accumulation and dispersion of pollutants, while station indicators (especially for ZG-3 and Siget) together with NO_2 point to the impact of local traffic and industrial activities. These results indicate that the variability of metal concentrations in Zagreb is determined by the combined effects of particle load, local emission sources and meteorological conditions, rather than by a single dominant factor.

The predictive power of the RF models varies across pollutants, with PAHs achieving higher R^2 values (~0.6–0.68) and metals exhibiting lower values (~0.24–0.42). The stronger predictive performance for

PAHs suggests that their concentrations are well explained by the selected predictor variables, likely due to their strong seasonal dependency, which enhances model predictability as the model relies on many time-based proxies for the missing emissions. In contrast, relatively low R^2 values for Pb, Mn, and Fe (~0.24–0.28) indicate that additional non-meteorological factors – such as industrial processes, urban dust resuspension, and localized emissions – contribute to their variability. Some specific emission sources were not directly captured in the available dataset, which may explain the lower predictive power for certain metals. The SHAP analysis confirms the conclusions from the RF models, revealing that meteorological conditions, particularly temperature, solar radiation, wind, and pressure – acting as proxies for missing direct emission data – have a main impact on PAH concentrations.

Kendall's Tau correlation heatmap show strong correlations among certain PAHs (BaP, BbF, and BkF) and between metals (Cd, Pb, and Zn), indicating shared influencing factors such as traffic emissions and heating activities which aligns with the NMF analysis. Identifying feature groups with strong correlations can help reduce model complexity while preserving predictive accuracy, enabling more effective interventions.

The feature importance analysis showed a strong overlap in main predictors for PAHs and metals, indicating shared sources. Temperature (temp_avg, temp_min, temp_max), humidity (rv_sred, rv_min, rv_max), and wind components were consistently influential across both pollutant groups, explaining the role of atmospheric dynamics in pollutant dispersion and transformation. The inclusion of NO_2 and PM_{10} in main variables explained the impact of traffic and industrial sources, while gas consumption indicators (Gas_Avg, Gas_Median, Gas_Sum) suggest a significant contribution from residential heating. Spatial variability was also evident, as station-related variables (ZG1, ZG3, Siget, IMI) were among the most important features, confirming location-dependent pollution sources. Additionally, meteorological variables related to precipitation, solar radiation, and pressure fluctuations indicate the influence of wet deposition, photochemical degradation, and atmospheric stability. Temperature-related variables (e.g., temp_avg, temp_min) are consistently the most influential factors across all PAHs. The strong negative correlation with temperature suggests that PAHs are largely influenced by residential heating, confirming the seasonal dependence of these pollutants. Solar radiation (clear_sky_direct_solar_radiation_at_surface) has a secondary but significant role, indicating the impact of photochemical degradation on PAH levels. Wind components (10m_u_component_of_wind, m/s) also show influence, confirming that atmospheric dispersion affects PAH concentrations. Metals are primarily associated with PM_{10} and traffic indicators. The strong correlation between PM_{10} and metals suggests that metals are largely transported via PM, which is expected given their association with combustion and industrial activities. NO_2 is another important factor, especially for Cu and Fe, explaining the connection between vehicular emissions and metal pollution. Location-based variables (station_Siget, station_ZG3) are highly ranked for some metals, suggesting that industrial and traffic-related sources contribute to localized pollution hotspots. The analysis revealed that station-related variables were much more important predictors for metals than for PAHs, suggesting that spatial variability have a greater role for metals. This likely reflects the influence of localized sources such as industry, traffic-related wear, and resuspension processes for metals, compared to the more diffuse distribution of PAHs driven by seasonal heating and widespread combustion activities. PAHs were better predicted overall, likely due to their strong dependence on meteorological conditions such as temperature and solar radiation, while the more sporadic and complex nature of metal emissions made them harder to model accurately.

The gas consumption indicators (Gas_Avg, Gas_Sum, Gas_Median) show moderate importance in metals and PAH models, particularly in the winter months, supporting their use as proxies for residential heating emissions. Traffic-related variables, particularly NO_2 and PM_{10}, were confirmed as strong proxies for vehicular pollution across multiple pollutants. However, the lack of direct in-city traffic monitoring might introduce some uncertainty in local traffic emission contributions.

NMF helped identify broad pollution sources (heating vs. traffic), while SHAP provided a more detailed quantification of how individual variables shape pollutant levels. Together, these analyses confirm that lower temperatures and stable, high-pressure winter conditions favour the accumulation of PAHs, whereas higher solar radiation and stronger winds promote their dispersion and degradation. Frequent winter inversion episodes in the Zagreb area likely contribute to the strong covariation of pollutants during the cold season, by suppressing vertical mixing and trapping pollutants near the surface. Precipitation- and ozone-related variables (snowfall, total precipitation, total_column_ozone) remained minor predictors in the models, suggesting that wet deposition and oxidative loss have a limited impact in controlling daily PAH variability under typical urban conditions.

These findings suggest that meteorological variables, PM_{10} concentrations, and ERA5-derived data remain the most influential predictors of pollutant variability. While station-specific influences also have a notable role, the contribution of traffic-related variables, although secondary in this analysis, shows the influence of high-traffic corridors near monitoring stations. To better capture the impact of urban traffic emissions in future studies, it would be beneficial to systematically monitor traffic activity within the city center and along major roads. Reducing heating-related emissions during colder months – such as through increased use of cleaner residential heating alternatives – could significantly lower PAH concentrations in urban areas. Additionally, traffic-related pollutants could be reduced through vehicle emissions regulations and improved public transportation infrastructure. Finally, industrial contributions, should be addressed through localized pollution controls, as certain metals exhibit strong site-specific trends.

4.1. *Merits and limitations of the methodological approach*

The location-based model works but requires further refinement, particularly regarding traffic data, which lacks direct in-city monitoring. ERA5 meteorological data proved highly relevant for pollutant prediction, particularly for PAHs, validating its integration into air quality modeling. Future studies should incorporate additional localized traffic data sources (e.g., GPS-based congestion data, street-level NO_2 monitoring) to improve the accuracy of traffic-related pollution modeling. The role of industrial contributions needs further exploration, as some metals (e.g., Mn, Pb) showed site-specific trends that may not be fully explained by the available variables. Several limitations of the applied techniques should be acknowledged. First, the approach relies on proxy variables for emissions, such as highway traffic counts and gas consumption in public buildings, rather than direct emission inventories. These proxies may not fully capture spatial differences in inner-city traffic or private heating practices, and thus introduce uncertainty into the attribution of sources. In the same way, the NMF analysis was designed to identify source-related components and compositional patterns, but it does not constitute a full receptor-model source apportionment with quantitative mass contributions and uncertainty estimates. Second, the RF models are trained on daily aggregated data, which limits the ability to resolve short-term pollution episodes and may smooth out peak emissions associated with specific meteorological conditions or traffic events. Methodologically, the use of backfill imputation for missing data and the presence of correlated predictors may influence feature importance estimates and SHAP values, even though such techniques are standard in environmental applications. RF and SHAP describe associations rather than causal effects; thus, the identified drivers should be interpreted as predictors of variability rather than definitive causal determinants. Furthermore, model evaluation is based on internal train–test splits within the same city and period, so the transferability of the models to other regions or future conditions is not directly assessed. These limitations highlight opportunities for further refinement, including the use of more detailed traffic and emission data, exploration of alternative factorization ranks and modelling approaches, and external validation in other urban settings. A comprehensive quantitative source apportionment study (e.g., using receptor models with detailed source profiles) is therefore identified as an important next step to build on the patterns and drivers identified in this work.

5. Conclusion

This study analyzed the meteorological, temporal, and spatial factors influencing pollutant concentrations in Zagreb, Croatia, with a focus on PAHs and metals. The findings confirm that temperature and solar radiation have a main role in explaining air pollution dynamics, particularly for PAHs, which show higher concentrations during colder months due to increased residential heating emissions. The strong relationship between BaP, BkF, and Chry and temperature-related variables shows the seasonal impact of heating activities, while solar radiation influences PAH degradation, particularly under high UV exposure. Additionally, PM_{10} and NO_2 emerged in models as main indicators of anthropogenic emissions, being among the most important predictors for metals such as As, Cd, and Pb, which mainly originate from traffic emissions, industrial activities, and other combustion processes. Spatial analysis revealed localized pollution hotspots, with Siget and ZG-3 exhibiting stronger links to traffic and industrial sources, whereas IMI and ZG-1 were more influenced by regional atmospheric conditions and heating emissions. The RF models successfully identified the dominant predictors of pollutant variability, with higher predictive performance for PAHs (R^2 ~0.60–0.68) compared to metals (R^2 ~0.24–0.42), suggesting that combustion-related sources (heating and traffic) follow more stable seasonal patterns, whereas metals are influenced by additional episodic or industrial processes. SHAP analysis further validated the main predictors, confirming that meteorological conditions, particularly wind speed, pressure, and temperature, influence pollutant transport and accumulation.

The results emphasize the need for seasonally adaptive air quality management strategies, particularly targeting residential heating emissions during winter and traffic and industrial pollution in high-density areas. By incorporating these findings into urban planning and policy decisions, targeted pollution control measures can be developed to improve air quality and minimize the health and environmental impacts of air pollution in Zagreb and similar urban environments.

CRediT authorship contribution statement

Nikolina Račić: Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Stanko Ružičić:** Writing – review & editing, Supervision. **Valentino Petrić:** Software. **Teo Terzić:** Software. **Mario Antunović:** Software. **Ivan Škaro:** Data curation. **Gordana Pehnec:** Writing – review & editing, Supervision. **Ivan Bešlić:** Data curation. **Ivana Jakovljević:** Data curation. **Zdravka Sever Štrukil:** Data curation. **Jasmina Rinkovec:** Data curation. **Silva Žužul:** Data curation. **Mario Lovrić:** Writing – review & editing, Visualization, Supervision, Methodology, Conceptualization.

Ethics approval and consent to participate

Since this study did not collect personal information, no ethical approval was required.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT (OpenAI, GPT-4) for support in developing and debugging code. The authors reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Funding

This study was funded by the European Union – Next Generation EU (Program Contract of December 8, 2023, Class: 643-02/23-01/00016, Reg. no. 533-03-23-0006)-EnvironPollutHealth. ML is supported by Horizon Europe (URBREATH), Grant ID: 101139711 granted to the Lisbon Council and European Union – Next Generation EU (BioAntroPop), Grant ID: IA-INT-2024 granted to the Institute for Anthropological Research, Zagreb, Croatia.

Declaration of competing interest

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

Acknowledgments

The authors express their gratitude to Hrvatske autoceste d.o.o. (http://www.hac.hr/hr) for providing essential traffic data, the Agency for Legal Transactions and Real Estate Mediation (http://apn.hr/) for their support, and the Information System for Energy Management (ISGE) (http://isge.hr/login.xhtml) for access to energy consumption data. This study was performed using the facilities and equipment funded within the European Regional Development Fund project KK.01.1. February 1, 0007 "Research and Education Centre of Environmental Health and Radiation Protection – Reconstruction and Expansion of the Institute for Medical Research and Occupational Health" and KK.06.2. February 1, 0001 "AIRQ – Expansion and Modernisation of the National Network for Continuous Air Quality Monitoring".

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aeaoa.2026.100413.

Data availability

Data will be made available on request.

References

Albayrak, L., Türksoy, V.A., Khalilov, R., Eftekhari, A., 2023. Investigation of heavy metal exposure and trace element levels in acute exacerbatıon of COPD. J. King Saud Univ. Sci. 35 (1), 102422. https://doi.org/10.1016/j.jksus.2022.102422.

Ambade, B., Kumar, A., Latif, M., 2021. Emission sources, characteristics and risk assessment of particulate bound polycyclic aromatic hydrocarbons (PAHs) from traffic sites. https://doi.org/10.21203/rs.3.rs-328364/v1.

Azmi, W.N.F.W., Pillai, T.R., Latif, M.T., Koshy, S., Shaharudin, R., 2023. Application of land use regression model to assess outdoor air pollution exposure: a review. Environ. Adv. 11, 100353. https://doi.org/10.1016/j.envadv.2023.100353.

Balcıoğlu, E.B., 2016. Potential effects of polycyclic aromatic hydrocarbons (PAHs) in marine foods on human health: a critical review. Toxin Rev. 35 (3–4), 98–105. https://doi.org/10.1080/15569543.2016.1201513.

Beslic, I., Burger, J., Cadoni, F., Centioli, D., Kranjc, I., Van Den Bril, B., Rinkovec, J., Sega, K., Zang, T., Zuzul, S., Gladtke, D., 2020. Determination of As, Cd, Ni and Pb in PM10 – comparison of different sample work-up and analysis methods/Bestimmung von As, Cd, Ni und Pb in PM10 – Vergleich verschiedener Probenaufbereitungs- und Analysenverfahren. Gefahrstoffe 80 (6), 227–233. https://doi.org/10.37544/0949-8036-2020-06-17.

Bhatnagar, A., 2022. Cardiovascular effects of particulate air pollution. Annu. Rev. Med. 73 (1), 393–406. https://doi.org/10.1146/annurev-med-042220-011549.

Chen, L.C., Lippmann, M., 2009. Effects of metals within ambient air Particulate Matter (PM) on human health. Inhal. Toxicol. 21 (1), 1–31. https://doi.org/10.1080/08958370802105405.

Chen, L.-C., Maciejczyk, P., D. Thurston, G., 2022. Metals and air pollution. In: Handbook on the Toxicology of Metals. Elsevier, pp. 137–182. https://doi.org/10.1016/B978-0-12-823292-7.00004-8.

Chen, Y., Bi, X., Mai, B., Sheng, G., Fu, J., 2004. Emission characterization of particulate/gaseous phases and size association for polycyclic aromatic hydrocarbons from residential coal combustion. Fuel 83 (7–8), 781–790. https://doi.org/10.1016/j.fuel.2003.11.003.

City of Zagreb, 2023. a. Statistički Ljetopis Grada Zagreb.

Copernicus Climate Change Service, 2018. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/CDS.ADBB2D47.
Dat, N.-D., Chang, M.B., 2017. Review on characteristics of PAHs in atmosphere, anthropogenic sources and control technologies. Sci. Total Environ. 609, 682–693. https://doi.org/10.1016/j.scitotenv.2017.07.204.
Echendu, A.J., Okafor, H.F., Iyiola, O., 2022. Air pollution, climate change and ecosystem health in the niger Delta. Soc. Sci. 11 (11), 525. https://doi.org/10.3390/socsci11110525.
Engraff, M., Solere, C., Smith, K.E.C., Mayer, P., Dahllöf, I., 2011. Aquatic toxicity of PAHs and PAH mixtures at saturation to benthic amphipods: linking toxic effects to chemical activity. Aquat. Toxicol. 102 (3–4), 142–149. https://doi.org/10.1016/j.aquatox.2011.01.009.
Fang, T., Fu, J., Gao, Y., Song, A., Zhang, Y., Zhang, Q., Wu, L., Peng, J., Wang, T., Mao, H., 2024. Real-world vehicular source indicators for exhaust and non-exhaust contribution to $PM_{2.5}$ during peak and off-peak hours using tunnel measurement. J. Geophys. Res. Atmos. 129 (13). https://doi.org/10.1029/2024JD040853 e2024JD040853.
Fazakas, E., Neamtiu, I.A., Gurzau, E.S., 2024. Health effects of air pollutant mixtures (volatile organic compounds, particulate matter, sulfur and nitrogen oxides) – a review of the literature. Rev. Environ. Health 39 (3), 459–478. https://doi.org/10.1515/reveh-2022-0252.
Gilmour, P.S., Ziesenis, A., Morrison, E.R., Vickers, M.A., Drost, E.M., Ford, I., Karg, E., Mossa, C., Schroeppel, A., Ferron, G.A., Heyder, J., Greaves, M., MacNee, W., Donaldson, K., 2004. Pulmonary and systemic effects of short-term inhalation exposure to ultrafine carbon Black particles. Toxicol. Appl. Pharmacol. 195 (1), 35–44. https://doi.org/10.1016/j.taap.2003.10.003.
Guo, C., Lv, S., Liu, Y., Li, Y., 2022. Biomarkers for the adverse effects on respiratory system health associated with atmospheric particulate matter exposure. J. Hazard Mater. 421, 126760. https://doi.org/10.1016/j.jhazmat.2021.126760.
Harrison, R.M., Yin, J., 2000. Particulate matter in the atmosphere: which particle properties are important for its effects on health? Sci. Total Environ. 249 (1–3), 85–101. https://doi.org/10.1016/S0048-9697(99)00513-6.
Idani, E., Geravandi, S., Akhzari, M., Goudarzi, G., Alavi, N., Yari, A.R., Mehrpour, M., Khavasi, M., Bahmaei, J., Bostan, H., Dobaradaran, S., Salmanzadeh, S., Mohammadi, M.J., 2020. Characteristics, sources, and health risks of atmospheric PM_{10} -bound heavy metals in a populated middle eastern city. Toxin Rev. 39 (3), 266–274. https://doi.org/10.1080/15569543.2018.1513034.
Jakovljević, I., Dvoršćak, M., Jagić, K., Klinčić, D., 2022. Polycyclic aromatic hydrocarbons in indoor dust in Croatia: levels, sources, and human health risks. Int. J. Environ. Res. Publ. Health 19 (19). https://doi.org/10.3390/ijerph191911848. Article 19.
Jakovljević, I., Pehnec, G., Vađić, V., Čačković, M., Tomašić, V., Jelinić, J.D., 2018. Polycyclic aromatic hydrocarbons in PM10, PM2.5 and PM1 particle fractions in an urban area. Air Qual. Atmos. Health 11 (7), 843–854. https://doi.org/10.1007/s11869-018-0603-3.
Jakovljević, I., Pehnec, G., Vadjić, V., Šišović, A., Davila, S., Bešlić, I., 2015. Carcinogenic activity of polycyclic aromatic hydrocarbons bounded on particle fraction. Environ. Sci. Pollut. Control Ser. 22 (20), 15931–15940. https://doi.org/10.1007/s11356-015-4777-z.
Jakovljević, I., Sever Štrukil, Z., Godec, R., Bešlić, I., Davila, S., Lovrić, M., Pehnec, G., 2020. Pollution sources and carcinogenic risk of PAHs in PM1 particle fraction in an urban area. Int. J. Environ. Res. Publ. Health 17 (24), 9587. https://doi.org/10.3390/ijerph17249587.
Jakovljević, I., Štrukil, Z.S., Godec, R., Davila, S., Pehnec, G., 2021. Influence of lockdown caused by the COVID-19 pandemic on air pollution and carcinogenic content of particulate matter observed in Croatia. Air Qual. Atmos. Health 14 (4), 467–472. https://doi.org/10.1007/s11869-020-00950-3.
Kamińska, J.A., 2018. The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: a case study in Wrocław. J. Environ. Manag. 217, 164–174. https://doi.org/10.1016/j.jenvman.2018.03.094.
Kaur, R., Pandey, P., 2021. Air pollution, climate change, and human health in Indian cities: a brief review. Front. Sustain. Cities 3, 705131. https://doi.org/10.3389/frsc.2021.705131.
Kazensky, L., Matković, K., Gerić, M., Žegura, B., Pehnec, G., Gajski, G., 2024. Impact of indoor air pollution on DNA damage and chromosome stability: a systematic review. Arch. Toxicol. https://doi.org/10.1007/s00204-024-03785-4.
Kecorius, S., Madueño, L., Lovric, M., Racic, N., Schwarz, M., Cyrys, J., Casquero-Vera, J. A., Alados-Arboledas, L., Conil, S., Sciare, J., Ondracek, J., Hallar, A.G., Gómez-Moreno, F.J., Ellul, R., Kristensson, A., Sorribas, M., Kalivitis, N., Mihalopoulos, N., Peters, A., et al., 2024. Atmospheric new particle formation identifier using longitudinal global particle number size distribution data. Sci. Data 11 (1), 1239. https://doi.org/10.1038/s41597-024-04079-1.
Khare, P., Baruah, B.P., 2010. Elemental characterization and source identification of PM2.5 using multivariate analysis at the suburban site of North-East India. Atmos. Res. 98 (1), 148–162. https://doi.org/10.1016/j.atmosres.2010.07.001.
Kim, K.-H., Jahan, S.A., Kabir, E., Brown, R.J.C., 2013. A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. Environ. Int. 60, 71–80. https://doi.org/10.1016/j.envint.2013.07.019.
Kljaković-Gašpić, Z., Tariba Lovaković, B., Smoljo, I., Jurič, A., Orct, T., Sekovanić, A., Brajenović, N., Brčić Karačonji, I., Pehnec, G., Lovrić, M., Pleština, M., Šprajc, E., Lovković, S., Jurasović, J., 2024. Metal(loid)s, phthalate esters and polycyclic aromatic hydrocarbons in Croatian natural mineral waters: regulatory compliance and associated health risk. Environ. Technol. Innovat. 34, 103570. https://doi.org/10.1016/j.eti.2024.103570.
Kumar, V., Kothiyal, N.C., Saruchi, Vikas, P., Sharma, R., 2016. Sources, distribution, and health effect of carcinogenic polycyclic aromatic hydrocarbons (PAHs) – current knowledge and future directions. J. Chin. Adv. Mater. Soc. 4 (4), 302–321. https://doi.org/10.1080/22243682.2016.1230475.
Li, W., Wang, W., 2024. Causal effects of exposure to ambient air pollution on cancer risk: insights from genetic evidence. Sci. Total Environ. 912, 168843. https://doi.org/10.1016/j.scitotenv.2023.168843.
Lovrić, M., Antunović, M., Šunić, I., Vuković, M., Kecorius, S., Kröll, M., Bešlić, I., Šimić, I., Pehnec, G., 2022. Insignificant changes in particulate matter during the COVID-19 lockdown: a machine learning study in Zagreb, Croatia. In: The 7th World Congress on Civil, Structural, and Environmental Engineering. https://doi.org/10.11159/iceptp22.187.
Lovrić, M., Gajski, G., Fernández-Agüera, J., Pöhlker, M., Gursch, H., Consortium, T.E., Borg, A., Switters, J., Mureddu, F., 2024a. Evidence-driven indoor air quality improvement: an innovative and interdisciplinary approach to improving indoor air quality. Biofactors. https://doi.org/10.1002/biof.2126 n/a(n/a).
Lovrić, M., Pavlović, K., Vuković, M., Grange, S.K., Haberl, M., Kern, R., 2021. Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. Environ. Pollut. 274, 115900. https://doi.org/10.1016/j.envpol.2020.115900.
Lovrić, M., Račić, N., Pehnec, G., Horvat, T., Lovrić Štefiček, M.J., Jakovljević, I., 2024b. Indoor polycyclic aromatic hydrocarbons—relationship to ambient air, risk estimation, and source apportionment based on household measurements. Atmosphere 15 (12), 1525. https://doi.org/10.3390/atmos15121525.
Ma, X., Zou, B., Deng, J., Gao, J., Longley, I., Xiao, S., Guo, B., Wu, Y., Xu, T., Xu, X., Yang, X., Wang, X., Tan, Z., Wang, Y., Morawska, L., Salmond, J., 2024. A comprehensive review of the development of land use regression approaches for modeling spatiotemporal variations of ambient air pollution: a perspective from 2011 to 2023. Environ. Int. 183, 108430. https://doi.org/10.1016/j.envint.2024.108430.
MacNee, W., Donaldson, K., 2003. Mechanism of lung injury caused by PM10 and ultrafine particles with special reference to COPD. Eur. Respir. J. 21 (Suppl. 40), 47S–51s. https://doi.org/10.1183/09031936.03.00403203.
Manalis, N., Grivas, G., Protonotarios, V., Moutsatsou, A., Samara, C., Chaloulakou, A., 2005. Toxic metal content of particulate matter (PM10), within the greater area of Athens. Chemosphere 60 (4), 557–566. https://doi.org/10.1016/j.chemosphere.2005.01.003.
Marchand, N., Besombes, J.L., Chevron, N., Masclet, P., Aymoz, G., Jaffrezo, J.L., 2004. Polycyclic aromatic hydrocarbons (PAHs) in the atmospheres of two French alpine valleys: sources and temporal patterns. Atmos. Chem. Phys. 4 (5), 1167–1181. https://doi.org/10.5194/acp-4-1167-2004.
Masiol, M., Hofer, A., Squizzato, S., Piazza, R., Rampazzo, G., Pavoni, B., 2012. Carcinogenic and mutagenic risk associated to airborne particle-phase polycyclic aromatic hydrocarbons: a source apportionment. Atmos. Environ. 60, 375–382. https://doi.org/10.1016/j.atmosenv.2012.06.073.
Michalicová, R., Hegrová, J., Svoboda, J., Ličbinský, R., 2024. Seasonal and spatial variations of arsenic and its species in particulate matter in an urban environment of brno, Czech Republic. Environ. Sci. Pollut. Control Ser. 31 (43), 55251–55262. https://doi.org/10.1007/s11356-024-34645-4.
Naydenova, St, Veli, A., Mustafa, Z., Hudai, S., Hristova, E., Gonsalvesh-Musakova, L., 2022. Atmospheric levels, distribution, sources, correlation with meteorological parameters and other pollutants and health risk of PAHs bound in $PM_{2.5}$ and PM_{10} in Burgas, Bulgaria – a case study. J. Environ. Sci. Health Part A 57 (4), 306–317. https://doi.org/10.1080/10934529.2022.2060669.
Nguyen, T.N.T., Kwon, H.-O., Lammel, G., Jung, K.-S., Lee, S.-J., Choi, S.-D., 2020. Spatially high-resolved monitoring and risk assessment of polycyclic aromatic hydrocarbons in an industrial city. J. Hazard Mater. 393, 122409. https://doi.org/10.1016/j.jhazmat.2020.122409.
Nováková, Z., Novák, J., Kitanovski, Z., Kukučka, P., Smutná, M., Wietzoreck, M., Lammel, G., Hilscherová, K., 2020. Toxic potentials of particulate and gaseous air pollutant mixtures and the role of PAHs and their derivatives. Environ. Int. 139, 105634. https://doi.org/10.1016/j.envint.2020.105634.
Olsson, A.C., Fevotte, J., Fletcher, T., Cassidy, A., 'T Mannetje, A., Zaridze, D., Szeszenia-Dabrowska, N., Rudnai, P., Lissowska, J., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Brennan, P., Boffetta, P., 2010. Occupational exposure to polycyclic aromatic hydrocarbons and lung cancer risk: a multicenter study in Europe. Occup. Environ. Med. 67 (2), 98–103. https://doi.org/10.1136/oem.2009.046680.
Pant, P., Harrison, R.M., 2013. Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: a review. Atmos. Environ. 77, 78–97. https://doi.org/10.1016/j.atmosenv.2013.04.028.
Patel, A.B., Shaikh, S., Jain, K.R., Desai, C., Madamwar, D., 2020. Polycyclic aromatic hydrocarbons: sources, toxicity, and remediation approaches. Front. Microbiol. 11, 562813. https://doi.org/10.3389/fmicb.2020.562813.
Patil, M., Jadhav, A.S., Singh, S.K., Ingle, S., 2024. Urban ambient PM10 analysis: characterization, sources, health risk and sustainable solutions. Environ. Chem. Ecotoxicol. 6, 192–205. https://doi.org/10.1016/j.enceco.2024.05.008.
Pehnec, G., Jakovljević, I., Šišović, A., Bešlić, I., Vađić, V., 2016. Influence of ozone and meteorological parameters on levels of polycyclic aromatic hydrocarbons in the air. Atmos. Environ. 131, 263–268. https://doi.org/10.1016/j.atmosenv.2016.02.009.
Petrić, V., Hussain, H., Časni, K., Vuckovic, M., Schopper, A., Andrijić, Ž.U., Kecorius, S., Madueno, L., Kern, R., Lovrić, M., 2024. Ensemble machine learning, deep learning, and time series forecasting: improving prediction accuracy for hourly concentrations of ambient air pollutants. Aerosol Air Qual. Res. 24, 230317. https://doi.org/10.4209/aaqr.230317.

Poelzl, M., Kern, R., Kecorius, S., Lovrić, M., 2025. Exploration of transfer learning techniques for the prediction of PM10. Sci. Rep. 15 (1), 2919. https://doi.org/10.1038/s41598-025-86550-6.

Popoola, L.T., Adebanjo, S.A., Adeoye, B.K., 2018. Assessment of atmospheric particulate matter and heavy metals: a critical review. Int. J. Environ. Sci. Technol. 15 (5), 935–948. https://doi.org/10.1007/s13762-017-1454-4.

Potter, N.A., Meltzer, G.Y., Avenbuan, O.N., Raja, A., Zelikoff, J.T., 2021. Particulate matter and associated metals: a link with neurotoxicity and mental health. Atmosphere 12 (4), 425. https://doi.org/10.3390/atmos12040425.

Raaschou-Nielsen, O., Andersen, Z.J., Hvidberg, M., Jensen, S.S., Ketzel, M., Sørensen, M., Hansen, J., Loft, S., Overvad, K., Tjønneland, A., 2011. Air pollution from traffic and cancer incidence: a Danish cohort study. Environ. Health 10 (1), 67. https://doi.org/10.1186/1476-069X-10-67.

Račić, N., Pehnec, G., Jakovljević, I., Štrukil, Z.S., Mureddu, F., Forsmann, M., Lovrić, M., 2025a. Machine learning analysis of drivers of differences in PAH content between PM1 and PM10 in Zagreb, Croatia. Atmos. Pollut. Res. 16 (7), 102541. https://doi.org/10.1016/j.apr.2025.102541.

Račić, N.,Ružičić, S., Pehnec, G., Jakovljević, I., Sever Štrukil, Z., Rinkovec, J., Žužul, S., Smoljo, I., Zgorelec, Ž., Lovrić, M., 2025b. Linking atmospheric and soil contamination: a comparative study of PAHs and metals in PM10 and surface soil near urban monitoring stations. Toxics 13 (10), 866. https://doi.org/10.3390/toxics13100866.

Račić, N., Ružičić, S., Terzić, T., Pehnec, G., Jakovljević, I., Štrukil, Z.S., Žužul, S., Rinkovec, J., Lovrić, M., 2024. Analyzing the relationship between gas consumption and airborne pollutants: case study of Zagreb, Croatia. Air Qual. Atmos. Health. https://doi.org/10.1007/s11869-024-01655-7.

Rajšić, S., Mijić, Z., Tasić, M., Radenković, M., Joksić, J., 2008. Evaluation of the levels and sources of trace elements in urban particulate matter. Environ. Chem. Lett. 6 (2), 95–100. https://doi.org/10.1007/s10311-007-0115-0.

Ramamoorthy, T., Nath, A., Singh, S., Mathew, S., Pant, A., Sheela, S., Kaur, G., Sathishkumar, K., Mathur, P., 2024. Assessing the global impact of ambient air pollution on cancer incidence and mortality: a comprehensive meta-analysis. JCO Glob. Oncol. 10, e2300427. https://doi.org/10.1200/GO.23.00427.

Ravindra, K., Sokhi, R., Vangrieken, R., 2008. Atmospheric polycyclic aromatic hydrocarbons: source attribution, emission factors and regulation. Atmos. Environ. 42 (13), 2895–2921. https://doi.org/10.1016/j.atmosenv.2007.12.010.

Reshmy, D.S., Swarnalatha, K., Gautam, S., Mathew, B.S., Hegde, P., 2025. Assessing air pollution in a coastal urban setting: contributions of PM_{10}, vehicular emissions, and public health impacts. J. Atmos. Chem. 82 (2), 18. https://doi.org/10.1007/s10874-025-09483-4.

Rubal, Kumar, D., 2018. Evolving differential evolution method with random forest for prediction of air pollution. Procedia Comput. Sci. 132, 824–833. https://doi.org/10.1016/j.procs.2018.05.094.

Safo-Adu, G., Attiogbe, F., Ofosu, F.G., Emahi, I., Palm, L.N.D.M., Enimil, E., 2025. Polycyclic aromatic compounds in ambient PM2.5 in the central region of Ghana: molecular distribution, origin and cancer risk assessment. Atmos. Environ. 343, 120973. https://doi.org/10.1016/j.atmosenv.2024.120973.

Sellami, F., Baati, H., Gautam, S., Azri, C., 2025a. Spatio-temporal patterns and source attribution of dry-deposited particulates in Sfax, Tunisia. Atmos. Pollut. Res., 102771 https://doi.org/10.1016/j.apr.2025.102771.

Sellami, F., Baati, H., Gautam, S., Azri, C., Elaksher, A.F., 2025b. Impact assessment of particulate deposits bound heavy metals on superficial soils along a street canyon at Sfax city (Tunisia): pollution degree and toxicity risks. Earth Sys. Environ. https://doi.org/10.1007/s41748-025-00719-7.

Shukla, K., Kumar, P., Mann, G.S., Khare, M., 2020. Mapping spatial distribution of particulate matter using kriging and inverse distance weighting at supersites of megacity Delhi. Sustain. Cities Soc. 54, 101997. https://doi.org/10.1016/j.scs.2019.101997.

Sielski, J., Kazirόd-Wolski, K., Jóźwiak, M.A., Jóźwiak, M., 2021. The influence of air pollution by PM2.5, PM10 and associated heavy metals on the parameters of out-of-hospital cardiac arrest. Sci. Total Environ. 788, 147541. https://doi.org/10.1016/j.scitotenv.2021.147541.

Šimić, I., Lovrić, M., Godec, R., Kröll, M., Bešlić, I., 2020. Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon. Environ. Pollut. 263, 114587. https://doi.org/10.1016/j.envpol.2020.114587.

Šišović, A., Škrbec, A., Vadjić, V., Kalinić, N., Hršak, J., 2002. PAH levels and profiles in the suspended particulate matter in Zagreb through four seasons. Environ. Monit. Assess. 74 (3), 217–224. https://doi.org/10.1023/A:1014228321207.

Šišović, A., Vadjić, Z., Šega, K., Bešlić, I., Vadjić, V., 2005. Comparison between PAH mass concentrations measured in PM10 and Pm2.5 particle fractions. Bull. Environ. Contam. Toxicol. 75 (1), 121–126. https://doi.org/10.1007/s00128-005-0727-7.

Sørensen, M., Autrup, H., Møller, P., Hertel, O., Jensen, S.S., Vinzents, P., Knudsen, L.E., Loft, S., 2003. Linking exposure to environmental pollutants with biological effects. Mutat. Res., Rev. Mutat. Res. 544 (2–3), 255–271. https://doi.org/10.1016/j.mrrev.2003.06.010.

Stading, R., Gastelum, G., Chu, C., Jiang, W., Moorthy, B., 2021. Molecular mechanisms of pulmonary carcinogenesis by polycyclic aromatic hydrocarbons (PAHs): implications for human lung cancer. Semin. Cancer Biol. 76, 3–16. https://doi.org/10.1016/j.semcancer.2021.07.001.

Suel, E., Sorek-Hamer, M., Moise, I., Von Pohle, M., Sahasrabhojanee, A., Asanjan, A.A., Arku, R.E., Alli, A.S., Barratt, B., Clark, S.N., Middel, A., Deardorff, E., Lingenfelter, V., Oza, N.C., Yadav, N., Ezzati, M., Brauer, M., 2022. What you see is what you breathe? Estimating air pollution spatial variation using street-level imagery. Remote Sens. 14 (14), 3429. https://doi.org/10.3390/rs14143429.

Thomaidis, N.S., Bakeas, E.B., Siskos, P.A., 2003. Characterization of lead, cadmium, arsenic and nickel in PM2.5 particles in the Athens atmosphere, Greece. Chemosphere 52 (6), 959–966. https://doi.org/10.1016/S0045-6535(03)00295-9.

Uzoekwe, S.A., Izah, S.C., Aigberua, A.O., 2021. Environmental and human health risk of heavy metals in atmospheric particulate matter (PM10) around gas flaring vicinity in Bayelsa state, Nigeria. Toxicol. Environ. Health Sci. 13 (4), 323–335. https://doi.org/10.1007/s13530-021-00085-7.

Verma, P.K., Sah, D., Dubey, J., Kumari, K.M., Lakhani, A., 2021. Mutagenic and cancer risk estimation of particulate bound polycyclic aromatic hydrocarbons from the emission of different biomass fuels. Chem. Res. Toxicol. 34 (3), 743–753. https://doi.org/10.1021/acs.chemrestox.0c00378.

Wang, H., Dong, J., Long, J., Jiang, J., Lin, C., 2024. Characteristics, sources analysis, and risk assessment of polycyclic aromatic hydrocarbon contamination in surface sediments surrounding tourist island. Mar. Pollut. Bull. 206, 116735. https://doi.org/10.1016/j.marpolbul.2024.116735.

Wei, W., Wu, B.-J., Wu, Y., Tong, Z.-T., Zhong, F., Hu, C.-Y., 2021. Association between long-term ambient air pollution exposure and the risk of breast cancer: a systematic review and meta-analysis. Environ. Sci. Pollut. Control Ser. 28 (44), 63278–63296. https://doi.org/10.1007/s11356-021-14903-5.

Wilcke, W., 2000. SYNOPSIS polycyclic aromatic hydrocarbons (PAHs) in soil—A review. J. Plant Nutr. Soil Sci. 163 (3), 229–248. https://doi.org/10.1002/1522-2624(200006)163:3<229::AID-JPLN229>3.0.CO;2-6.

Wu, Y., Li, G., An, T., 2022. Toxic metals in particulate matter and health risks in an E-Waste dismantling park and its surrounding areas: analysis of three PM size groups. Int. J. Environ. Res. Publ. Health 19 (22). https://doi.org/10.3390/ijerph192215383.

Yang, L., Zhang, H., Zhang, X., Xing, W., Wang, Y., Bai, P., Zhang, L., Hayakawa, K., Toriba, A., Tang, N., 2021. Exposure to atmospheric particulate matter-bound polycyclic aromatic hydrocarbons and their health effects: a review. Int. J. Environ. Res. Publ. Health 18 (4), 2177. https://doi.org/10.3390/ijerph18042177.

Yebra-Pimentel, I., Fernández-González, R., Martínez-Carballo, E., Simal-Gándara, J., 2015. A critical review about the health risk assessment of PAHs and their metabolites in foods. Crit. Rev. Food Sci. Nutr. 55 (10), 1383–1405. https://doi.org/10.1080/10408398.2012.697497.

Zhang, L., Tian, X., Zhao, Y., Liu, L., Li, Z., Tao, L., Wang, X., Guo, X., Luo, Y., 2021. Application of nonlinear land use regression models for ambient air pollutants and air quality index. Atmos. Pollut. Res. 12 (10), 101186. https://doi.org/10.1016/j.apr.2021.101186.

Zhang, W., Zhang, S., Wan, C., Yue, D., Ye, Y., Wang, X., 2008. Source diagnostics of polycyclic aromatic hydrocarbons in urban road runoff, dust, rain and canopy throughfall. Environ. Pollut. 153 (3), 594–601. https://doi.org/10.1016/j.envpol.2007.09.004.

5.4 Publication III

- *Račić, N.; Ružičić, S.; Pehnec, G.; Jakovljević, I.; Sever Štrukil, Z.; Rinkovec, J.; Žužul, S.; Smoljo, I.; Zgorelec, Ž.; Lovrić, M. (2025) Linking Atmospheric and Soil Contamination: A Comparative Study of PAHs and Metals in PM10 and Surface Soil near Urban Monitoring Stations. Toxics 2025, 13, 866. https://doi.org/10.3390/toxics13100866*

toxics

MDPI

Article

Linking Atmospheric and Soil Contamination: A Comparative Study of PAHs and Metals in PM_{10} and Surface Soil near Urban Monitoring Stations

Nikolina Račić [1,2], Stanko Ružičić [3], Gordana Pehnec [1], Ivana Jakovljević [1], Zdravka Sever Štrukil [1], Jasmina Rinkovec [1], Silva Žužul [1], Iva Smoljo [1], Željka Zgorelec [4] and Mario Lovrić [2,5,6,*]

1 Institute for Medical Research and Occupational Health, 10000 Zagreb, Croatia; nracic@imi.hr (N.R.); gpehnec@imi.hr (G.P.); ijakovljevic@imi.hr (I.J.); zsever@imi.hr (Z.S.Š.); jrinkovec@imi.hr (J.R.); szuzul@imi.hr (S.Ž.); ismoljo@imi.hr (I.S.)
2 The Lisbon Council, 1040 Brussels, Belgium
3 Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, 10000 Zagreb, Croatia; stanko.ruzicic@rgn.unizg.hr
4 Faculty of Agriculture, University of Zagreb, 10000 Zagreb, Croatia; zzgorelec@agr.hr
5 Institute for Anthropological Research, 10000 Zagreb, Croatia
6 Faculty of Food Technology Osijek, Josip Juraj Strossmayer University of Osijek, 31000 Osijek, Croatia
* Correspondence: mario.lovric@inantro.hr

Abstract

Understanding how atmospheric pollutants interact with soil pollution is essential for assessing long-term environmental and human health risks. This study compares concentrations of polycyclic aromatic hydrocarbons (PAHs) and potentially toxic elements (PTEs) in PM_{10} and surface soil near air quality monitoring stations in Zagreb, Croatia. While previous work identified primary emission sources affecting PM_{10} composition in the area, this study extends the analysis to investigate potential pollutant transfer and accumulation in soils. Multivariate statistical tools, including correlation analysis and principal component analysis (PCA), were employed to gain a deeper understanding of the sources and behavior of pollutants. Results reveal significant correlations between air and soil concentrations for several PTEs and PAHs, particularly when air pollutant data are averaged over extended periods (up to 6 months), indicating cumulative deposition effects. Σ11PAH concentrations in soils ranged from 1.2 to 524 μg/g, while mean BaP in PM_{10} was 2.2 ng/m^3 at traffic-affected stations. Strong positive air–soil correlations were found for Pb and Cu, whereas PAH associations strengthened at longer averaging windows (3–6 months), especially at 10 cm depth. Seasonal variations were observed, with stronger associations in autumn, reflecting intensified emissions and atmospheric conditions that facilitate pollutant transfer. PCA identified similar pollutant groupings in both air and soil matrices, suggesting familiar sources such as traffic emissions, industrial activities, and residential heating. The integrated PCA approach, which jointly analyzed air and soil pollutants, showed coherent behaviour for heavier PAHs and several PTEs (e.g., Pb, Cu), as well as divergence in more volatile or mobile species (e.g., Flu, Zn). Spatial differences among monitoring sites show localized influences on pollutant accumulation. Furthermore, this work demonstrates the value of coordinated air–soil monitoring in urban environments and provides an understanding of pollutant distributions across different components of the environment.

Keywords: air pollution; metals; PAHs; soil; particulate matter

Academic Editor: Rafael Clemente

Received: 12 September 2025
Revised: 8 October 2025
Accepted: 10 October 2025
Published: 12 October 2025

Citation: Račić, N.; Ružičić, S.; Pehnec, G.; Jakovljević, I.; Sever Štrukil, Z.; Rinkovec, J.; Žužul, S.; Smoljo, I.; Zgorelec, Ž.; Lovrić, M. Linking Atmospheric and Soil Contamination: A Comparative Study of PAHs and Metals in PM_{10} and Surface Soil near Urban Monitoring Stations. *Toxics* **2025**, *13*, 866. https://doi.org/10.3390/toxics13100866

1. Introduction

Environmental pollution with polycyclic aromatic hydrocarbons (PAHs) and potentially toxic elements (PTEs) has been widely studied, particularly within individual environmental components such as air, water, or soil [1–7]. Numerous studies have focused on characterizing the presence, sources, and behavior of these pollutants in either the atmospheric or terrestrial environment. However, comparative analyses and correlations between compartments, particularly the complex relationships through which airborne pollutants are deposited into soil, remain relatively underexplored [1,8,9].

PAHs and PTEs are among the most prevalent pollutants found in urban air, primarily associated with particulate matter (PM), especially the PM_{10} fraction (particles with aerodynamic diameters ≤10 µm) [10–12]. PAHs are organic compounds composed of multiple fused aromatic rings, commonly produced during incomplete combustion of organic matter [10,13]. Major anthropogenic sources of atmospheric PAHs include motor vehicle exhaust, residential wood or coal burning, industrial processes, waste incineration, and, to a lesser extent, natural sources such as wildfires [14–17].

PTEs, including lead (Pb), cadmium (Cd), zinc (Zn), copper (Cu), arsenic (As), and nickel (Ni), can enter the atmosphere through similar pathways. In urban environments, these metals are typically emitted from traffic-related sources (e.g., fuel combustion, brake and tire wear), industrial activities, and the resuspension of contaminated dust from roads or construction sites [16,18–21].

The health impacts of PAHs and PTEs in ambient air are well established, with exposure linked to severe health conditions, including cancer, cardiovascular disease, and developmental disorders [5,22]. Many PAHs, such as benzo[a]pyrene (BaP), are classified as probable or known human carcinogens and have been associated with genotoxicity, oxidative stress, and respiratory diseases [23,24]. Chronic exposure to PM-bound metals is connected to a range of health effects, including neurotoxicity, nephrotoxicity, cardiovascular dysfunction, and cancers [25–27]. Due to their ability to penetrate deep into the lungs when adsorbed onto fine particles, these pollutants are considered a major public health concern, particularly in densely populated urban areas. Airborne particles serve as a transport medium for both PAHs and metals, facilitating their atmospheric dispersion and eventual deposition onto terrestrial surfaces through both wet and dry deposition processes [28–30]. While atmospheric monitoring provides essential information on current exposure levels and regulatory compliance, it captures only part of the environmental picture, the pollutants in active circulation. Once deposited from the atmosphere, PAHs and metals can accumulate in soils, where different dynamics govern their environmental behavior compared to when they are in the air. Soils can act as both sinks and sources of pollutants, depending on the characteristics of the pollutant, land use, and environmental conditions. PAHs in soil are often adsorbed to organic matter and exhibit slow degradation rates, particularly under anaerobic or cold conditions [6,31,32]. As a result, they can persist in the environment for extended periods and may become a source of contamination for groundwater, plants, and biota. PTEs, in contrast, are non-degradable and tend to persist indefinitely in soils [33,34]. Their mobility and bioavailability depend on several factors, including soil pH, texture, organic matter content, and redox conditions [35,36]. In contaminated soils, metals may pose risks to human health through direct contact, inhalation of resuspended dust, or uptake by crops and subsequent ingestion [35]. This highlights the role of soil as a secondary source of airborne pollutants, capable of reintroducing previously deposited metals and associated particulates back into the atmosphere [37,38]. In urban areas with high pedestrian activity, traffic, and ongoing construction, such re-emission pathways are especially relevant. Additionally, uptake by crops or homegrown vegetables in gardens near roads or industrial areas presents another route of chronic exposure through ingestion [39,40]. These mechanisms explain the importance of monitoring not just atmospheric emissions but also soil pollution and its potential to prolong or amplify human exposure, even after primary emission sources are reduced.

Although many studies have addressed the sources, concentrations, and health effects of PAHs and metals in air or soil separately, comparative investigations that assess their concentrations in both environments simultaneously are still limited. Very few studies attempt to evaluate whether the pollutant profiles in the air are reflected in the underlying soils, or whether the same emission sources contribute similarly to both media. Additionally, the degree to which current air pollution contributes to ongoing soil contamination through atmospheric deposition is not well understood. Soil sampling in the vicinity of air quality monitoring stations provides a unique opportunity to assess these relationships. If the same pollutants measured in PM_{10} are found in elevated concentrations in nearby soils, and if patterns align spatially or statistically, this can indicate either direct deposition pathways or shared source signatures. Such analyses provide valuable insights into pollutant transport and accumulation processes, as well as long-term environmental exposure risks. Croatian soils are characterized by a diverse geochemical background due to the country's complex geological structure. According to the national Geochemical Atlas [21], concentrations of trace metals, such as Fe, Mn, Zn, Pb, and Cu, vary significantly across different regions, reflecting both natural lithological variation and anthropogenic influences. For example, elevated Zn and Pb levels in urban soils have been linked to historic industrial activities and traffic emissions, while As and Fe concentrations are more strongly controlled by geological substrates and soil-forming processes [21].

Building on previous work that identified primary sources of PAHs and metals in PM_{10} in the urban environment of Zagreb [16], this study expands the focus to include the soil component. The main objectives are

- To determine the concentrations and distribution of selected PAHs and PTEs in PM10 and surface soil samples collected near urban air quality monitoring stations over three years.
- To investigate spatial and vertical variability in soil pollutant concentrations (0–5 cm vs. 5–10 cm) and explore seasonal differences in pollutant accumulation.
- To compare pollutant profiles across air and soil compartments and assess the extent of their correlation, to evaluate atmospheric deposition as a potential pathway for soil contamination.
- To apply descriptive and multivariate statistical techniques to identify common pollution sources and better understand the inter-compartmental behavior of pollutants.

The idea of this study is to advance the understanding of pollutant behavior by integrating atmospheric and soil data, thereby contributing to a more comprehensive assessment of environmental contamination processes in urban environments. To the best of our knowledge, this is the first study conducted in this region to examine and compare PAHs and metal concentrations in both air (PM_{10}) and soil, exploring their potential interactions and deposition trends.

2. Materials and Methods

2.1. Study Area and Pollutant Characterization

The city of Zagreb, situated in the northwestern part of Croatia, represents the country's most significant urban, economic, and administrative centre. With nearly 800,000 residents, the city represents a typical urban environment, including a mix of residential, industrial, and transportation-related emission sources [41]. Seasonal variations in heating demand, connected with urban traffic, contribute to fluctuating levels of air pollutants, notably PM, PAHs, and PTEs. To capture spatial variability in urban air pollution and its potential influence on surrounding soils, three sites were selected within the city of Zagreb: Ksaverska cesta (IMI), Center (CEN), and Siget (SIG) (Figure 1).

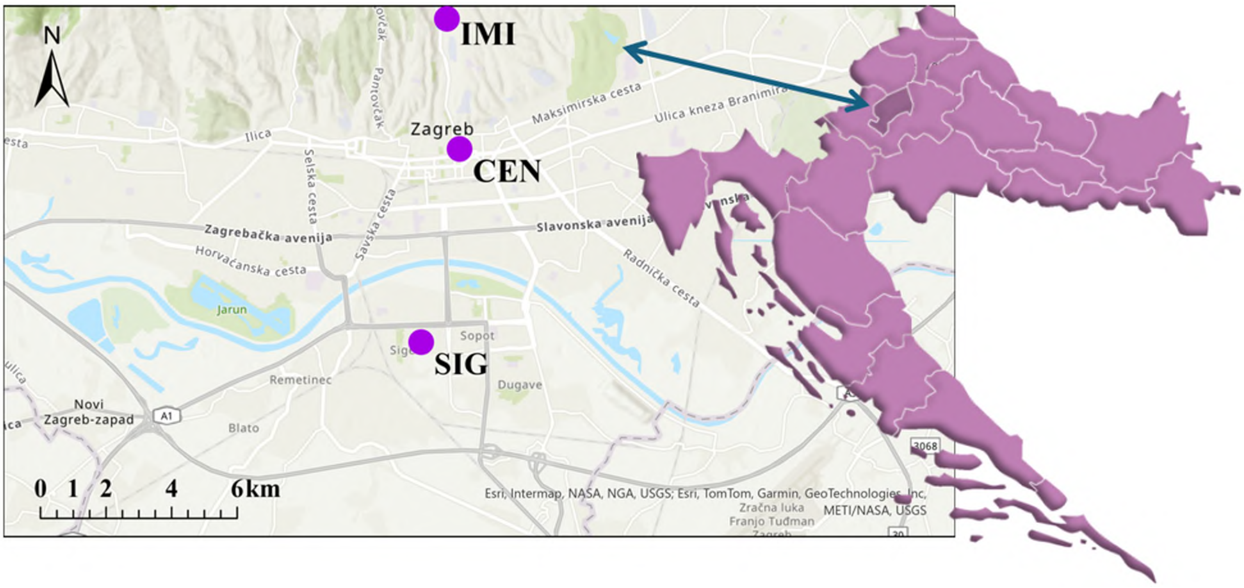


Figure 1. Site locations. Figure was made using ArcGIS Pro (version 3.2.0) and Adobe Illustrator (version 29.5.1). IMI (altitude 116 m a.s.l., 45°50′7′′ N, longitude 15°58′42′′ E); SIG (altitude 116 m a.s.l., latitude 45°46′25′′ N, longitude 15°59′04′′ E), CEN (altitude 113 m a.s.l., 45°48′37′′ N, longitude 15°59′4′′ E).

Site selection was based on the type of monitoring station (urban background or traffic-influenced), their geographic distribution across the city, and the availability of long-term pollutant concentration data. At all three locations, there are air quality monitoring stations within Zagreb's official air quality monitoring network, which is operated with support from the City Office for Economy, Energy, and Environmental Protection. At the IMI site, there is Ksaverska cesta monitoring station, which is categorized as an urban background site and is located in the northern residential zone of Zagreb. The station is located approximately 20 m from a road with a moderate traffic volume and is surrounded predominantly by low-rise family housing units. Heating in this area is primarily based on gas, with supplemental use of wood and oil, particularly during the winter season. The SIG site, by contrast, is located in a densely populated southern neighbourhood. The air quality monitoring station SIG is situated approximately 4 m above ground and roughly 30 m from a main road with intense traffic. The area is characterized by high-rise buildings, mostly connected to the central heating system (approximately 5 km east-southeast of SIG). Still, it is also influenced by dominant northerly winds carrying emissions from the city center and residential heating systems using various fuel types. Soil samples were collected in the immediate vicinity of the SIG station. The CEN station (air quality monitoring station Đorđićeva ulica) is positioned near the urban center, with proximity to heavily trafficked roads.

For the CEN site, PAH concentrations were not directly available at the station; therefore, data from the nearest national network air quality monitoring station (Zagreb-1) in proximity were used to represent PAH levels. The built environment in this sector consists of multi-story buildings, most of which are heated by natural gas. This location represents one of the historically most densely populated and traffic-affected parts of the city. The wider Zagreb area is situated within the southwestern part of the Pannonian Basin, characterized by Quaternary alluvial deposits overlying Neogene marls, clays, sands, and gravel layers, with local occurrences of Miocene limestones and dolomites [42]. These unconsolidated alluvial and fluvial sediments dominate the upper soil horizons across the city, providing a silicate-rich geogenic background for elements such as Fe, Mn, and Cr. Variations in soil texture and mineralogy across the urban area reflect both natural substrate heterogeneity and anthropogenic modifications due to construction, filling, and land use. This geological context is crucial for understanding the distinction between lithogenic and anthropogenic sources of trace elements in the collected soil samples [43].

At the selected monitoring stations, PM_{10} was measured daily in accordance with EU standardized reference methodologies (CEN: EN 12341:2023, EN 14902:2005 and AC:2006, EN 15549:2008, and CEN/TS 16645:2014) [44–48]. For this study, the period from 1 January 2018 to 31 December 2020 was chosen. During this time, continuous 24-h PM_{10} sampling was conducted using low-volume samplers (Sven Leckel, ~55 m^3/day), with particles collected on quartz or membrane filters depending on the analytical requirements. The chemical composition of PM_{10} was assessed through analysis of selected trace metals and PAHs, including those not included in routine monitoring. Targeted metals included arsenic (As), cadmium (Cd), lead (Pb), manganese (Mn), iron (Fe), copper (Cu), and zinc (Zn). At the same time, the PAH fraction comprised benzo[a]pyrene (BaP), benzo[a]anthracene (BaA), benzo[ghi]perylene (BghiP), dibenzo[ah]anthracene (DahA), chrysene (Chry), benzo[k]fluoranthene (BkF), benzo[b]fluoranthene (BbF), benzo[j]fluoranthene (BjF), fluoranthene (Flu), pyrene (Pyr), and indeno[1,2,3-cd]pyren (IP). PAH compounds were extracted from quartz filters using ultrasonic-assisted extraction with a cyclohexane-toluene solvent mixture. Following centrifugation and evaporation, the extracts were reconstituted in acetonitrile and analyzed via high-performance liquid chromatography (HPLC) using an Agilent 1260 Infinity system (Agilent Technology, Santa Clara, CA, USA) equipped with a fluorescence detector, according to protocols adapted from Jakovljević et al. [49]. Metal concentrations were determined after microwave-assisted acid digestion (Ultraclave IV, Milestone Srl) of the filters, followed by analysis using inductively coupled plasma mass spectrometry (ICP-MS; Agilent 7500cx, Agilent Technology, Santa Clara, CA, USA), as described in Beslic et al. [18]. The monitoring stations and analysis were previously described in [50–54].

In parallel with air monitoring activities, surface soil samples were collected near the three air quality monitoring stations over three years (2018–2020). Sampling was carried out twice annually, in spring and autumn, to capture potential seasonal variations in pollutant deposition and accumulation. At each site, soil was sampled at two depths: 0–5 cm and 5–10 cm, targeting layers most influenced by atmospheric deposition. Composite samples were formed by combining multiple subsamples within a defined radius of each station to ensure spatial representativeness. Five individual subsamples were collected at each depth and then homogenized to form one composite sample per depth and season. This procedure was repeated at all three stations during spring and autumn campaigns, yielding a total of 36 composite samples (3 stations $\times$ 2 depths $\times$ 2 seasons $\times$ 3 years). The collected soil was air-dried, homogenized, and before chemical analysis, the samples were finely ground using a Retsch RS 200 vibratory disc mill (Retsch GmbH, Haan, Germany) to achieve uniform particle size and analytical consistency. Metal concentrations were then determined using a portable X-ray fluorescence (XRF) analyzer (HITACHI X-MET8000 Expert GEO, Hitachi High-Tech Analytical Science, Oxford, UK), which enables rapid, non-destructive elemental analysis of prepared soil samples. An Accelerated Solvent Extraction system (ASE 350, Dionex, Thermo Fisher Scientific, Sunnyvale, CA, USA) was used to extract PAHs from soil samples. Extractions were performed at 1500 psi and 125 °C, using a flush volume of 70% of the extraction cell volume. A 1:1 (v/v) mixture of n-hexane and acetone was used as the extraction solvent. The final extracts were concentrated to 1 mL, and 10 ng of the internal standard, perylene-d_{12}, was added to each sample. Quantification of 11 PAHs was carried out using gas chromatography coupled with tandem mass spectrometry (GC–MS/MS; Agilent 7890B–7000C, Agilent Technology, Santa Clara, CA, USA) in multiple reaction monitoring (MRM) mode. The GC was equipped with a DB-EUPAH capillary column (20 m × 180 μm i.d., 0.14 μm film thickness), using helium as the carrier gas.

All analytical procedures were performed under strict QA/QC protocols. For PM_{10} samples, calibration was carried out using commercial standards, and certified reference materials (e.g., NIST SRM 1649a for PAHs, ERM-CZ120 for metals) were analyzed in parallel. Calibration curve linearity, recovery, and repeatability were checked routinely, with recovery rates for both PAHs and metals falling within 85–115%. The same QA/QC principles were applied to soil PAH analysis. Calibration standards and surrogate spikes were used to assess method performance, with recoveries within accepted ranges. All concentrations were corrected for recovery. Procedural, laboratory, and field blanks were included to monitor potential contamination, and replicate analyses confirmed reproducibility.

2.2. Data Processing and Statistical Analysis

The dataset for this study comprises daily records of pollution data from three air quality monitoring stations as well as data from soil samples. The analysis was conducted using the Python programming language (www.python.org, v3.10). Air pollutant concentrations were aggregated by calculating median values over defined temporal windows of 1, 2, and 3 months before each soil sampling season to capture potential lagged deposition effects. Soil pollutant concentrations were transformed to a long format and harmonized with air data by matching pollutant type, station, season, and year. Spearman rank correlation coefficients were then computed for each pollutant to assess the relationships between airborne and soil-bound concentrations, focusing on the matched sampling periods and locations. This approach enabled the examination of temporal and spatial relationships between air and soil contamination, providing information about pollutant behavior across different environmental media.

To explore pollutant sources and trends, principal component analysis (PCA) was conducted separately on the standardized datasets of air and soil pollutants. PCA helped identify groups of pollutants with common variance patterns, which can be interpreted as indicative of shared sources or similar environmental behavior in different environmental matrices. To explore the relationships and grouping between individual air and soil pollutants, we performed a PCA using concentration data from paired air and topsoil (5 cm depth) samples. Before PCA, non-numeric variables and metadata columns (e.g., station, year, medium) were excluded. To compare pollutant behavior across media, the dataset was transposed so that each pollutant became a separate row (observation) and each sample became a variable. The transposed data matrix was then standardized using z-score normalization, ensuring each pollutant was centered and scaled to unit variance. PCA was performed on the standardized matrix using the scikit-learn PCA implementation. PCA results were visualized in a two-dimensional score plot, with each point representing a pollutant-medium pair. Because soils integrate inputs over longer periods than instantaneous air concentrations, we paired each soil campaign with rolling 1-, 3-, and 6-month medians of prior airborne concentrations to capture seasonal-scale deposition signals. The 0–5 cm and 5–10 cm depths were analyzed to differentiate recent vs. historical accumulation signatures.

3. Results and Discussion

3.1. General Pollution Levels in Air and Soil

Across the three monitoring stations in Zagreb, PAH and metal concentrations in PM_{10} showed spatial variability. For the overall period, SIG exhibited the highest mean concentrations for all PAHs and metals, including BaP (2.15 ng/m^3), BghiP (2.14 ng/m^3), and BbF (2.44 ng/m^3). IMI and CEN stations generally showed lower PAH levels, although CEN had notable levels of BaP (1.25 ng/m^3) and BghiP (1.26 ng/m^3), suggesting the impact of urban central emissions. Similar trends were observed for PTEs: SIG reported the highest concentrations of Fe (543.51 ng/m^3), Zn (28.14 ng/m^3), and Cu (17.37 ng/m^3), which are likely linked to traffic-related sources, such as brake and tire wear. In contrast, the IMI site showed the lowest levels for most metals, possibly due to lower direct emissions or dispersion effects.

Based on the dataset of 11 PAH compounds measured in topsoil (0–5 cm) and subsoil (5–10 cm) samples collected in spring and autumn of 2018–2020 at three stations, the results show notable spatial, seasonal, and vertical trends (Figure S1 in Supplementary Materials). To our knowledge, these are the first systematically collected and analysed data on PAHs in soils from Zagreb and among the few such datasets available for Croatia. The Andrija Štampar Teaching Institute of Public Health occasionally conducts measurements of PAHs and metals in Zagreb soils as part of its environmental monitoring activities, with results accessible via the City of Zagreb's Eco Map portal (https://ekokartazagreb.stampar.hr/, accessed on 10 August 2025). However, this monitoring is occasional rather than systematic, is conducted at varying locations without repeated sampling at the same sites, and the data is not regularly processed or published in the scientific literature, serving instead as public information on whether guideline values are met. In this study, Σ11PAH concentrations in Zagreb urban soils ranged from 1.23 to 524,098 ng/g, with BaP between 0.05 and 20,375 ng/g. These values align with the global urban soil ranges summarized by Wilcke [55], where background urban sites typically contain 10^2–10^5 ng/g ΣPAH, and industrial hotspots can exceed 10^6 ng/g. Similarly, Li et al. [56] reported urban soil ΣPAHs spanning <100 to >106 ng/g, with traffic and residential combustion as dominant sources. Compared to regional data, Zagreb soils overlap with the lower to mid-range of values from Novi Sad, Serbia (180–24,500 ng/g Σ16PAH [57]). Still, they are substantially below the high contamination levels found in Glasgow, Torino, and Ljubljana [6]. Overall, the Zagreb results indicate moderate urban PAH contamination, consistent with the ranges reported for cities influenced primarily by traffic and domestic heating emissions, rather than by heavy industrial activity.

Across sites, Σ11PAH concentrations showed both seasonal and depth-related differences (Figure 2). At CEN, values were higher at 10 cm than at 5 cm for most seasons. At IMI, seasonal means were more balanced, with slightly higher values at 10 cm in both spring and autumn, but no major depth effect was observed. At SIG, concentrations remained uniformly low regardless of depth or season, with only minor variation between 5 cm and 10 cm layers.

The IMI site exhibited moderately high PAH concentrations, particularly in spring 2018 at a 5 cm depth (Sum11PAH = 46.448 μg/g), with a dominance of BaA (10.433 μg/g), Chry (8.751 μg/g), and BbF (8.394 μg/g). Over time, concentrations at IMI declined, particularly by the spring of 2020, when sum11PAH dropped to 17.884 μg/g at 5 cm. The CEN station shows the highest PAH contamination, particularly in spring 2020, with elevated values at 10 cm depth (Sum11PAH = 524.098 μg/g), driven by extremely high BaA (142.913 μg/g), Pyr (94.192 μg/g), Chry (74.555 μg/g), and BbF (72.905 μg/g). Even at 5 cm depth, spring 2020 values reached 105.714 μg/g, again highlighting BbF, BaA, and Chry as key contributors.

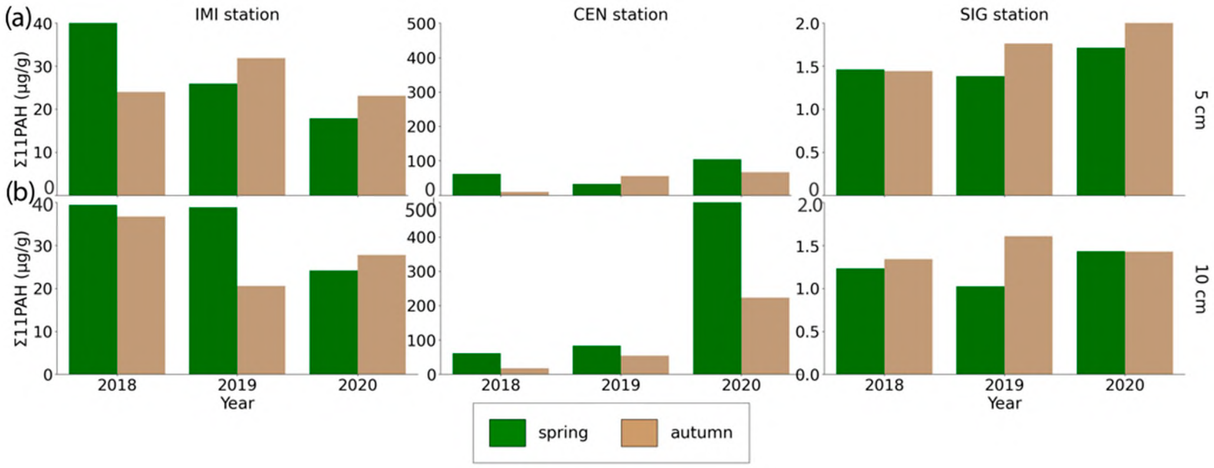


Figure 2. Annual variation (*x*-axis, 2018–2020) of sum PAH concentrations in soil at depths (**a**) 0–5 cm; (**b**) 5–10 cm at monitoring sites IMI, CEN, and SIG, in μg/g, mean for both seasons.

The exceptionally high PAH concentrations observed at CEN in spring 2020 could reflect a combination of local emission events and environmental disturbances. In particular, the March 2020 Zagreb earthquake and subsequent clean-up activities may have contributed to resuspension of contaminated dust in the city center, while increased residential heating during the unusually cold early spring period may have further elevated emissions [53]. Episodic inputs of this kind are consistent with the observed sharp interannual variability at CEN but not at other sites. Autumn levels were generally lower, but still substantial, e.g., 67.838 μg/g at 5 cm in autumn 2020. In contrast, the SIG station consistently exhibited very low PAH concentrations. For example, spring 2018 values were only 1.502 μg/g at 5 cm and 1.276 μg/g at 10 cm. Even by the 2020 autumn, concentrations remained low at 2.086 μg/g and 1.475 μg/g at 5 and 10 cm, respectively.

In Figure S1 in the Supplementary Materials, seasonal differences are not consistent: in some years and stations, spring values are higher (e.g., IMI 2018, CEN 2020), while in others, autumn is higher or the difference is minimal. Therefore, the overall pattern is better represented by averaging concentrations by season (Figure 3), which reveals a clear decreasing trend at IMI, a pronounced increase at CEN (driven mainly by 2020), and persistently low but slightly increasing levels at the SIG site. This pattern aligns with previous studies, which show that in urban soils, interannual variability and episodic inputs (such as heating/combustion, traffic, and atmospheric deposition) often override a stable seasonal cycle. Winter and spring inputs are frequently elevated but are strongly influenced by local sources and meteorological conditions. The dominance of 4–6-ring PAHs and the observed concentration ranges are comparable to other moderately impacted European cities [6], supporting the interpretation that local sources and atmospheric deposition are the primary drivers of variability. These findings revealed the impact of site-specific characteristics, seasonal atmospheric inputs, and pollutant mobility in explaining PAH distribution in soils.

The concentrations of selected metals (Cu, Zn, As, Cd, and Pb) in soil varied considerably across stations, depths, and sampling periods (Figure 4).

Among all metals, iron (Fe) shows the highest concentrations, with values ranging from 8950 μg/g to 54,215 μg/g, depending on the station and depth. Mn concentrations also showed significant variability, spanning from 212 μg/g to over 1130 μg/g, with generally higher values observed in CEN and IMI stations compared to SIG. Cu and Zn concentrations were highest at SIG, where Zn reached up to 219 μg/g and Cu exceeded 120 μg/g in some samples. Arsenic levels ranged from 4.3 μg/g to approximately 22 μg/g, indicating moderate fluctuations across locations and depths. Cd, while generally present at low concentrations (often below five μg/g), occasionally peaked at 8.5 μg/g, particularly at SIG in 2019. Lead concentrations showed a broad range (from 21 μg/g to over 388 μg/g), with particularly elevated values observed at IMI in 2019 and CEN in 2020. The IMI peak likely reflects dust from demolition and roadworks for the new Institute building near the station, while the CEN peak may be linked to dust resuspension after the March 2020 Zagreb earthquake and post-quake clean-up in the city centre [58]. These values are comparable to, or exceed, those previously reported for Zagreb soils, where Pb ranged 1.5–139 μg/g, Zn 15–277 μg/g, Cu 4–183 μg/g, Mn 79–1282 μg/g, and Fe 5.8–51.8 g/kg, with spatial patterns reflecting both anthropogenic influences (traffic emissions, industrial activities, waste disposal) and lithogenic controls [59]. Other studies in the Zagreb region similarly identified Zn, Pb, Cd, and Cu as markers of urban and industrial contamination, with short-range variability linked to local emission sources and long-range variability reflecting geological background [60]. Elevated levels of several metals in earlier surveys were also associated with traffic, industrial history, and riverine deposition from the Sava [54]. Overall, higher concentrations of several metals (notably Cu, Zn, and Pb) tended to occur in topsoil (5 cm) and during spring campaigns, suggesting possible influences from seasonal deposition and proximity to anthropogenic sources (Figures S2–S4 in Supplementary Materials).

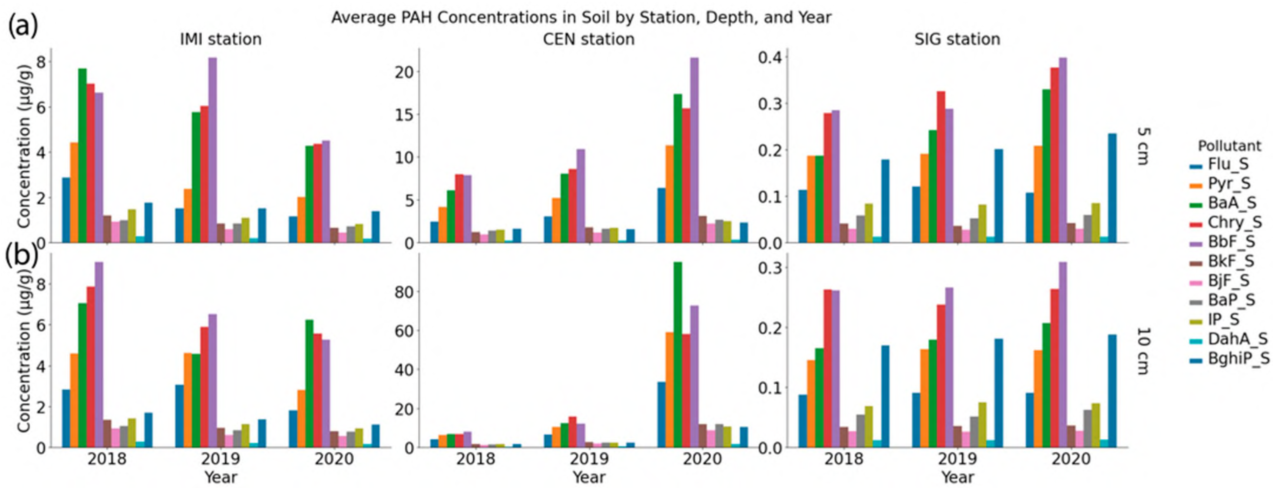


Figure 3. Annual variation (*x*-axis, 2018–2020) of PAH concentrations in soil at depths (**a**) 0–5 cm; (**b**) 5–10 cm at monitoring sites IMI, CEN, and SIG, in μg/g, mean for both seasons.

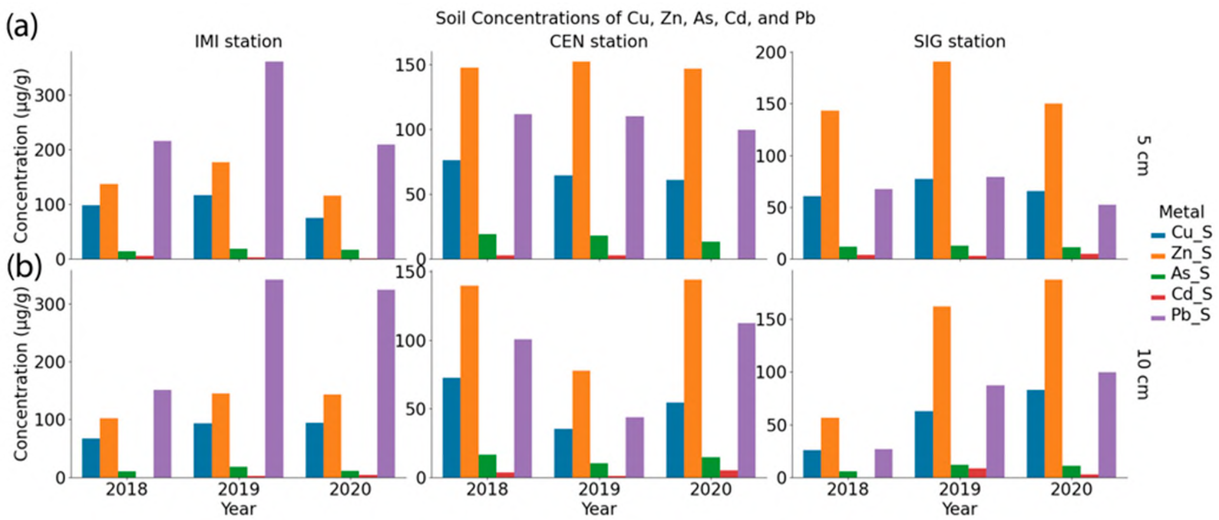


Figure 4. Annual variation (*x*-axis, 2018–2020) of metal concentrations in soil at depths (**a**) 0–5 cm; (**b**) 5–10 cm at monitoring sites IMI, CEN, and SIG, in μg/g, mean for both seasons.

3.2. Relationships Between Air and Soil Pollutants

As a first step, the relationships between air and soil concentrations of selected PAHs and trace metals were investigated by correlating seasonal soil pollutant levels with median air pollutant concentrations for that same season. The heatmap visualization of all pollutants indicated some clustering of metals with positive cross-correlations, while many PAHs exhibited negative or negligible correlations between air and soil phases (Figure 5). PTEs such as Pb, Mn, Fe, As, and Cu showed strong positive correlations between air and soil, suggesting that they may share similar seasonal emission and deposition patterns, potentially resulting from traffic, construction activities, and the resuspension of contaminated dust. However, As and Cd exhibited negative correlations between air and soil concentrations, indicating possible differences in source profiles, atmospheric behavior, or soil retention dynamics. The negative correlations observed for several PAHs between air and soil likely reflect differences in environmental behavior compared to those of trace metals. Many PAHs are semi-volatile and subject to bidirectional exchange between soil and atmosphere, especially under higher temperatures, which can result in re-emission rather than net accumulation. Atmospheric degradation (e.g., photolysis and reactions with OH radicals or ozone) further reduces the fraction available for deposition. In soil, microbial degradation and sorption–desorption dynamics also influence persistence; lighter PAHs (e.g., Flu, Pyr) are more mobile and degrade faster, leading to weaker or even inverse associations with recent airborne levels. Moreover, site-specific soil properties, such as organic matter content, texture, and pH, affect retention, with soils of lower organic carbon content showing reduced PAH binding and greater losses. Together, these processes can explain why metals mostly tend to show positive air–soil correlations, while PAHs often display weaker or negative relationships. This suggests that complex deposition, degradation, or source factors influence pollutant behavior in soil [61]. Comparable findings have been reported in other European urban soil studies, where self-organizing map analyses highlighted the distinct clustering of lithogenic versus traffic-related elements and underscored the complex behavior of PAHs in surface soils [62].

The Spearman correlation analysis revealed an association between air and soil pollutant concentrations at two soil depths (0–5 cm and 5–10 cm, Table S1 in Supplementary Materials). At 0–5 cm depth, the strongest positive correlations were observed for Pb ($R = 0.94$, $p = 0.005$), Cd ($R = 0.88$, $p = 0.020$), and Mn ($R = 0.77$, $p = 0.072$), indicating a close association between their airborne concentrations and accumulation in the surface soil layer. Fe ($R = 0.77$) and Cu ($R = 0.71$) also showed moderately strong positive correlations. At the same time, most PAHs displayed moderate to strong negative correlations (e.g., Flu, DahA), suggesting different environmental behaviors or source–sink dynamics for these compounds. At 5–10 cm depth, Pb ($R = 0.83$, $p = 0.042$), Mn ($R = 0.94$, $p = 0.005$), and Fe ($R = 0.89$, $p = 0.019$) remained strongly positively correlated with their airborne concentrations. For PAHs, negative correlations persisted but were generally weaker than in the surface layer, indicating attenuation of the air–soil linkage with depth, likely due to the reduced influence of recent deposition and the greater dominance of historical accumulation or degradation processes. These findings are consistent with other urban studies, which show trace metal accumulation in topsoil, particularly Pb and Zn, due to legacy emissions and their affinity for organic matter and clays [63,64].

As a next step, the analysis of Spearman correlation coefficients between airborne pollutant concentrations and soil pollutant levels across different averaging windows (1, 3, and 6 months before soil sampling, Table S2 in Supplementary Materials) reveals important information about the temporal dynamics of pollutant deposition and accumulation in soil. The results indicate that for several pollutants, particularly PTEs such as Pb and Cu, correlations with soil concentrations are already high at 1 month and remain consistently strong or slightly decrease with longer averaging windows. Pb, for example, shows robust positive correlations across all windows (R = 0.825 at 5 cm and R = 0.727 at 10 cm for the 1-month average), confirming that soil acts as a cumulative sink for airborne Pb over extended periods. In contrast, most PAHs display weak or inconsistent correlations at 1 month but strengthen markedly over 3–6 months, particularly at 10 cm depth for compounds such as Chry, BaP, BbF, and Flu (R often > 0.6, $p < 0.05$), suggesting that seasonal-scale integration better captures their deposition into subsurface soil layers. The generally weaker correlations at short windows likely reflect the influence of episodic events, seasonal variability, atmospheric degradation, and differences in deposition pathways. These findings support the view that for persistent metals, both short- and long-term airborne levels are relevant to soil contamination, whereas for PAHs, longer averaging periods provide more meaningful soil–air relationships due to slower accumulation and greater atmospheric variability. Soil typically integrates pollutants over time due to limited mobility and slower degradation compared to the atmosphere [65]. Not all pollutants respond similarly to the averaging window length. Some PAHs, such as Pyr, exhibit weak or even slightly negative correlations at shorter windows but improve modestly at longer windows, suggesting that episodic events or seasonal variations may influence atmospheric concentrations, complicating short-term relationships. The weaker correlations at shorter windows may also result from differential atmospheric transport, degradation rates, or variable emission sources. For metals with relatively stable emissions and slow soil dynamics, longer averaging windows yield more meaningful correlations. Conversely, for more volatile or episodic compounds, additional factors beyond simple temporal averaging may better explain soil–air relationships.

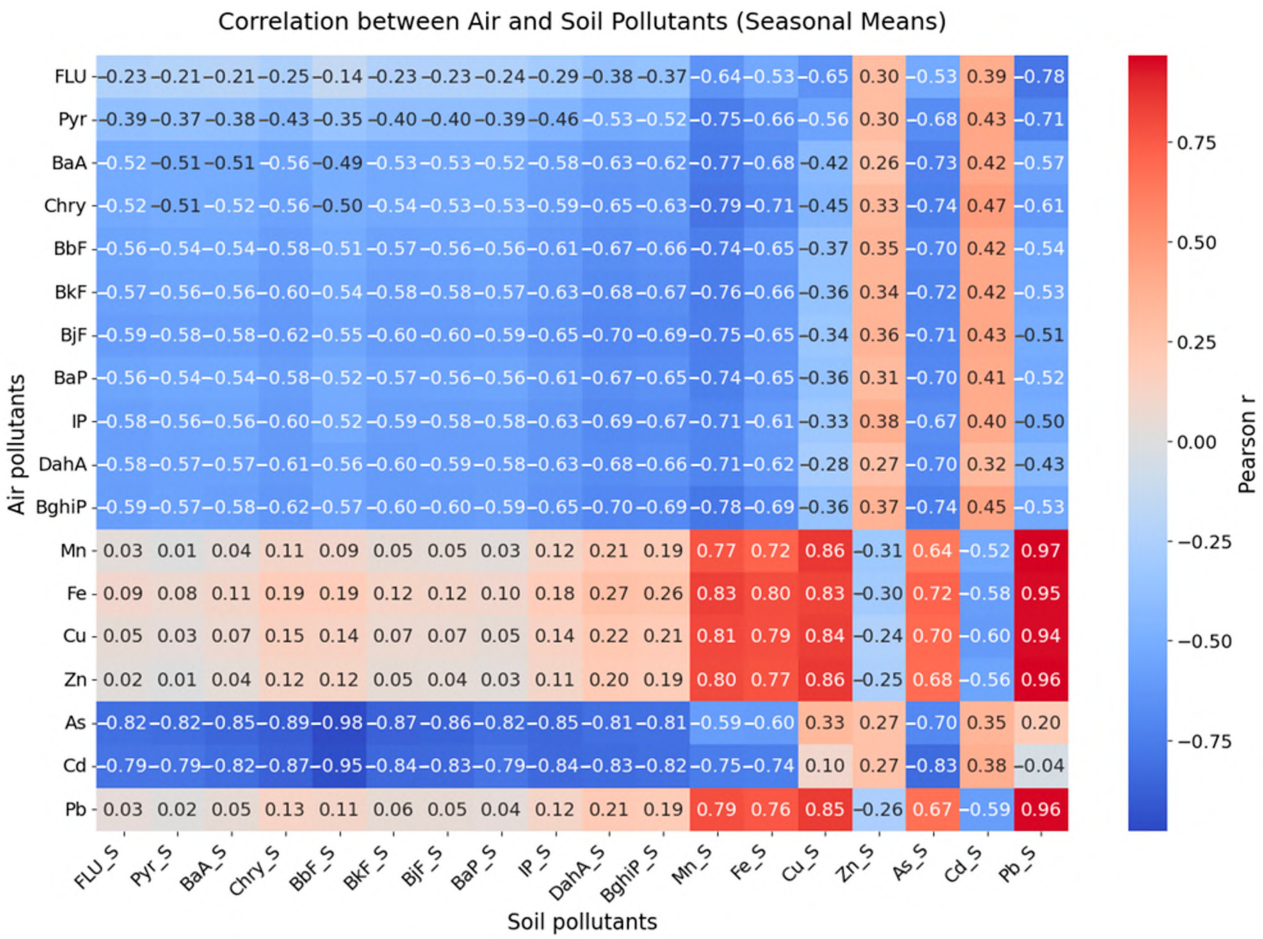


Figure 5. Correlation heatmap between air and soil pollutants (air: seasonal mean concentrations; soil: mean across both seasons and depths).

To explore potential seasonal differences, Spearman correlations were calculated separately for the spring and autumn seasons, for both soil depths (Table S3 in the Supplementary Materials). Using the 6-month window, autumn shows strong positive soil–air correlations for metals, led by Pb (R = 0.900, p = 0.001 at 5 cm; R = 0.767, p = 0.016 at 10 cm) and Cu at 5 cm (R = 0.750, p = 0.020), with Fe and Mn positive but borderline at 10 cm (Fe R = 0.633, p = 0.067; Mn R = 0.650, p = 0.058); Arsenic is strongly negative at 5 cm (R = −0.783, p = 0.013), and many PAHs are negative (e.g., IP R = −0.700, p = 0.036 at 5 cm; R = −0.650, p = 0.058 at 10 cm; BjF R = −0.650, p = 0.058 at 5 cm; BkF R = −0.636, p = 0.066 at 5 cm). In spring, PAHs remain mostly negative at 5 cm, BghiP (R = −0.733, p = 0.025), BaP (R = −0.667, p = 0.050), BkF (R = −0.667, p = 0.050), BjF (R = −0.700, p = 0.036), while metals are mixed: Pb stays positive but weaker (R = 0.650, p = 0.058 at 5 cm; R = 0.467, p = 0.205 at 10 cm), Fe is modest (R = 0.550, p = 0.125 at 5 cm), and Cu is weak (R = 0.300, p = 0.433 at 5 cm). Using a 6-month window, the seasonal analysis lines up with pollutant chemistry. Metals, especially Pb and partly Cu, exhibit a stable positive soil–air relationship, strongest in autumn, reflecting their persistence and strong sorption that drives long-term accumulation into deeper layers. PAHs, being semi-volatile and more reactive, show mostly negative within-season correlations, strongest at 5 cm, due to faster turnover, degradation, and temperature-driven exchange. This explains why earlier non-seasonal positive PAH signals at 10 cm likely came from between-season contrasts rather than true within-season co-variation.

The PCA biplots for both soil (mean for all periods of measurement and both depths) and air (PM_{10}, full dataset) data across the IMI, CEN, and SIG stations show clear groupings of PAHs and metals that suggest common sources and transport pathways (Figure S5 in Supplementary Materials). In soil, PAHs cluster together, likely reflecting combustion or industrial emissions, while metals form a distinct group indicating shared origins such as traffic emissions or geogenic background. The air PCA similarly shows separation between PAHs and metals, with consistent clustering of metals and PAHs across stations, confirming their atmospheric sources are primarily related to combustion processes. The alignment of pollutant groupings between soil and air suggests that atmospheric deposition is the primary mechanism for transferring airborne contaminants to soils. Although the patterns are generally coherent, differences in pollutant loadings and PCA scores across stations indicate that site-specific factors influence pollutant distribution, such as local emission sources or soil properties. Overall, these PCA results support the impact of atmospheric pollution as a soil contaminant and emphasize the value of integrated air and soil monitoring for understanding pollutant sources and dynamics in the region.

To further explore sources, separate PCA analyses were conducted for PAHs and for metals across all stations. These individual biplots (Figures S6 and S7 in the Supplementary Materials) allow more straightforward interpretation of pollutant groupings, highlighting how specific PAHs or metals contribute to variability across sites. The PCA biplots for soil PAHs reveal both common trends and compound-specific variations across the IMI, CEN, and SIG stations. At IMI, most PAHs, such as BbF, BkF, DahA, and IP, cluster tightly, suggesting a common source likely related to combustion. In contrast, BaA and BghiP project more distinctly, indicating slightly different behavior or source influence. At CEN, the PAHs group is even more closely aligned along PC1, especially BaP, DahA, and IP, highlighting a dominant shared source, possibly high-temperature combustion or industrial emissions. Chry and BbF show moderate separation, suggesting some variability in input. Similar source-related PAH patterns have also been reported in previous studies from the Zagreb area [16,50,66].

In contrast, SIG shows the most dispersion among PAHs: BaA, Chry, BghiP, and BbF align with PC2, while Flu, Pyr, DahA, and BkF deviate downward, implying mixed sources or altered retention in soil, potentially due to local soil or emission characteristics. These differences suggest that while most PAHs are co-emitted and behave similarly, certain compounds exhibit site-specific behaviors likely driven by localized conditions or additional sources. The PCA biplots for metals in soil show consistent grouping patterns across stations, with some metal-specific distinctions. At IMI, Zn, Cu, Fe, and Pb cluster along PC1, suggesting a shared origin, likely from traffic or urban emissions. In contrast, As and Mn show stronger projections on PC2, indicating different geochemical behaviors or additional sources. At CEN, metals are tightly grouped, especially Pb and Zn, indicating a dominant, likely anthropogenic source; As, however, separates clearly, suggesting a possible geogenic origin. At SIG, most metals form a compact group along PC1, confirming common urban sources, while As and Pb display different orientations, possibly reflecting site-specific inputs or soil retention differences. Overall, the patterns suggest that while most metals in soil originate from shared sources, As and, to some extent, Pb exhibit unique spatial behavior across stations.

The PCA results for both PAHs and metals in air show consistent groupings indicative of shared sources. High molecular weight PAHs such as BaP, BbF, BkF, DahA, and BghiP cluster tightly at all stations, suggesting dominant sources from combustion processes (e.g., traffic, heating), while lighter PAHs (Flu, Pyr) show separation along PC2, especially at CEN, indicating variable atmospheric behavior [16]. In contrast, the PCA of metals shows more pronounced inter-station differences. At IMI, Mn and Fe are clearly separated from other metals along PC2, indicating a more substantial contribution from crustal or industrial sources. In contrast, Pb, Zn, Cd, and Cu cluster together along PC1, suggesting a connection to traffic-related emissions. At CEN, metals are more tightly grouped, but As and Cu stand apart with strong PC2 loading, likely reflecting industrial input or long-range transport. SIG exhibits a similar grouping structure to CEN, but with more dispersed (wider) loading vectors, indicating that the same types of sources are present, albeit with greater variation in their relative contributions or influence. These results align with earlier studies in Zagreb, which identified vehicular emissions and residential heating as the leading contributors to airborne PAHs and metals [16,51]. They are also consistent with broader European evidence, which shows that Ni, Cr, and Co are primarily controlled by geological background. In contrast, Cd, Zn, and Cu tend to reflect anthropogenic enrichment and higher mobility [66,67].

To explore the relationships between pollutant concentrations in air and soil, a modified PCA was performed on a transposed dataset, where pollutants were treated as observations. In this approach, the average concentrations of PAHs and trace metals in both air and soil were used to construct a matrix, where each row corresponds to a pollutant and each column represents a sampling event. This PCA (Figure 6) illustrates how individual pollutants behave across various environmental media and sampling conditions, enabling the visual clustering of pollutants that exhibit similar environmental distribution patterns. Unlike conventional PCA, which groups similar samples, this approach enables the direct comparison of air–soil pairs of the same compound, thereby explaining their environmental stability, partitioning behavior, and potential common emission sources.

The results show that many PAHs form different air and soil pairs in the PCA space, suggesting consistent behavior across environmental components. Heavier PAHs are typically associated with combustion-related emissions such as vehicle exhaust, residential heating, and industrial processes, and are strongly particle-bound. Their semi-volatile nature allows them to travel through the atmosphere adsorbed onto fine particulate matter (PM), before depositing and accumulating in topsoil, where they persist due to low water solubility and slow degradation rates. Flu and Pyr exhibit an interesting divergence in the media. In the air, they are spatially separated from the cluster of heavier PAHs (like BaA, BbF, BkF, etc.), likely due to their higher volatility, lower molecular weight, and greater tendency to remain in the gas phase. This reduces their association with particle-bound transport and makes them less likely to behave like the heavier PAHs in atmospheric conditions. However, in soil, their positioning within or near the PAH cluster suggests that once deposited, they can accumulate similarly to heavier compounds, possibly due to adsorption to soil organic matter or repeated seasonal deposition. This asymmetry between air and soil behavior reflects efficient atmospheric transport and episodic deposition on one hand, and strong sorption and environmental persistence in soil on the other.

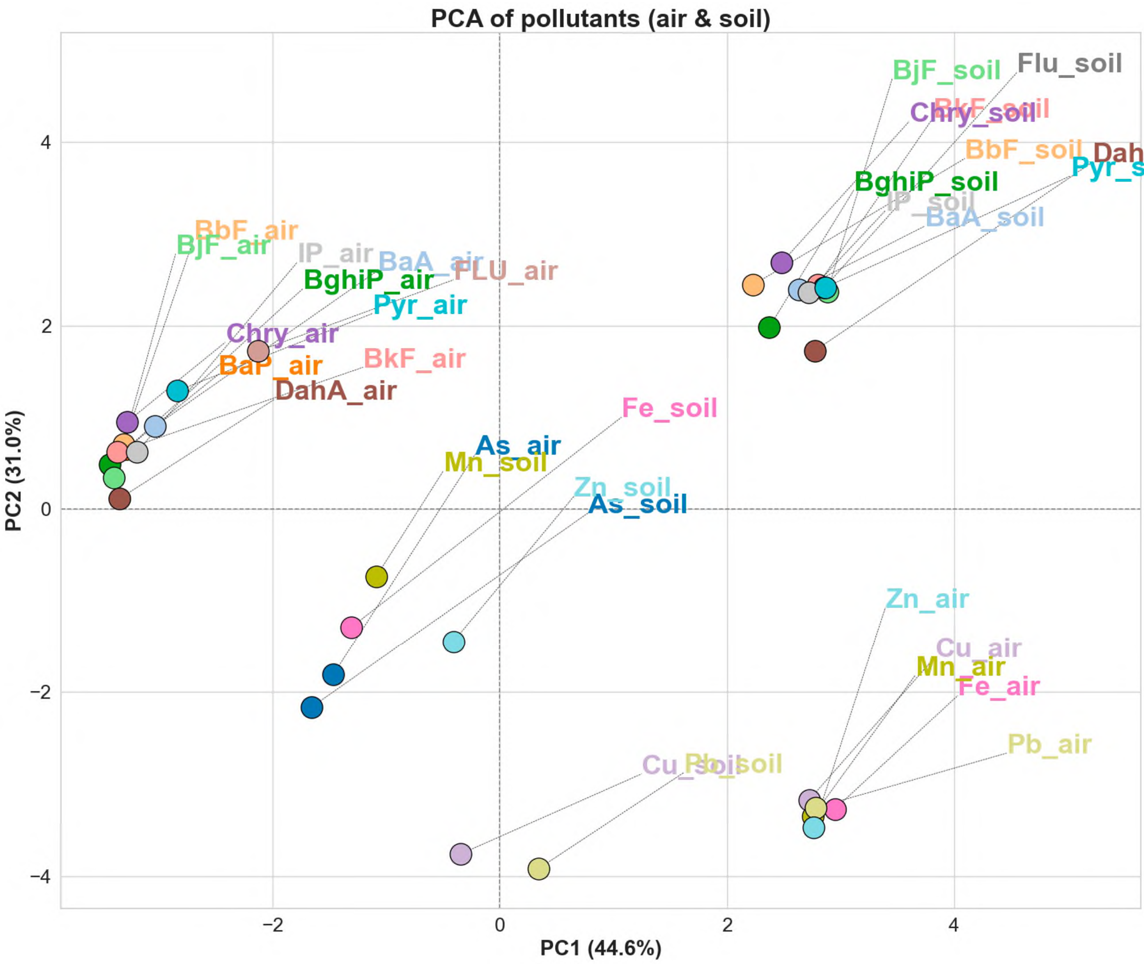


Figure 6. PCA score plot of air and soil pollutants, based on standardized concentrations across all samples.

For trace metals, the PCA indicates varied behavior between elements. Fe and Mn, both of which are naturally abundant crustal elements, show relatively balanced air–soil associations. In urban settings, they are also connected to resuspended road dust and industrial activity. Their proximity in PCA space suggests both geogenic background and anthropogenic enhancement. Pb and Cu also demonstrate coherent air–soil clustering, consistent with their well-established links to traffic-related sources, particularly brake and tire wear. These metals, often associated with finer particles, can be efficiently deposited in soil and are relatively stable, especially in surface layers where they bind to organic matter and clays. Elevated levels of Pb and Cu in Croatian urban soils are clearly associated with anthropogenic influences, especially in high-traffic zones such as Zagreb [21].

Zn, on the other hand, shows marked divergence between its air and soil forms. In the air, Zn is often associated with fine particulate matter emitted from tire wear, lubricating oils, and galvanized surfaces, making it relatively mobile and subject to short-term variability driven by traffic intensity and meteorological conditions. In contrast, Zn in soil may reflect a combination of recent deposition and legacy contamination. Still, its mobility in the soil environment is higher than that of Pb or Cu, particularly in acidic or organic-poor soils. This can result in less retention and spatial variability in soil concentrations.

Additionally, Zn's tendency to leach or redistribute within the soil profile may further decouple its air–soil association compared to more strongly bound metals, such as Pb and Cu. Arsenic shows strong proximity between its air and soil forms in the PCA space, suggesting similar sources and stable environmental partitioning. This coherence may reflect widespread regional emissions from combustion, with efficient deposition and retention in soil. Arsenic's behavior in soil is also influenced by pH and redox conditions, which can stabilize it and reduce its mobility, allowing it to persist and mirror airborne inputs.

Overall, this PCA confirms that while PAHs, especially heavier ones, tend to display coherent behavior across air and soil due to their shared emission sources and stable sorption to particles, trace metals exhibit more variable relations depending on their source, mobility, atmospheric reactivity, and soil retention characteristics. Soil often reflects long-term cumulative deposition and retention, whereas air may represent more recent or transient emissions. This transposed PCA approach provides a novel perspective for simultaneously visualizing pollutant behavior across environmental compartments, underscoring the importance of integrated monitoring in evaluating pollutant fate and transport. Future studies should include long-term and seasonal soil monitoring, as well as direct measurements of particulate and total deposition, to better understand pollutant transfer. Expanding to other media, such as vegetation or dust, and integrating modeling could help assess accumulation trends and long-term risks.

PCA biplots show a clear separation between soil and PM_{10} at all stations, indicating limited cross-compartment coherence for PAHs; even heavier PAHs cluster primarily within each compartment rather than jointly across air and soil. Metals display mixed patterns consistent with source heterogeneity, mobility, and soil retention, but again do not form a unified air–soil cluster. Interpretations have several limitations: the air side relies only on PM_{10} (appropriate for health relevance and data availability, but less direct for soil transfer), no deposition fluxes were measured, and site histories of soil disturbance (e.g., excavation or imported material) are unknown. Another limitation is the short atmospheric record compared with the longer integration times of soils. This restricts assessment of decadal accumulation, though the use of multi-month air averages and two soil depths partly addresses recent deposition signals. Longer air records and deposition flux data would enhance the evaluation of long-term trends, which can be applied and expanded upon in future research.

Despite these limitations, the dataset provides the first published PAH measurements in Zagreb with systematic, multi-year seasonal soil sampling, offering a valuable baseline. Future work should pair soil campaigns with deposition measurements, include other size fractions and co-located metadata on soil management, and extend to additional media (e.g., vegetation, settled dust) and process-based modeling to resolve transfer pathways and long-term accumulation.

4. Conclusions

In this study, concentrations of various PAHs and metals were measured in both ambient air (PM_{10}) and soil samples collected across multiple monitoring stations in Zagreb, Croatia. Soil samples were collected from two depths and analyzed for the same suite of pollutants to investigate the extent and patterns of pollutant deposition and accumulation.

Statistical analyses, including Spearman correlation and PCA, were applied to explore relationships between air and soil pollutant concentrations and to identify main sources.

Seasonal differences were evident, with a 6-month window, as metals, especially Pb, showed a robust air–soil correlation. In contrast, PAHs generally exhibited negative within-season correlations, which were stronger at 5 cm, indicating surface-biased, season-dependent correlations. PCA separated soil from PM_{10}, with PAHs and metals clustering within each compartment rather than jointly, revealing soil's longer-term retention versus more transient atmospheric variability. Spatial variability among monitoring stations explained localized factors affecting pollutant distributions. This study presents the first systematically collected dataset of PAH concentrations in soils of Zagreb, combined with co-located multi-year PM_{10} measurements, allowing for a direct investigation of air–soil pollutant linkages in an urban setting. Concentrations of Σ11PAHs in soils (1.2–524 μg/g) indicate moderate but site-specific contamination. In contrast, metal enrichment (Pb up to 388 μg/g, Zn up to 219 μg/g) reflects both traffic and heating sources. In PM_{10}, mean BaP reached 2.15 ng/m^3 at traffic-affected sites, with Fe, Zn, and Cu dominating the trace metal fraction. Strong seasonal and depth-dependent air–soil correlations were observed for Pb, Mn, and Cu, supporting deposition as a key transfer pathway. In contrast, PAHs showed weaker or negative short-term correlations due to their semi-volatile nature and degradation. The novelty of this work lies in linking multi-year atmospheric and soil datasets to reveal both cumulative and short-term changes in deposition signals. Future studies should extend monitoring to longer timeframes, include direct deposition fluxes and additional media (e.g., vegetation, settled dust), and apply process-based modeling to quantify pollutant transfer pathways and long-term accumulation better. Overall, this integrated approach helps in understanding pollutant transfer from air to soil and provides valuable insights for environmental monitoring, source identification, and risk assessment studies.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/toxics13100866/s1. Figure S1. PAH concentrations in soil by season, depth and station. Figure S2. Metals concentration in soil by season, depth and station. Figure S3. Fe concentration in soil by season, depth and station. Figure S4. Mn concentration in soil by season, depth and station. Table S1. Correlation analysis for air (PM_{10}) and soil (for different depth) pollutants. Table S2. Correlation analysis for air (different time windows) and soil pollutants. Table S3. Correlation analysis for air and soil pollutants for different seasons and depths (for time windows of 6 months for air (PM_{10}) pollutant concentrations). Figure S5. PCA for the soil pollutants (a) and air (PM_{10}) pollutants (b) for different monitoring stations. Figure S6. PCA for PAHs and metals in soil samples for different monitoring stations (mean for both depth and periods). Figure S7. PCA for PAHs and metals air samples for different monitoring stations.

Author Contributions: Conceptualization, N.R., S.R., M.L.; Data curation, I.J., Z.S.Š., J.R., S.Ž., I.S., Ž.Z.; Formal analysis, N.R.; Investigation, N.R.; Methodology, N.R., M.L., S.R.; Resources, M.L., G.P.; Supervision, M.L., S.R., G.P.; Visualization, N.R.; Writing—original draft, N.R.; Writing—review & editing, M.L., S.R., G.P., I.J., J.R., S.Ž., I.S., Ž.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the European Union—Next Generation EU (Program Contract of 8 December 2023, Class: 643-02/23-01/00016, Reg. no. 533-03-23-0006)- EnvironPollutHealth and by Horizon Europe (URBREATH), Grant ID: 101139711 granted to the Lisbon Council. Soil samples were collected under the IMROH internal project "Levels of Platinum Group Elements (PGE) Near Roads (2018–2022)".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, [M.L.], upon reasonable request.

Acknowledgments: This study was performed using the facilities and equipment funded within the European Regional Development Fund project KK.01.1.1.02.0007 "Research and Education Centre of Environmental Health and Radiation Protection—Reconstruction and Expansion of the Institute for Medical Research and Occupational Health". Measurements of metals and some PAHs within the local and national air quality monitoring networks were funded by the City of Zagreb and the Croatian Environmental Protection and Energy Efficiency Fund, respectively.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Bezberdaya, L.A.; Kasimov, N.S.; Chernitsova, O.V.; Tkachenko, A.N.; Lychagin, M.Y. Heavy metals and Metalloids in Soils, Road Dust, and Their PM10 Fractions in Sebastopol: Levels, Sources, and Pollution Risk. *Eurasian Soil Sci.* **2022**, *55*, 1871–1890. [CrossRef]
2. Ciarkowska, K.; Gambus, F.; Antonkiewicz, J.; Koliopoulos, T. Polycyclic Aromatic Hydrocarbon and Heavy Metal Contents in the Urban Soils in Southern Poland. *Chemosphere* **2019**, *229*, 214–226. [CrossRef] [PubMed]
3. Di Vaio, P.; Cocozziello, B.; Corvino, A.; Fiorino, F.; Frecentese, F.; Magli, E.; Onorati, G.; Saccone, I.; Santagada, V.; Settimo, G.; et al. Level, Potential Sources of Polycyclic Aromatic Hydrocarbons (PAHs) in Particulate Matter (PM10) in Naples. *Atmos. Environ.* **2016**, *129*, 186–196. [CrossRef]
4. Fazeli, G.; Karbassi, A.; Khoramnejadian, S.; Nasrabadi, T. Evaluation of Urban Soil Pollution: A Combined Approach of Toxic Metals and Polycyclic Aromatic Hydrocarbons (PAHs). *Int. J. Environ. Res.* **2019**, *13*, 801–811. [CrossRef]
5. Li, Y.; Liu, J.; Cao, Z.; Lin, C.; Yang, Z. Spatial Distribution and Health Risk of Heavy Metals and Polycyclic Aromatic Hydrocarbons (PAHs) in the Water of the Luanhe River Basin, China. *Environ. Monit. Assess.* **2010**, *163*, 1–13. [CrossRef]
6. Morillo, E.; Romero, A.S.; Maqueda, C.; Madrid, L.; Ajmone-Marsan, F.; Grcman, H.; Davidson, C.M.; Hursthouse, A.S.; Villaverde, J. Soil Pollution by PAHs in Urban Soils: A Comparison of Three European Cities. *J. Environ. Monit.* **2007**, *9*, 1001. [CrossRef]
7. Peng, K.; Li, Z.; Gao, T.-R.; Lv, J.; Wang, W.-J.; Zhan, P.; Yao, W.-C.; Zhao, H.; Wang, H.; Xu, D.-X.; et al. Polycyclic Aromatic Hydrocarbon Exposure Burden: Individual and Mixture Analyses of Associations with Chronic Obstructive Pulmonary Disease Risk. *Environ. Res.* **2023**, *222*, 115334. [CrossRef]
8. Liu, J.; Jia, J.; Grathwohl, P. Dilution of Concentrations of PAHs from Atmospheric Particles, Bulk Deposition to Soil: A Review. *Environ. Geochem. Health* **2022**, *44*, 4219–4234. [CrossRef]
9. Mwangi, J.K.; Degrendele, C.; Bandowe, B.A.M.; Bohlin-Nizzetto, P.; Halse, A.K.; Šmejkalová, A.H.; Kim, J.-T.; Kukučka, P.; Martiník, J.; Nežiková, B.P.; et al. Air-Soil Cycling of Oxygenated, Nitrated, and Parent Polycyclic Aromatic Hydrocarbons in Source and Receptor Areas. *Sci. Total Environ.* **2024**, *921*, 170495. [CrossRef]
10. Ambade, B.; Kumar, A.; Latif, M. Emission Sources, Characteristics and Risk Assessment of Particulate Bound Polycyclic Aromatic Hydrocarbons (PAHs) from Traffic Sites. *Res. Sq.* **2021**. [CrossRef]
11. Chen, L.-C.; Maciejczyk, P.; Thurston, G.D. Metals and Air Pollution. In *Handbook on the Toxicology of Metals*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 137–182, ISBN 978-0-12-823292-7.
12. Turnbull, A.B.; Harrison, R.M. Major Component Contributions to PM10 Composition in the UK Atmosphere. *Atmos. Environ.* **2000**, *34*, 3129–3137. [CrossRef]
13. Hassanvand, M.S.; Naddafi, K.; Faridi, S.; Nabizadeh, R.; Sowlat, M.H.; Momeniha, F.; Gholampour, A.; Arhami, M.; Kashani, H.; Zare, A.; et al. Characterization of PAHs and Metals in Indoor/Outdoor PM10/PM2.5/PM1 in a Retirement Home and a School Dormitory. *Sci. Total Environ.* **2015**, *527–528*, 100–110. [CrossRef] [PubMed]
14. Callén, M.S.; López, J.M.; Iturmendi, A.; Mastral, A.M. Nature and Sources of Particle Associated Polycyclic Aromatic Hydrocarbons (PAH) in the Atmospheric Environment of an Urban Area. *Environ. Pollut.* **2013**, *183*, 166–174. [CrossRef]
15. Dat, N.-D.; Chang, M.B. Review on Characteristics of PAHs in Atmosphere, Anthropogenic Sources and Control Technologies. *Sci. Total Environ.* **2017**, *609*, 682–693. [CrossRef] [PubMed]
16. Račić, N.; Ružičić, S.; Terzić, T.; Pehnec, G.; Jakovljević, I.; Štrukil, Z.S.; Žužul, S.; Rinkovec, J.; Lovrić, M. Analyzing the Relationship between Gas Consumption and Airborne Pollutants: Case Study of Zagreb, Croatia. *Air Qual. Atmos. Health* **2024**, *18*, 507–519. [CrossRef]
17. Ravindra, K.; Sokhi, R.; Vangrieken, R. Atmospheric Polycyclic Aromatic Hydrocarbons: Source Attribution, Emission Factors and Regulation. *Atmos. Environ.* **2008**, *42*, 2895–2921. [CrossRef]
18. Beslic, I.; Burger, J.; Cadoni, F.; Centioli, D.; Kranjc, I.; Van Den Bril, B.; Rinkovec, J.; Sega, K.; Zang, T.; Zuzul, S.; et al. Determination of As, Cd, Ni and Pb in PM10—Comparison of Different Sample Work-up and Analysis Methods/Bestimmung

von As, Cd, Ni Und Pb in PM10—Vergleich Verschiedener Probenaufbereitungs- Und Analysenverfahren. *GrdL* **2020**, *80*, 227–233. [CrossRef]
19. Idani, E.; Geravandi, S.; Akhzari, M.; Goudarzi, G.; Alavi, N.; Yari, A.R.; Mehrpour, M.; Khavasi, M.; Bahmaei, J.; Bostan, H.; et al. Characteristics, Sources, and Health Risks of Atmospheric PM10-Bound Heavy metals in a Populated Middle Eastern City. *Toxin Rev.* **2020**, *39*, 266–274. [CrossRef]
20. Lee, B.K.; Hieu, N.T. Seasonal Variation and Sources of Heavy Metals in Atmospheric Aerosols in a Residential Area of Ul-san, Korea. *Aerosol Air Qual. Res.* **2011**, *11*, 679–688. [CrossRef]
21. Halamic, J.; Miko, S.; Peh, Z.; Galović, L. *Geochemical Atlas of the Republic of Croatia*; Halamic, J., Miko, S., Eds.; HGI Hrvatski Geoloski Institut: Zagreb, Croatia, 2009; ISBN 978-953-6907-18-2.
22. Castel, R.; Bertoldo, R.; Lebarillier, S.; Noack, Y.; Orsière, T.; Malleret, L. Toward an Interdisciplinary Approach to Assess the Adverse Health Effects of Dust-Containing Polycyclic Aromatic Hydrocarbons (PAHs) and Metal(Loid)s on Preschool Children. *Environ. Pollut.* **2023**, *336*, 122372. [CrossRef]
23. Bukowska, B.; Mokra, K.; Michałowicz, J. Benzo[a]Pyrene—Environmental Occurrence, Human Exposure, and Mechanisms of Toxicity. *Int. J. Mol. Sci.* **2022**, *23*, 6348. [CrossRef]
24. Guerreiro, C.B.B.; Horálek, J.; De Leeuw, F.; Couvidat, F. Benzo(a)Pyrene in Europe: Ambient Air Concentrations, Population Exposure and Health Effects. *Environ. Pollut.* **2016**, *214*, 657–667. [CrossRef]
25. Kebe, M.; Traore, A.; Sow, M.; Fall, S.; Tahri, M. Human Health Risk Evaluation of Particle Air Pollution (PM10 and PM2.5) and Heavy metals in Dakar's Two Urban Areas. *Asian J. Atmos. Environ.* **2025**, *19*, 7. [CrossRef]
26. Kumari, S.; Jain, M.K.; Elumalai, S.P. Assessment of Pollution and Health Risks of Heavy metals in Particulate Matter and Road Dust Along the Road Network of Dhanbad, India. *J. Health Pollut.* **2021**, *11*, 210305. [CrossRef] [PubMed]
27. Uzoekwe, S.A.; Izah, S.C.; Aigberua, A.O. Environmental and Human Health Risk of Heavy metals in Atmospheric Particulate Matter (PM10) around Gas Flaring Vicinity in Bayelsa State, Nigeria. *Toxicol. Environ. Health Sci.* **2021**, *13*, 323–335. [CrossRef]
28. Desboeufs, K. Trace Metals and Contaminants Deposition. In *Atmospheric Chemistry in the Mediterranean Region*; Dulac, F., Sauvage, S., Hamonou, E., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 345–369, ISBN 978-3-030-82384-9.
29. Jiang, S.; Dong, X.; Han, Z.; Zhao, J.; Zhang, Y. Emissions and Atmospheric Dry and Wet Deposition of Trace Metals from Natural and Anthropogenic Sources in Mainland China. *Atmosphere* **2024**, *15*, 402. [CrossRef]
30. Xia, W.; Liang, B.; Chen, L.; Zhu, Y.; Gao, M.; Chen, J.; Wang, F.; Chen, Y.; Tian, M. Atmospheric Wet and Dry Depositions of Polycyclic Aromatic Compounds in a Megacity of Southwest China. *Environ. Res.* **2022**, *204*, 112151. [CrossRef] [PubMed]
31. Balcioğlu, E.B.; Çevik, F.E.; Aksu, A. Source Determination and Seasonal Distribution of Polycyclic Aromatic Hydrocarbons (PAHs) in Urban Soil of the Megacity Istanbul. *Polycycl. Aromat. Compd.* **2021**, *41*, 626–634. [CrossRef]
32. Wang, H.; Yang, Y.; Walker, T.R.; Wang, Y.; Wu, H.; Wang, X.; Luo, Q. Characterization, Source Apportionment, and Risk Assessment of Polycyclic Aromatic Hydrocarbons (PAHs) in Urban Soils from 23 Cities in China. *Environ. Sci. Pollut. Res.* **2022**, *29*, 73401–73413. [CrossRef]
33. Binner, H.; Sullivan, T.; Jansen, M.A.K.; McNamara, M.E. Metals in Urban Soils of Europe: A Systematic Review. *Sci. Total Environ.* **2023**, *854*, 158734. [CrossRef]
34. Zheng, F.; Guo, X.; Tang, M.; Zhu, D.; Wang, H.; Yang, X.; Chen, B. Variation in Pollution Status, Sources, and Risks of Soil Heavy metals in Regions with Different Levels of Urbanization. *Sci. Total Environ.* **2023**, *866*, 161355. [CrossRef]
35. Khan, Y.K.; Toqeer, M.; Shah, M.H. Mobility, Bioaccessibility, Pollution Assessment and Risk Characterization of Potentially Toxic Metals in the Urban Soil of Lahore, Pakistan. *Environ. Geochem. Health* **2023**, *45*, 1391–1412. [CrossRef] [PubMed]
36. Rate, A.W. Urban Soil as a Source and Sink. In *Urban Soils*; Rate, A.W., Ed.; Progress in Soil Science; Springer International Publishing: Cham, Switzerland, 2022; pp. 293–317, ISBN 978-3-030-87315-8.
37. Nipen, M.; Vogt, R.D.; Bohlin-Nizzetto, P.; Borgå, K.; Mwakalapa, E.B.; Borgen, A.R.; Jørgensen, S.J.; Ntapanta, S.M.; Mmochi, A.J.; Schlabach, M.; et al. Spatial Trends of Chlorinated Paraffins and Dechloranes in Air and Soil in a Tropical Urban, Suburban, and Rural Environment. *Environ. Pollut.* **2022**, *292*, 118298. [CrossRef] [PubMed]
38. Ravi, S.; Li, J.; Meng, Z.; Zhang, J.; Mohanty, S. Generation, Resuspension, and Transport of Particulate Matter From Biochar-Amended Soils: A Potential Health Risk. *GeoHealth* **2020**, *4*, e2020GH000311. [CrossRef]
39. Boahen, E. Heavy Metal Contamination in Urban Roadside Vegetables: Origins, Exposure Pathways, and Health Implications. *Discov. Environ.* **2024**, *2*, 145. [CrossRef]
40. Izquierdo-Díaz, M.; Hansen, V.; Barrio-Parra, F.; De Miguel, E.; You, Y.; Magid, J. Assessment of Lettuces Grown in Urban Areas for Human Consumption and as Bioindicators of Atmospheric Pollution. *Ecotoxicol. Environ. Saf.* **2023**, *256*, 114883. [CrossRef]
41. City of Zagreb Statistički Ljetopis Grada Zagreba. 2023. Available online: https://www.zagreb.hr/UserDocsImages/001/SLJZG23_web.pdf (accessed on 9 October 2025).
42. Velić, J.; Saftić, B.; Malvić, T. Lithologic composition and stratigraphy of quaternary sediments in the area of the "Jakuševec" Waste Depository, Zagreb, Northern Croatia. *Geol. Croat.* **1999**, *52*, 119–130. [CrossRef]

43. Račić, N.; Malvić, T. Relation between air and soil pollution based on statistical analysis and interpolation of Nickel (Ni) and Lead (Pb): Case study of Zagreb, Croatia. *Min. Miner. Depos.* **2023**, *17*, 112–120. [CrossRef]
44. *EN 12341:2023*; Ambient Air—Standard Gravimetric Measurement Method for the Determination of the PM_{10} or $PM_{2.5}$ Mass Concentration of Suspended Particulate Matter. European Committee for Standardization (CEN): Brussels, Belgium, 2023.
45. *EN 14902:2005*; Ambient Air Quality—Standard Method for the Measurement of Pb, Cd, As and Ni in the PM_{10} Fraction of Suspended Particulate Matter. European Committee for Standardization (CEN): Brussels, Belgium, 2005.
46. *EN 14902:2005/AC:2006*; Ambient Air Quality—Standard Method for the Measurement of Pb, Cd, As and Ni in the PM_{10} Fraction of Suspended Particulate Matter—Technical Corrigendum 1. European Committee for Standardization (CEN): Brussels, Belgium, 2006.
47. *EN 15549:2008*; Ambient Air Quality—Standard Method for the Measurement of the Concentration of Benzo[a]pyrene in Ambient Air. European Committee for Standardization (CEN): Brussels, Belgium, 2008.
48. *CEN/TS 16645:2014*; Ambient Air—Method for the Measurement of Benz[a]anthracene, Benzo[b]fluoranthene, Benzo[j]fluoranthene, Benzo[k]fluoranthene, Dibenz[a,h]anthracene, Indeno[1,2,3-cd]pyrene and Benzo[ghi]perylene. European Committee for Standardization (CEN): Brussels, Belgium, 2014.
49. Jakovljević, I.; Sever Štrukil, Z.; Godec, R.; Bešlić, I.; Davila, S.; Lovrić, M.; Pehnec, G. Pollution Sources and Carcinogenic Risk of PAHs in PM1 Particle Fraction in an Urban Area. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9587. [CrossRef]
50. Jakovljević, I.; Pehnec, G.; Vađić, V.; Čačković, M.; Tomašić, V.; Jelinić, J.D. Polycyclic Aromatic Hydrocarbons in PM_{10}, $PM_{2.5}$ and PM_1 Particle Fractions in an Urban Area. *Air Qual. Atmos. Health* **2018**, *11*, 843–854. [CrossRef]
51. Jakovljević, I.; Štrukil, Z.S.; Pehnec, G.; Horvat, T.; Sanković, M.; Šumanovac, A.; Davila, S.; Račić, N.; Gajski, G. Ambient Air Pollution and Carcinogenic Activity at Three Different Urban Locations. *Ecotoxicol. Environ. Saf.* **2025**, *289*, 117704. [CrossRef] [PubMed]
52. Lovrić, M.; Antunović, M.; Šunić, I.; Vuković, M.; Kecorius, S.; Kröll, M.; Bešlić, I.; Šimić, I.; Pehnec, G. Insignificant Changes in Particulate Matter during the COVID-19 Lockdown: A Machine Learning Study in Zagreb, Croatia. In Proceedings of the 7th World Congress on Civil, Structural, and Environmental Engineering, CSEE 2022, Virtual Conference, 10–12 April 2022.
53. Pehnec, G.; Jakovljević, I.; Šišović, A.; Bešlić, I.; Vađić, V. Influence of Ozone and Meteorological Parameters on Levels of Polycyclic Aromatic Hydrocarbons in the Air. *Atmos. Environ.* **2016**, *131*, 263–268. [CrossRef]
54. Šišović, A.; Pehnec, G.; Jakovljević, I.; Šilović Hujić, M.; Vađić, V.; Bešlić, I. Polycyclic Aromatic Hydrocarbons at Different Crossroads in Zagreb, Croatia. *Bull. Environ. Contam. Toxicol.* **2012**, *88*, 438–442. [CrossRef]
55. Wilcke, W. SYNOPSIS Polycyclic Aromatic Hydrocarbons (PAHs) in Soil—A Review. *J. Plant Nutr. Soil Sci.* **2000**, *163*, 229–248. [CrossRef]
56. Li, G.; Sun, G.-X.; Ren, Y.; Luo, X.-S.; Zhu, Y.-G. Urban Soil and Human Health: A Review. *Eur. J. Soil Sci.* **2018**, *69*, 196–215. [CrossRef]
57. Škrbić, B.D.; Đurišić-Mladenović, N.; Tadić, Đ.J.; Cvejanov, J.Đ. Polycyclic Aromatic Hydrocarbons in Urban Soil of Novi Sad, Serbia: Occurrence and Cancer Risk Assessment. *Environ. Sci. Pollut. Res.* **2017**, *24*, 16148–16159. [CrossRef]
58. Stepinac, M.; Lourenço, P.B.; Atalić, J.; Kišiček, T.; Uroš, M.; Baniček, M.; Šavor Novak, M. Damage Classification of Residential Buildings in Historical Downtown after the ML5.5 Earthquake in Zagreb, Croatia in 2020. *Int. J. Disaster Risk Reduct.* **2021**, *56*, 102140. [CrossRef]
59. Romic, M.; Romic, D. Heavy metals Distribution in Agricultural Topsoils in Urban Area. *Environ. Geol.* **2003**, *43*, 795–805. [CrossRef]
60. Sollitto, D.; Romic, M.; Castrignanò, A.; Romic, D.; Bakic, H. Assessing Heavy Metal Contamination in Soils of the Zagreb Region (Northwest Croatia) Using Multivariate Geostatistics. *CATENA* **2010**, *80*, 182–194. [CrossRef]
61. Dotaniya, M.L.; Dotaniya, C.K.; Kumar, K.; Doutaniya, R.K.; Meena, H.M. Type of Soil Pollutant and Their Degradation: Methods and Challenges. In *Handbook of Nanomaterials and Nanocomposites for Energy and Environmental Applications*; Springer International Publishing: Cham, Switzerland, 2021; pp. 3103–3134, ISBN 978-3-030-36267-6.
62. Guagliardi, I.; Astel, A.M.; Cicchella, D. Exploring Soil Pollution Patterns Using Self-Organizing Maps. *Toxics* **2022**, *10*, 416. [CrossRef]
63. Montaño-López, F.; Biswas, A. Are Heavy metals in Urban Garden Soils Linked to Vulnerable Populations? A Case Study from Guelph, Canada. *Sci. Rep.* **2021**, *11*, 11286. [CrossRef]
64. Tóth, Z.; Dombos, M.; Hornung, E. Urban Soil Quality Deteriorates Even with Low Heavy Metal Levels: An Arthropod-based Multi-indices Approach. *Ecol. Appl.* **2023**, *33*, e2848. [CrossRef] [PubMed]
65. Silva, H.F.; Silva, N.F.; Oliveira, C.M.; Matos, M.J. Heavy metals Contamination of Urban Soils—A Decade Study in the City of Lisbon, Portugal. *Soil Syst.* **2021**, *5*, 27. [CrossRef]

66. Cheng, K.; Tian, H.Z.; Zhao, D.; Lu, L.; Wang, Y.; Chen, J.; Liu, X.G.; Jia, W.X.; Huang, Z. Atmospheric emission inventory of cadmium from anthropogenic sources. *Int. J. Environ. Sci. Technol.* **2014**, *11*, 605–616. [CrossRef]
67. Dinter, T.C.; Gerzabek, M.H.; Puschenreiter, M.; Strobel, B.W.; Couenberg, P.M.; Zehetner, F. Heavy metal contents, mobility and origin in agricultural topsoils of the Galápagos Islands. *Chemosphere* **2021**, *272*, 129821. [CrossRef]

5.5 Supplementary materials

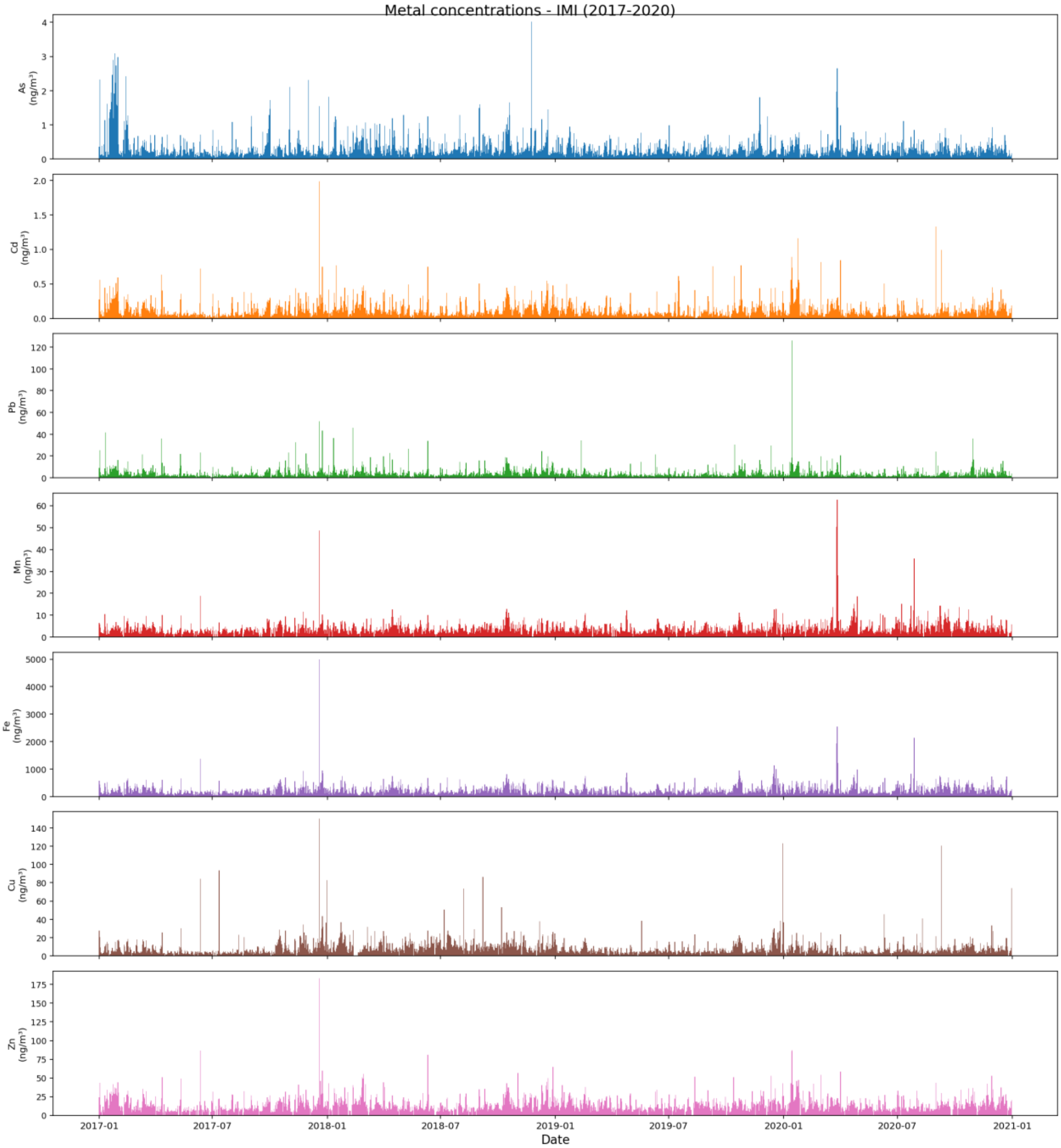


Figure S1. Daily concentrations of selected PTEs in PM_{10} at the IMI monitoring station during the period 2017–2020 (ng/m^3).

5.6 Supplementary materials with publications

5.6.1 Supplementary materials with publication 2:

SUPPLEMENTARY MATERIAL

Contributors to PAHs and metals in PM_{10} using temporal, spatial and emission proxy data in machine learning

The ERA5-Land data is subject to the Copernicus licence from following source where every is described: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=overview, it includes following variables (described at the end of this supplement):
'10m_u_component_of_wind',
'10m_v_component_of_wind',
'2m_dewpoint_temperature',
'clear_sky_direct_solar_radiation_at_surface',
'cloud_base_height',
'convective_precipitation',
'convective_snowfall',
'instantaneous_surface_sensible_heat_flux',
'mean_sea_level_pressure',
'snowfall',
'snowmelt',
'snow_density',
'snow_depth',
'surface_net_solar_radiation',
'surface_net_thermal_radiation',
'temperature_of_snow_layer',
'top_net_thermal_radiation',
'total_cloud_cover',
'total_column_cloud_liquid_water',
'total_column_ozone',
'total_column_rain_water',
'total_column_snow_water',
'total_precipitation',
'precipitation_type_encoded',

Table S1. Relative contribution (%) of individual PAHs to ΣPAH at each monitoring station in Zagreb

Pollutant	IMI	SIG	ZG1	ZG3
BaA	6.58	7.67	7.24	7.61
BaP	12.23	13.06	12.09	12.36
BbF	15.47	15.07	14.95	14.97
BghiP	12.86	13.00	12.16	12.03

BjF	9.51	8.98	8.61	8.83
BkF	6.14	5.89	5.76	5.81
Chry	11.12	12.34	12.21	12.17
DahA	1.84	1.77	1.59	1.58
Flu	5.79	4.92	6.80	6.40
IP	12.75	11.98	12.00	12.30
Pyr	5.71	5.32	6.58	5.92
As	0.12	33.59	0.07	0.13
Cd	0.05	16.89	0.03	0.07
Cu	3.10	1.43	2.98	2.81
Fe	88.62	44.46	90.92	86.83
Mn	1.44	0.58	1.10	1.43
Pb	1.66	0.65	1.10	2.65
Zn	5.02	2.39	3.82	6.08

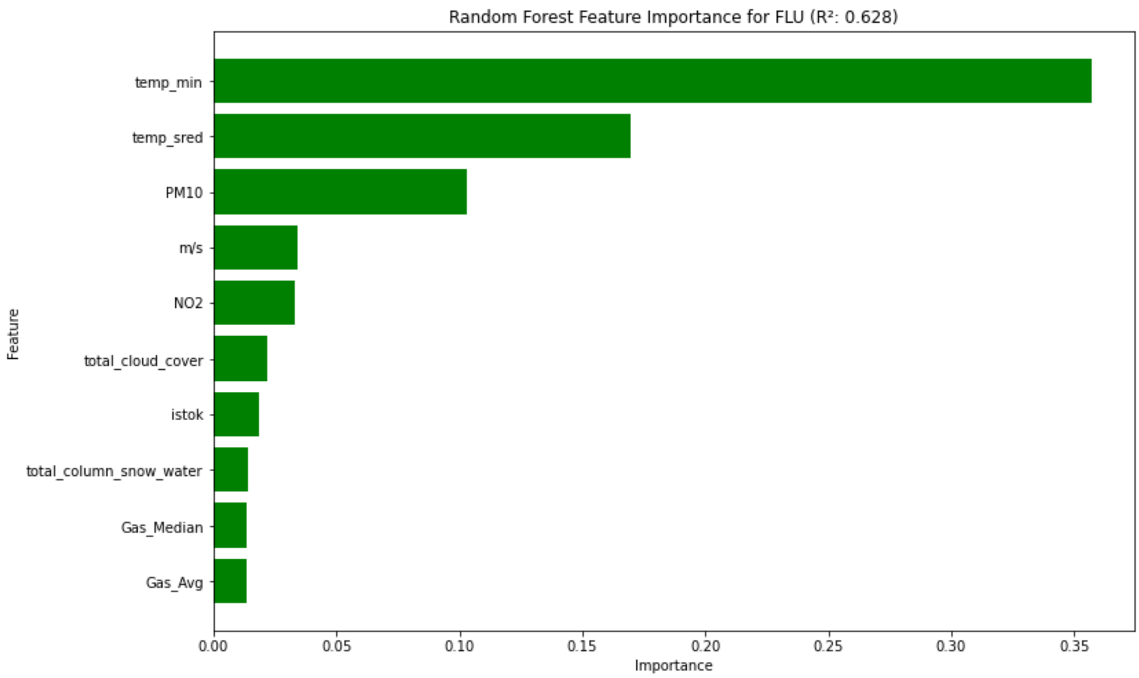


Figure S1. Feature importance plot for Flu

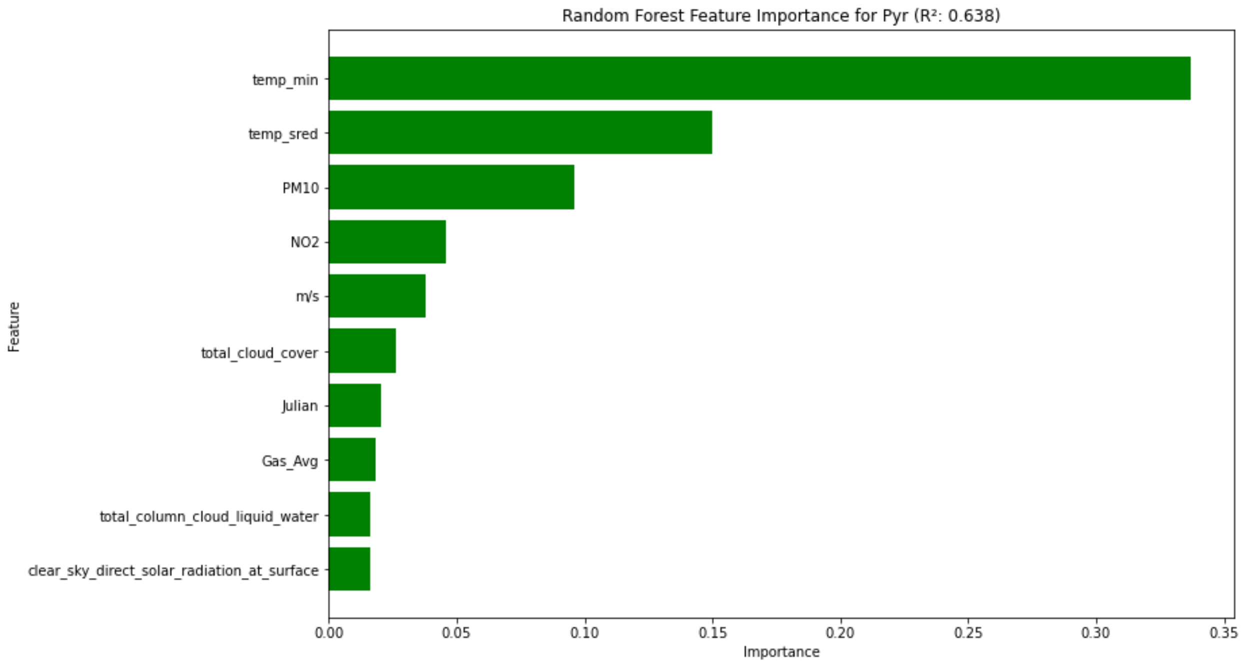


Figure S2. Feature importance plot for Pyr

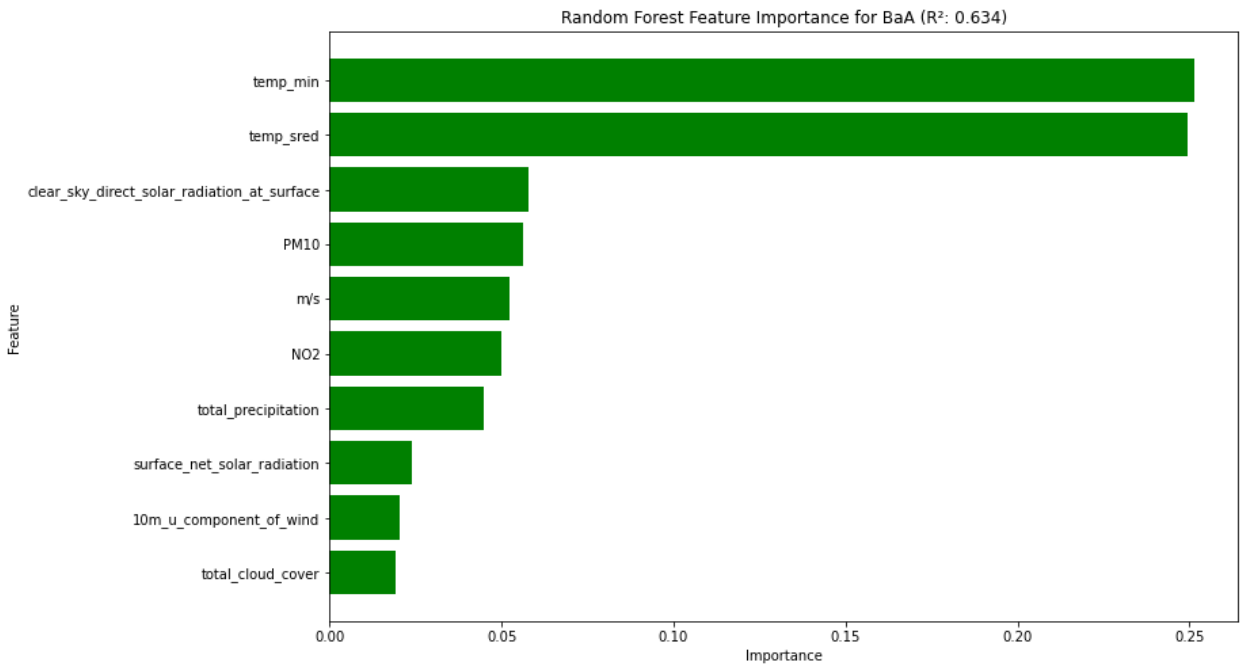


Figure S3. Feature importance plot for BaA

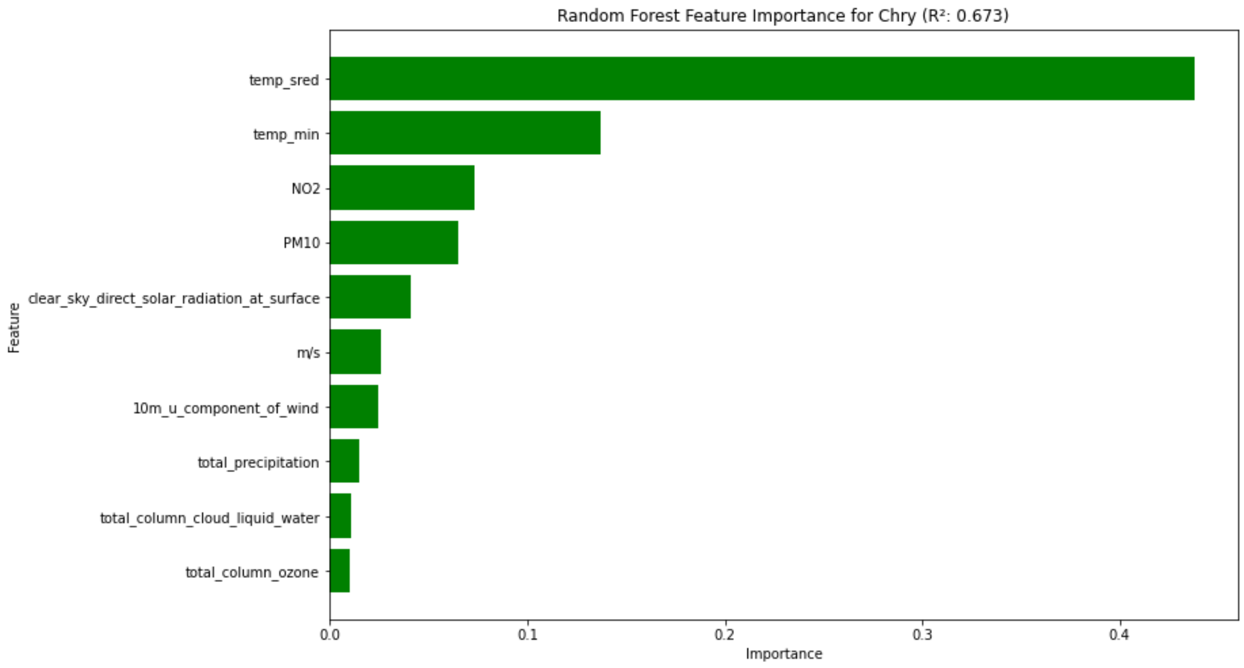


Figure S4. Feature importance plot for Chry

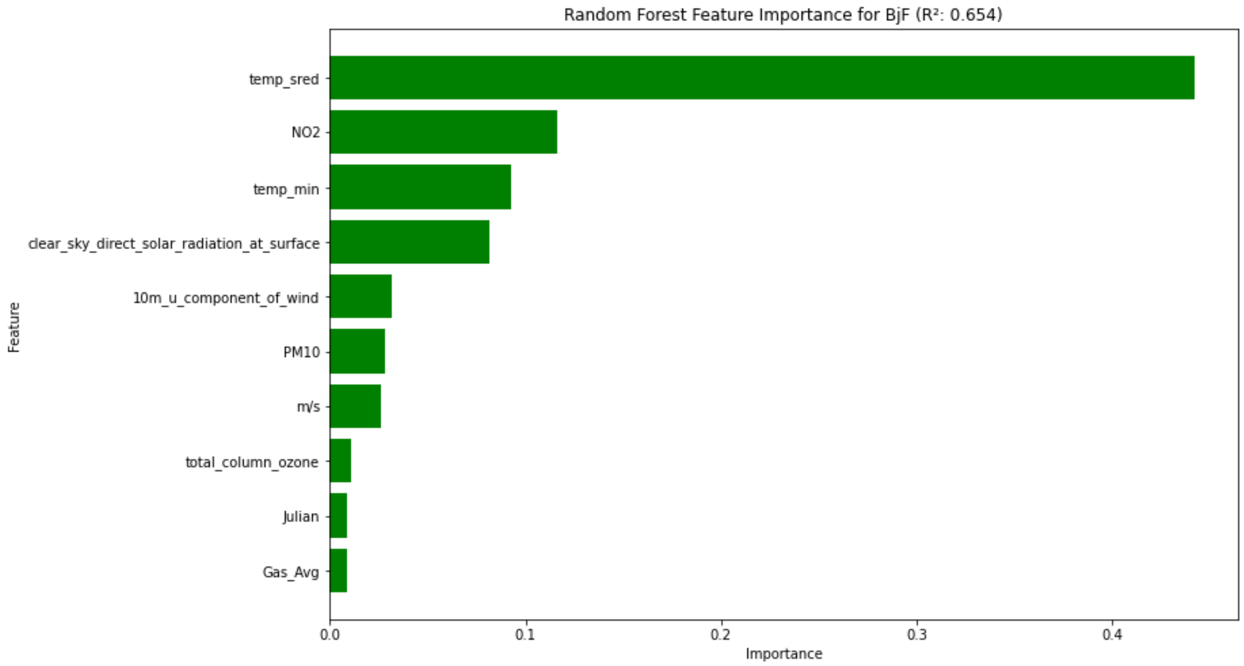


Figure S5. Feature importance plot for BjF

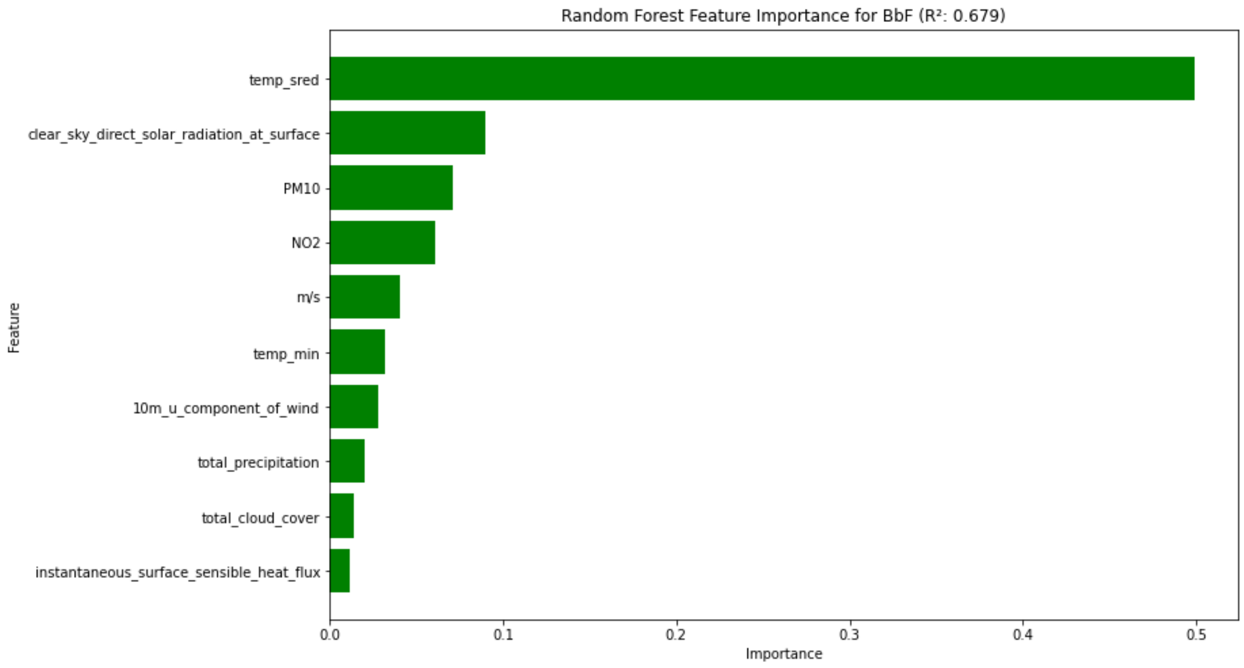


Figure S6. Feature importance plot for BbF

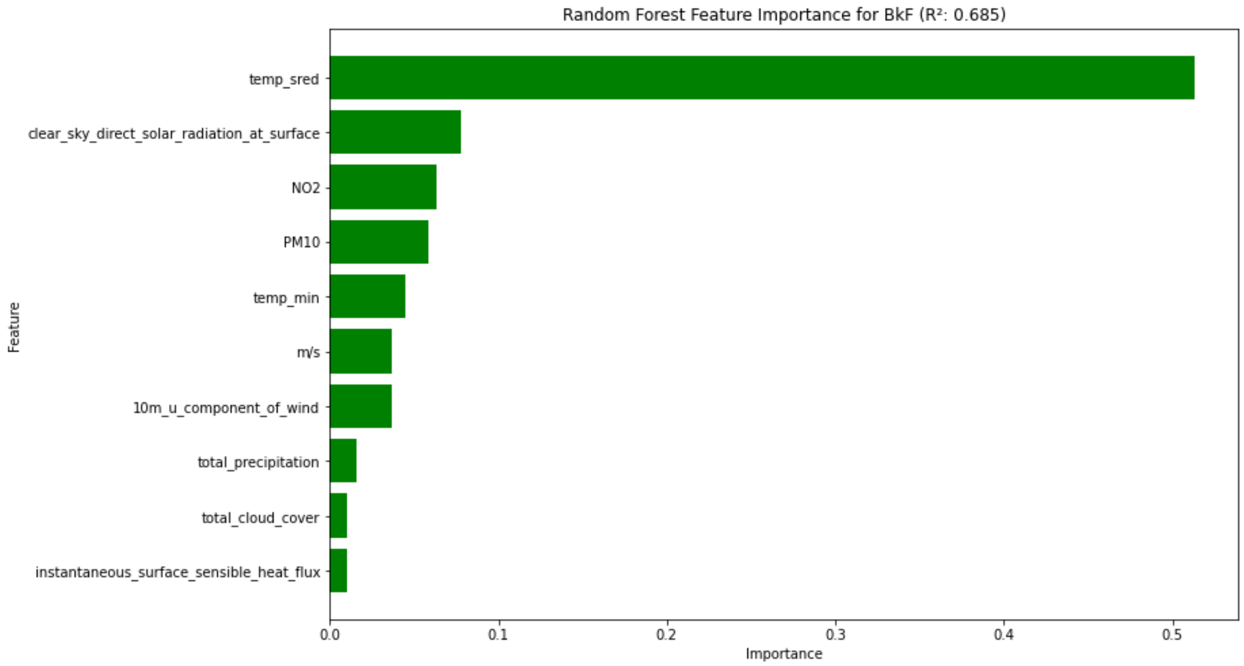


Figure S7. Feature importance plot for BkF

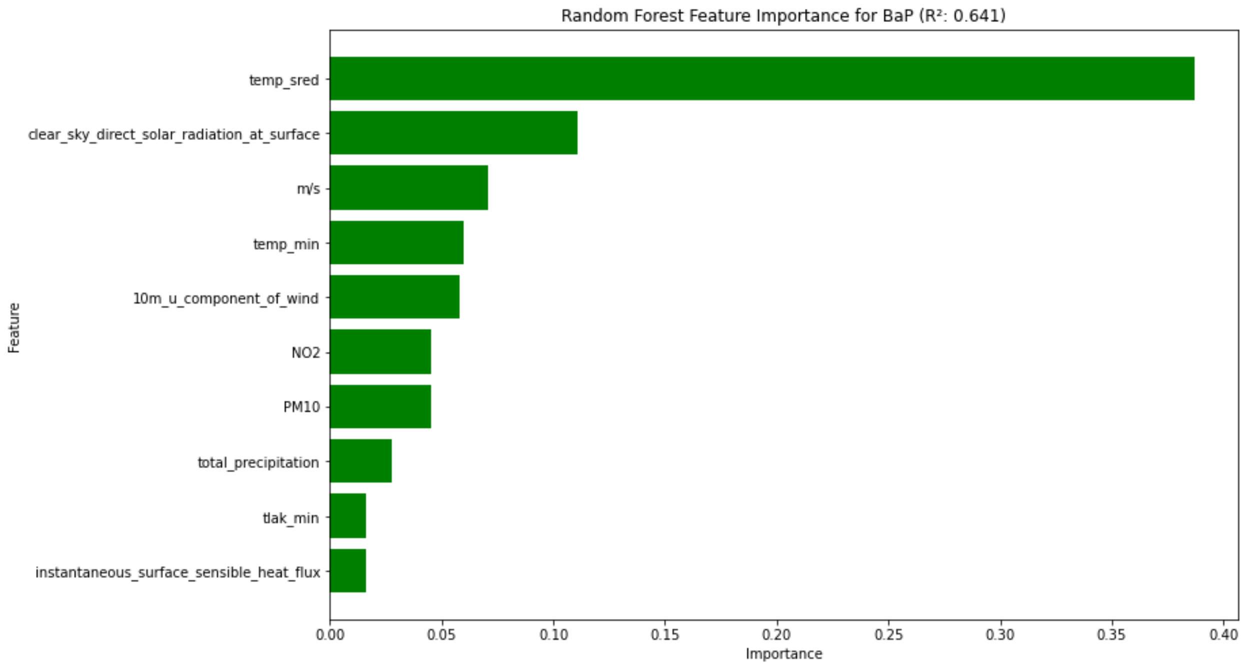


Figure S8. Feature importance plot for BaP

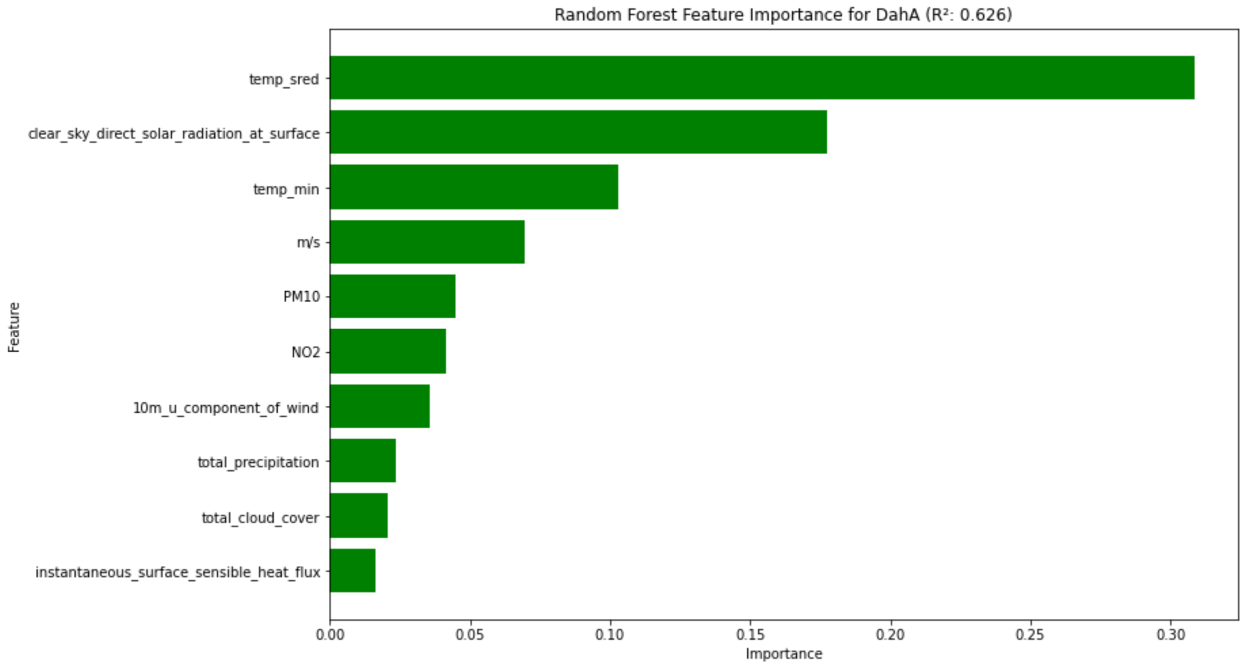


Figure S9. Feature importance plot for DahA

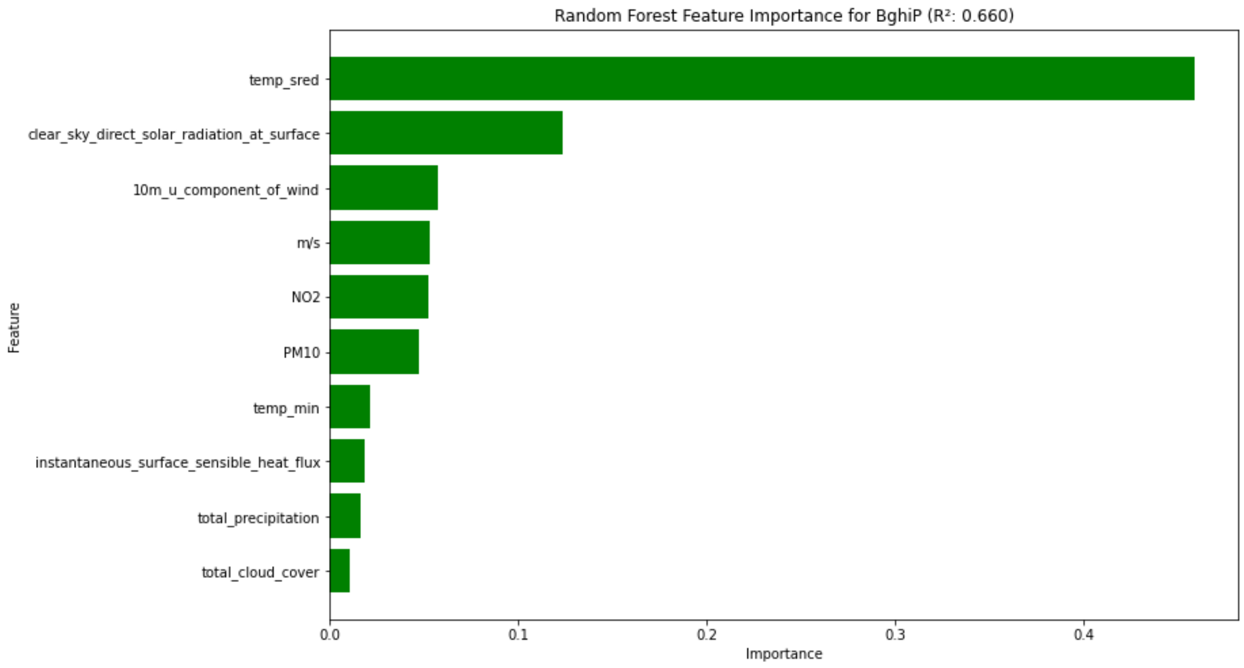


Figure S10. Feature importance plot for BghiP

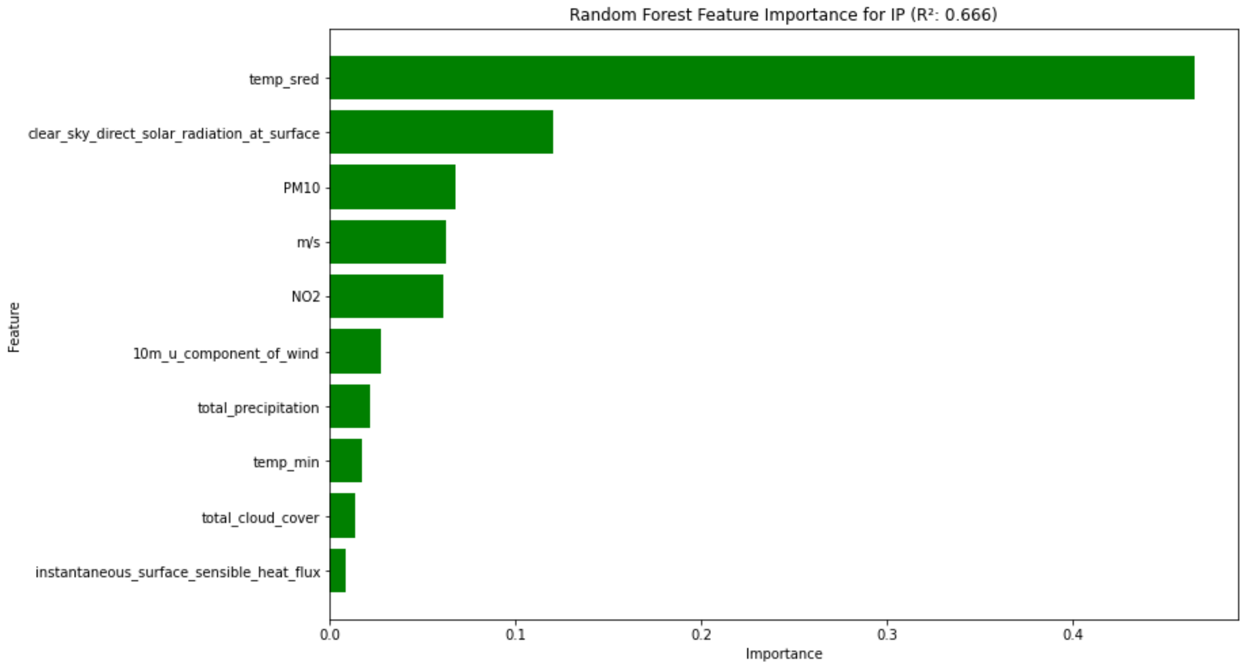


Figure S11. Feature importance plot for IP

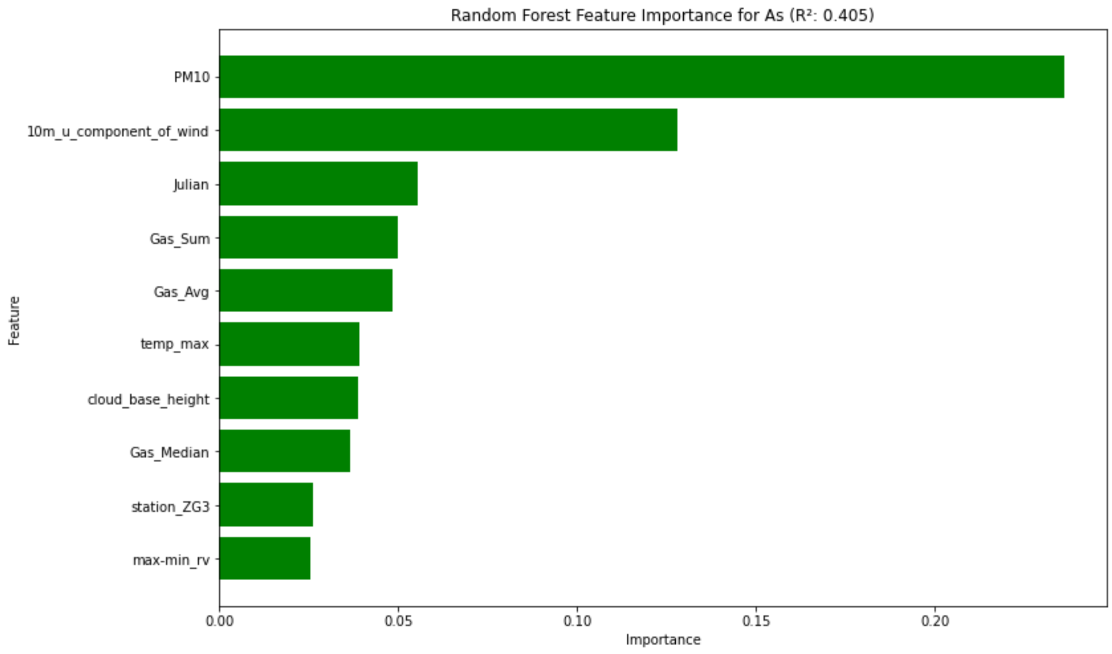


Figure S12. Feature importance plot for As

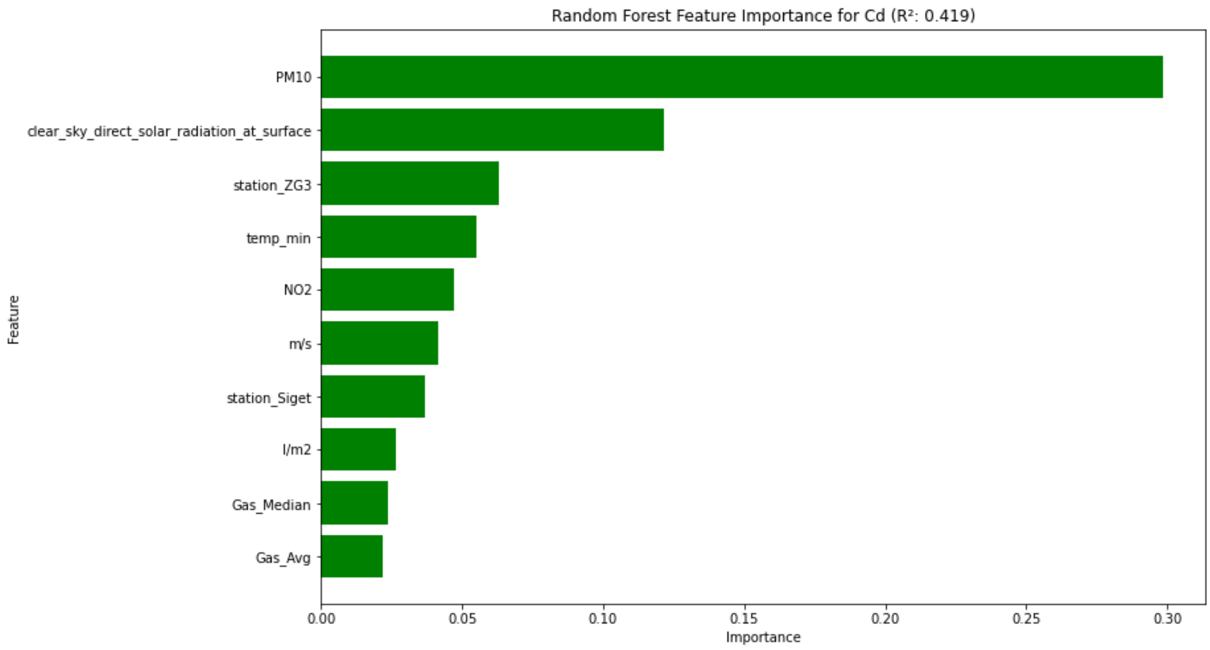


Figure S13. Feature importance plot for Cd

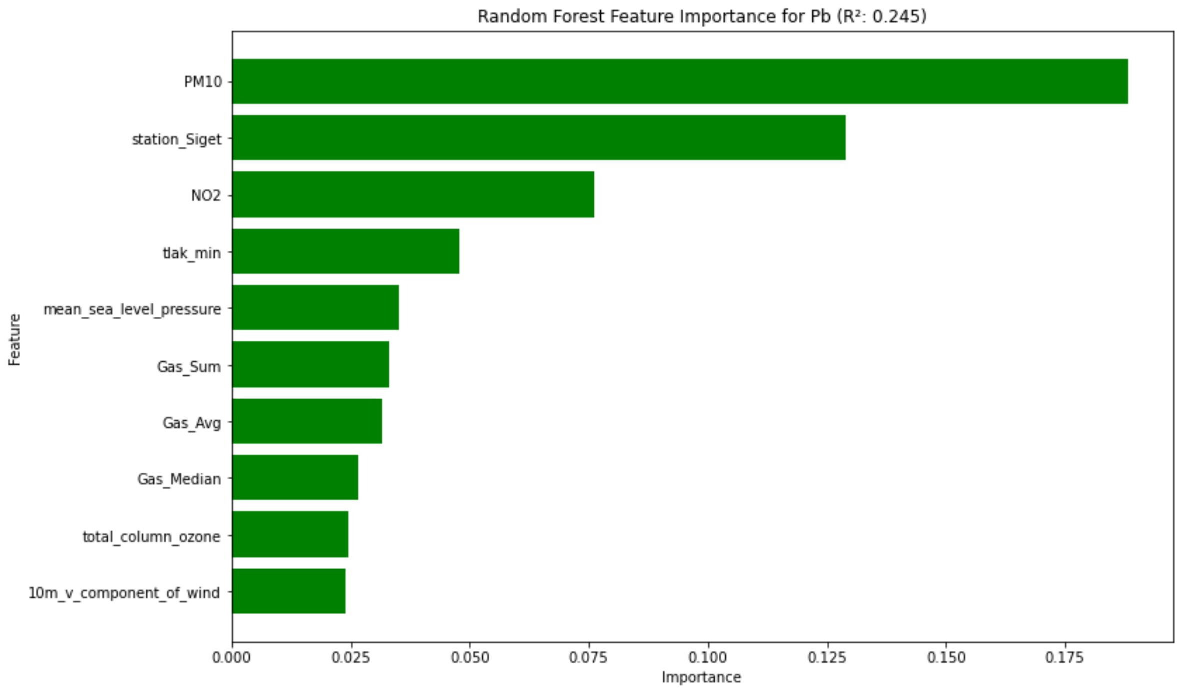


Figure S14. Feature importance plot for Pb

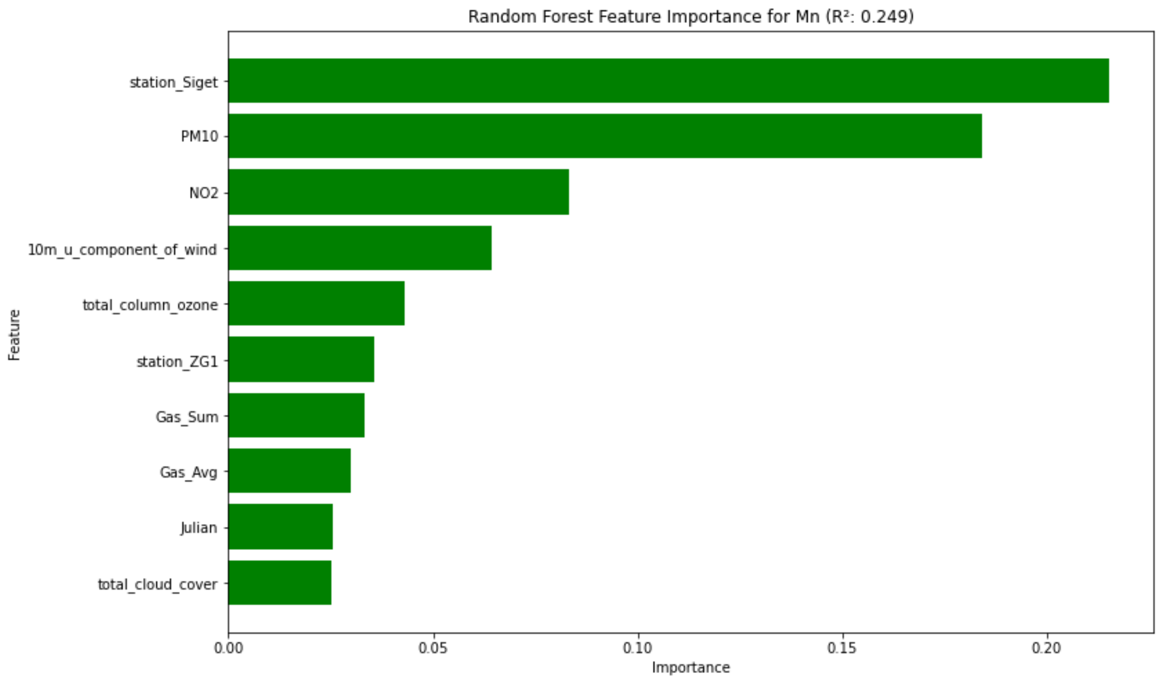


Figure S15. Feature importance plot for Mn

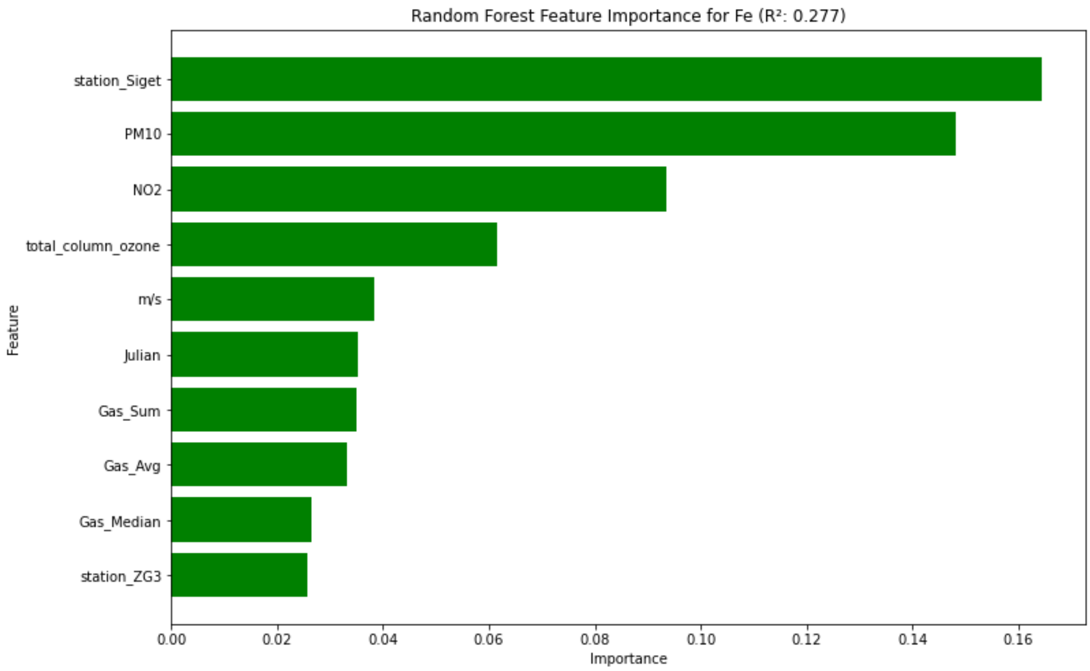


Figure S16. Feature importance plot for Fe

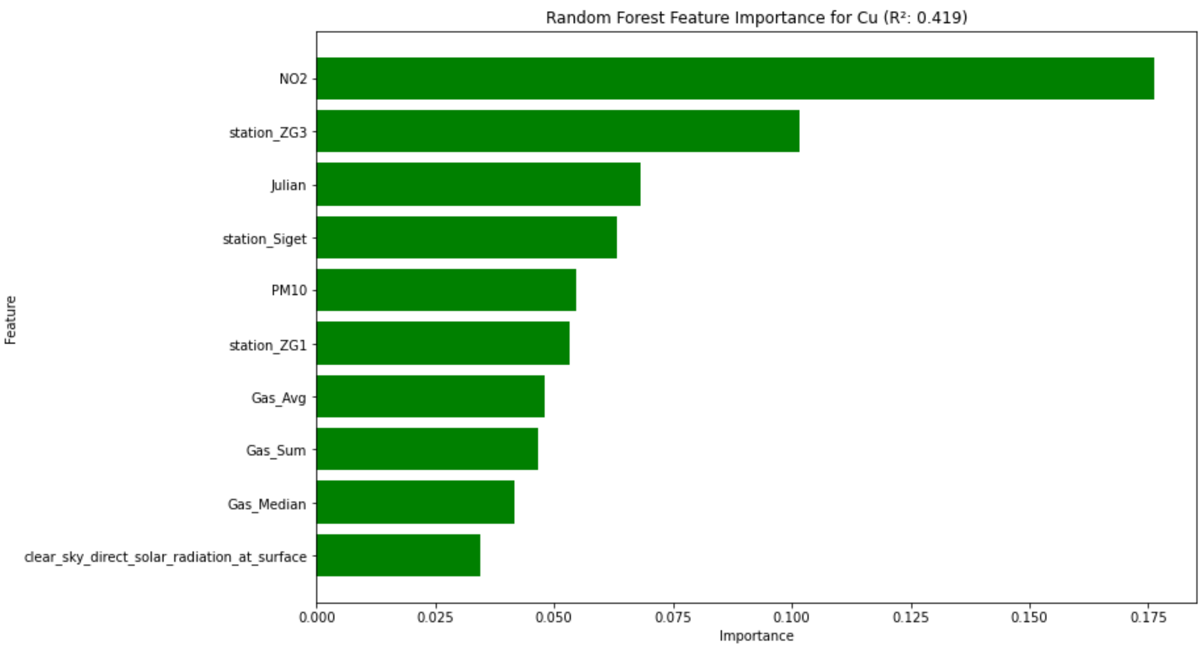


Figure S17. Feature importance plot for Cu

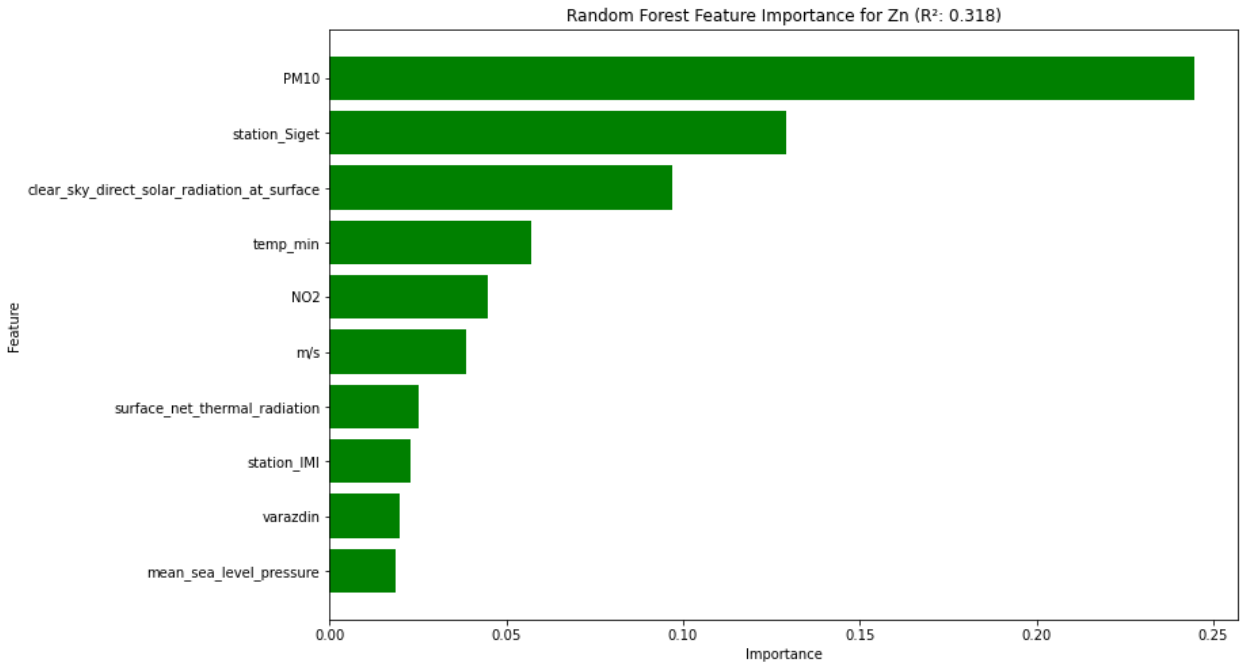


Figure S18. Feature importance plot for Zn

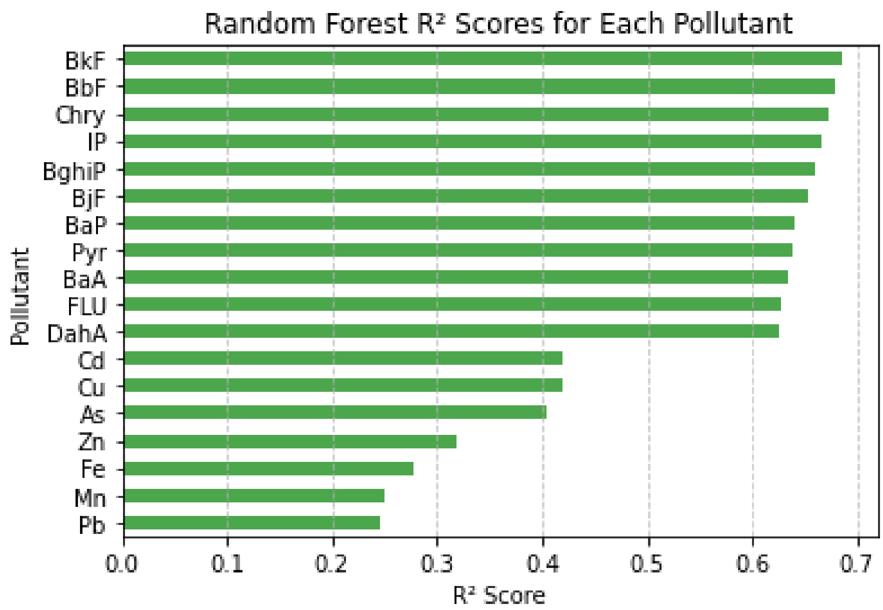


Figure S19. RF scores for every pollutant

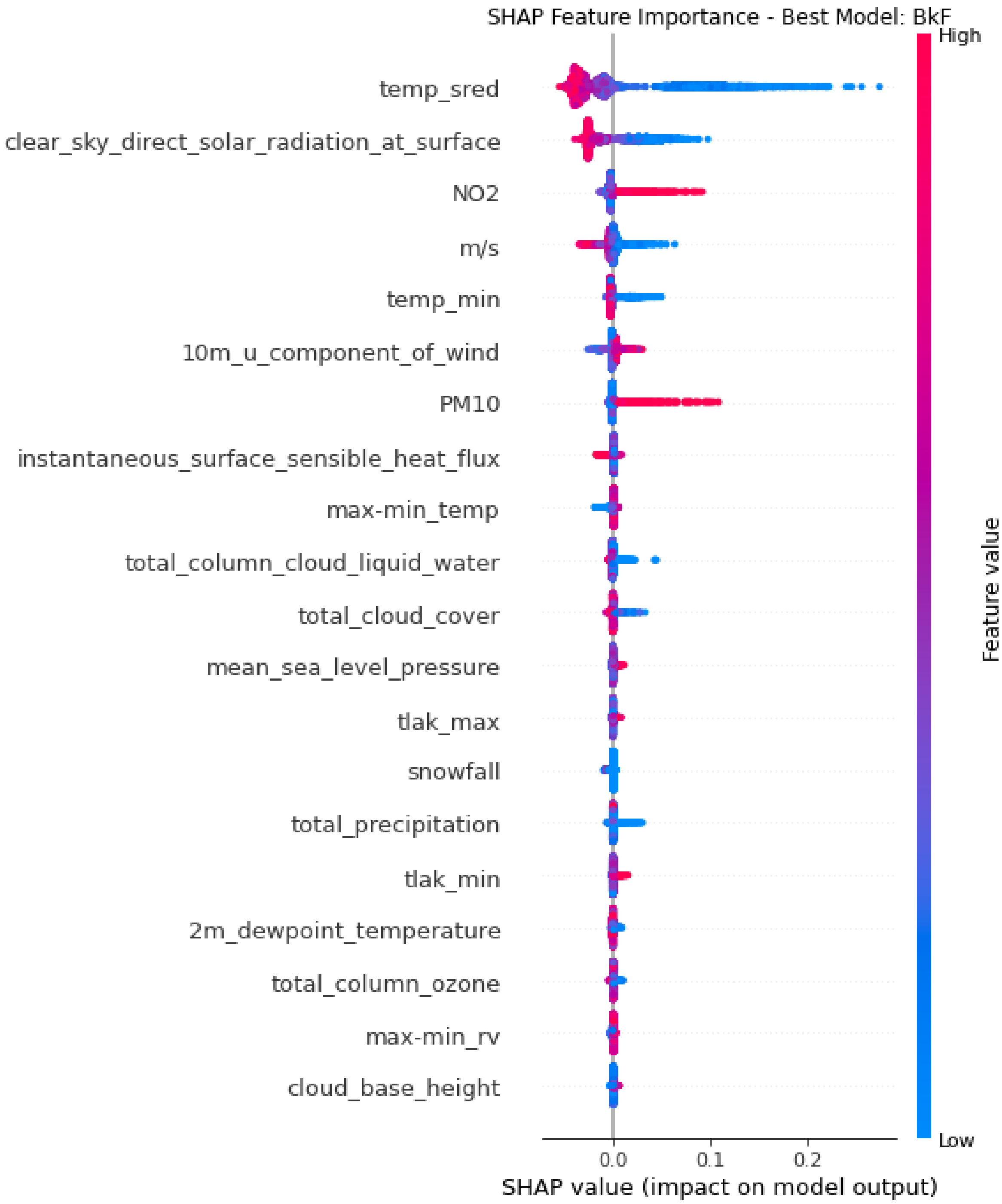


Figure S20. SHAP feature importance plot for BkF (best model)

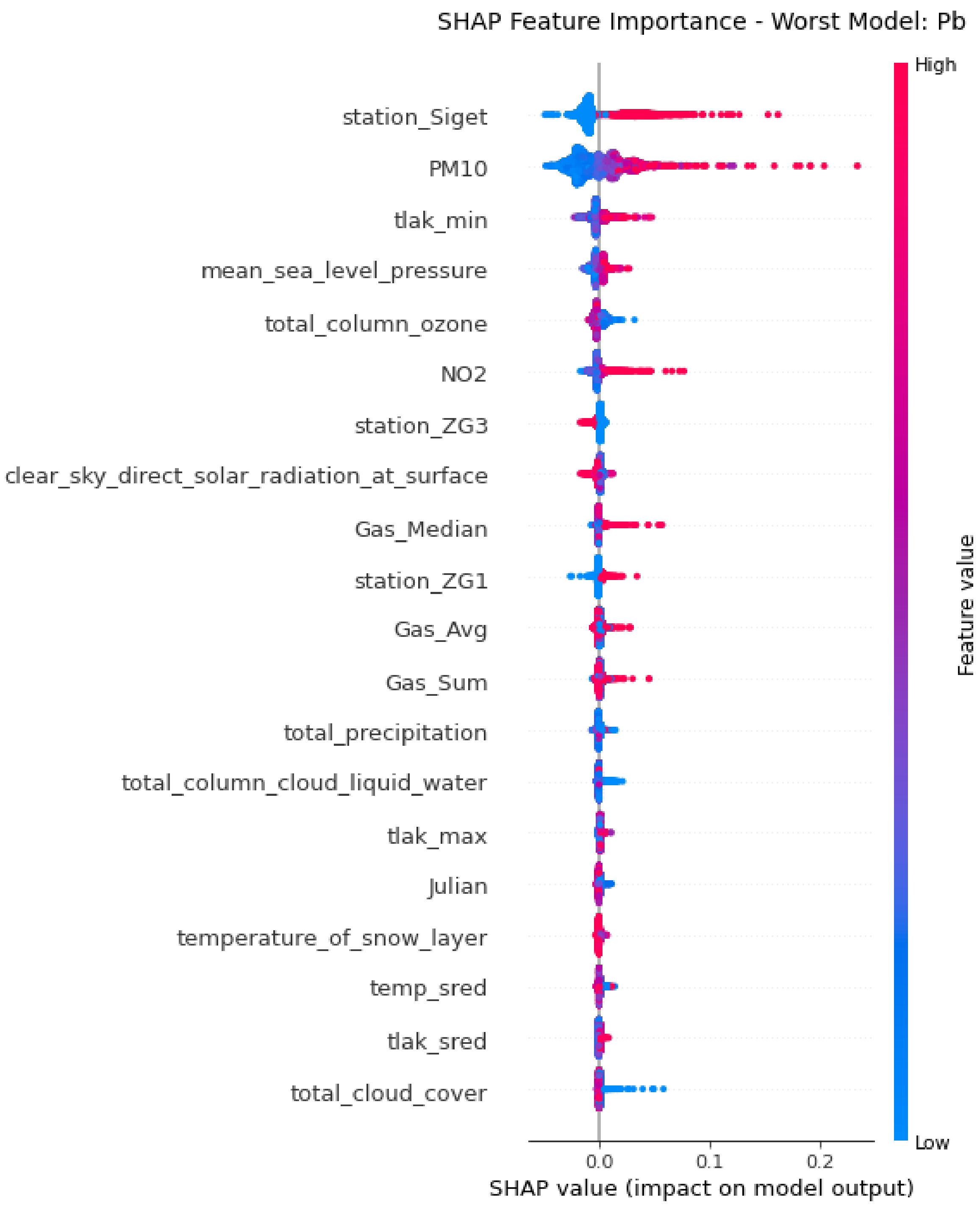


Figure S21. SHAP feature importance plot for Pb (worst model)

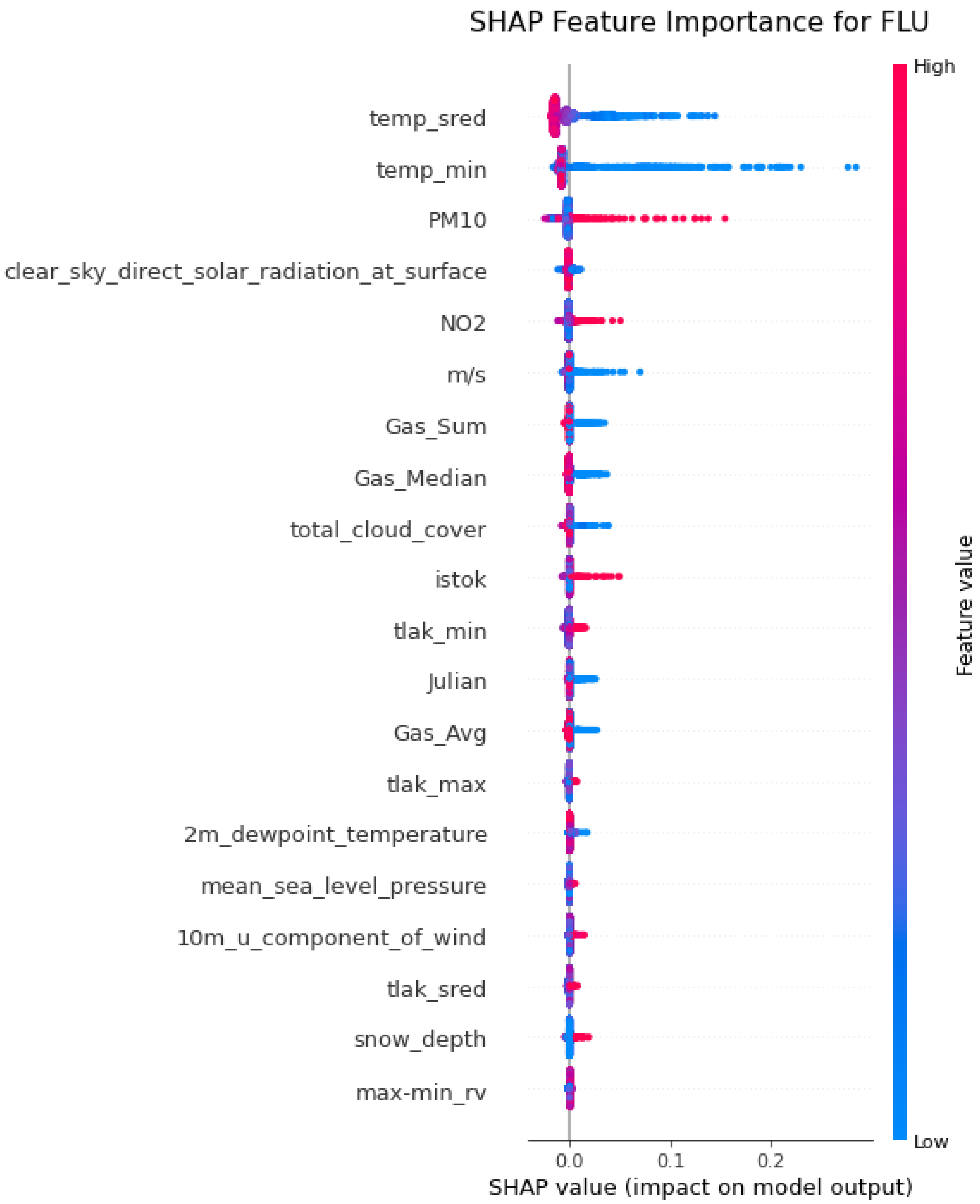


Figure S22. SHAP feature importance plot for Flu

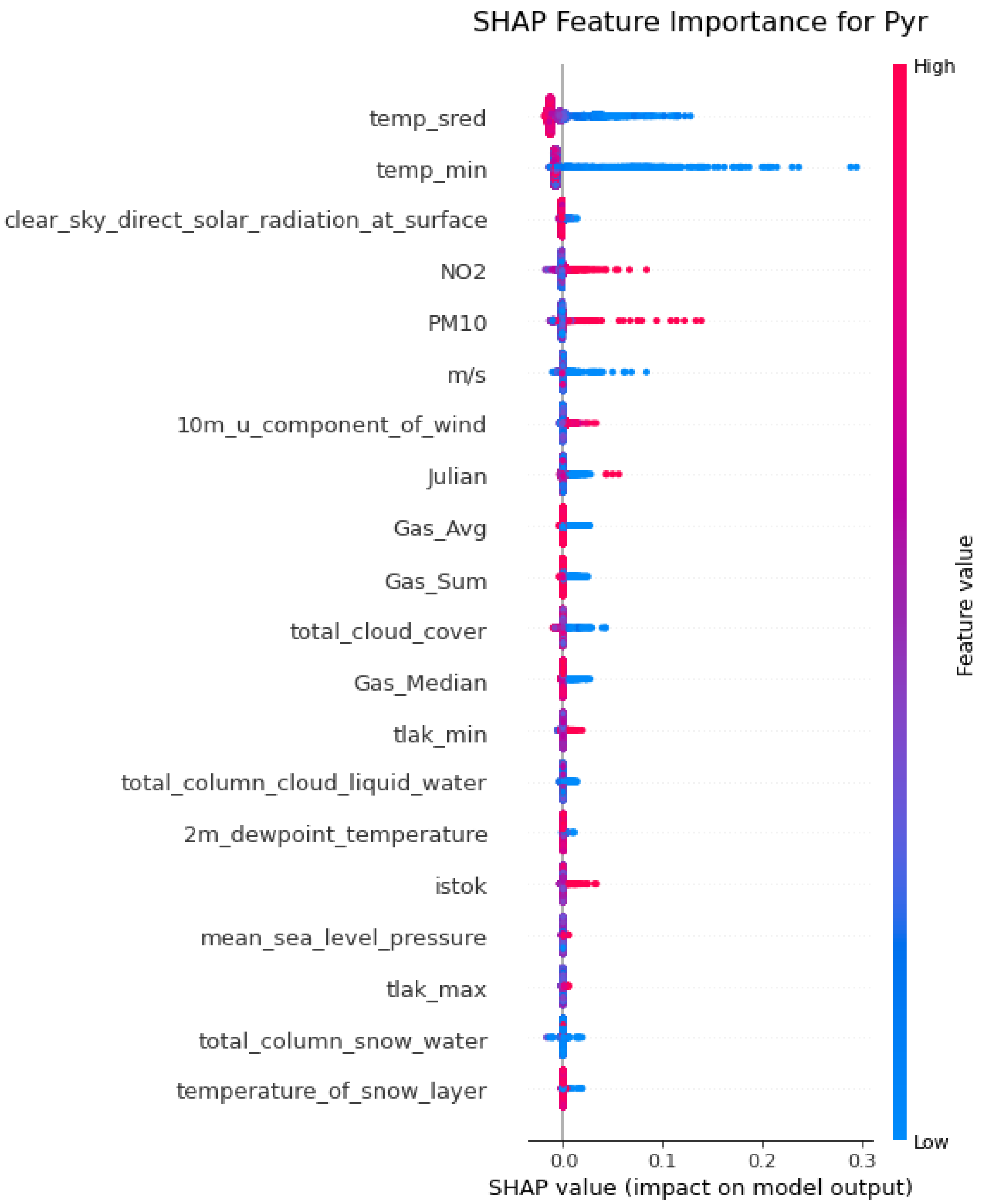


Figure S23. SHAP feature importance plot for Pyr

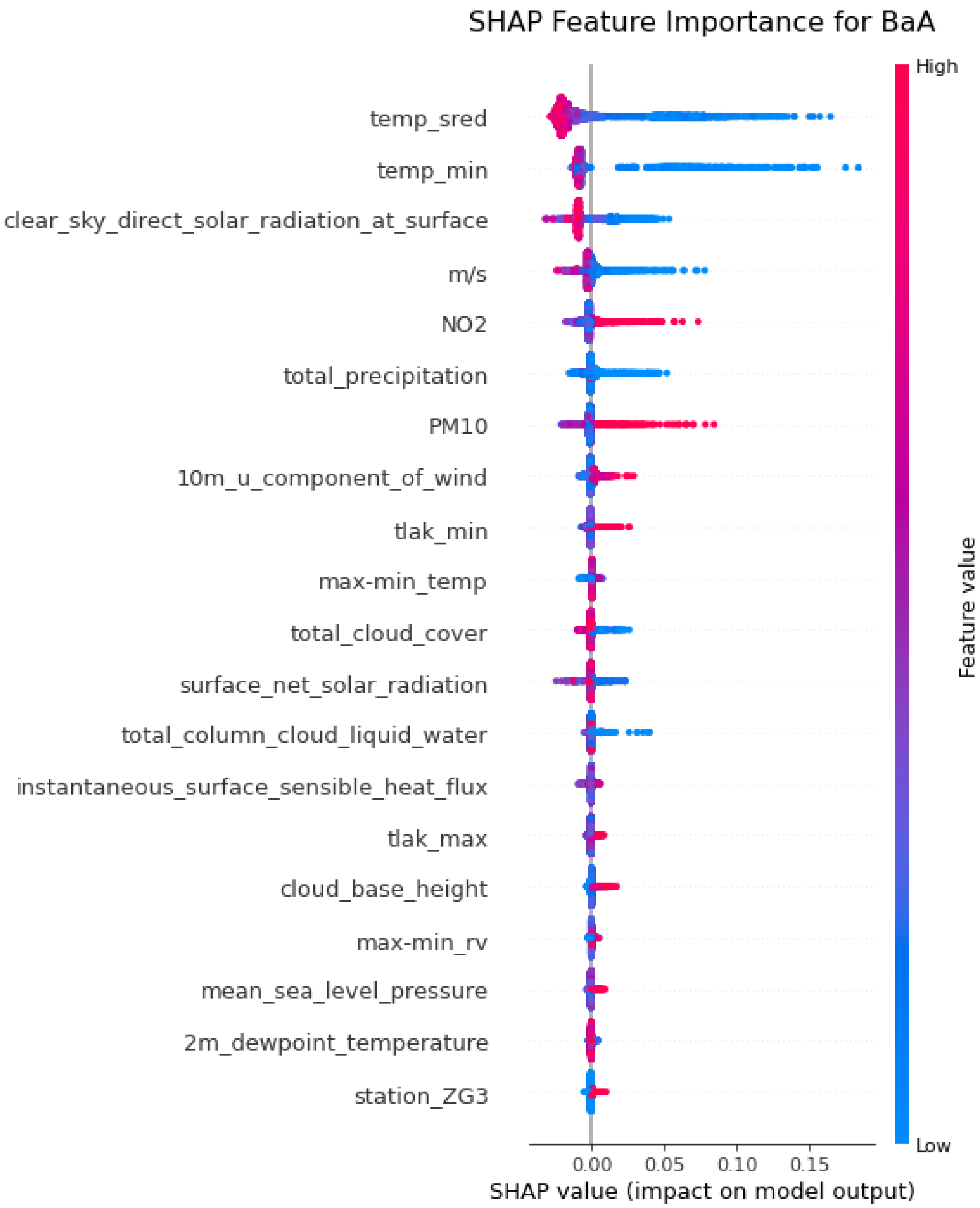


Figure S24. SHAP feature importance plot for BaA

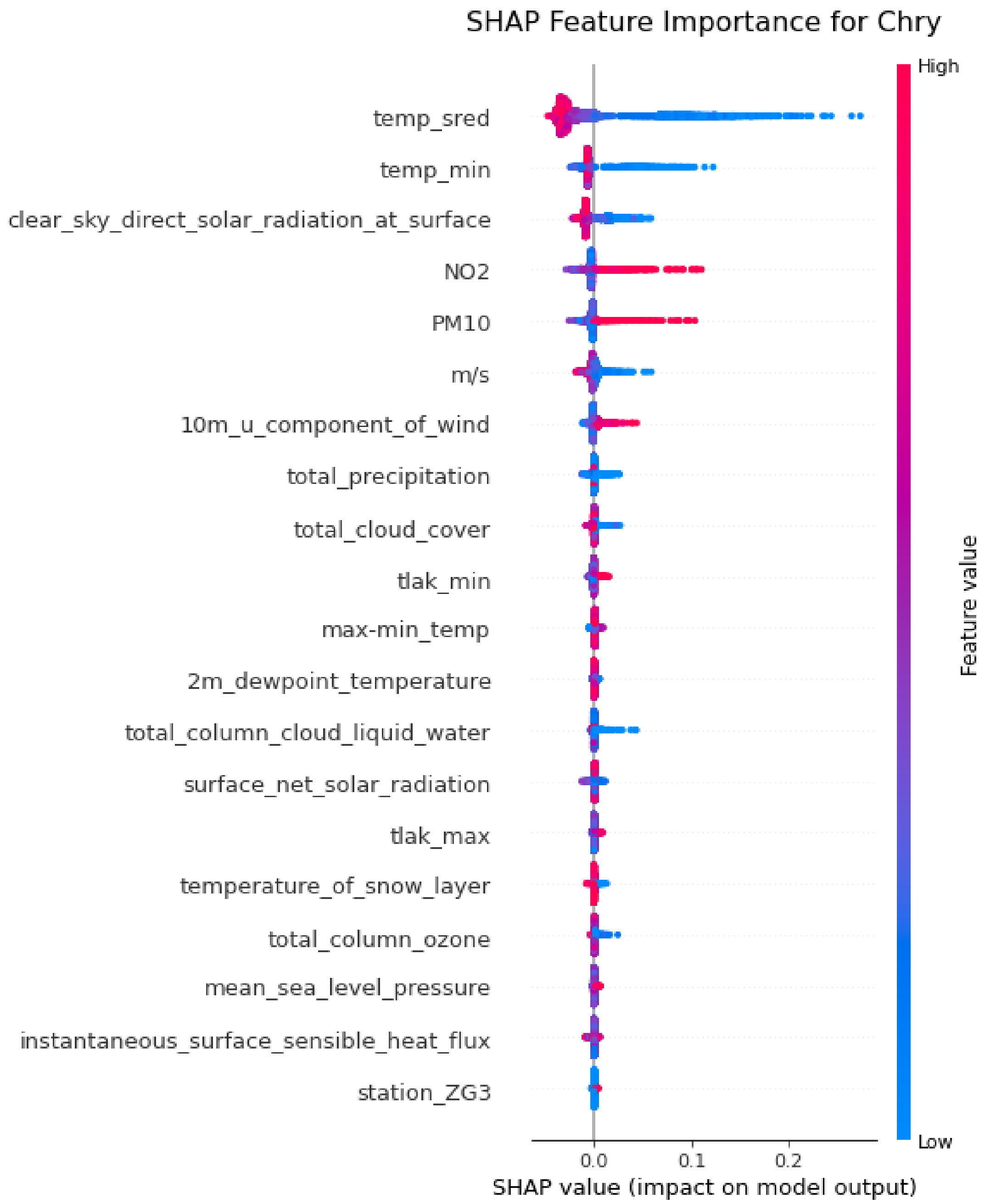


Figure S25. SHAP feature importance plot for Chry

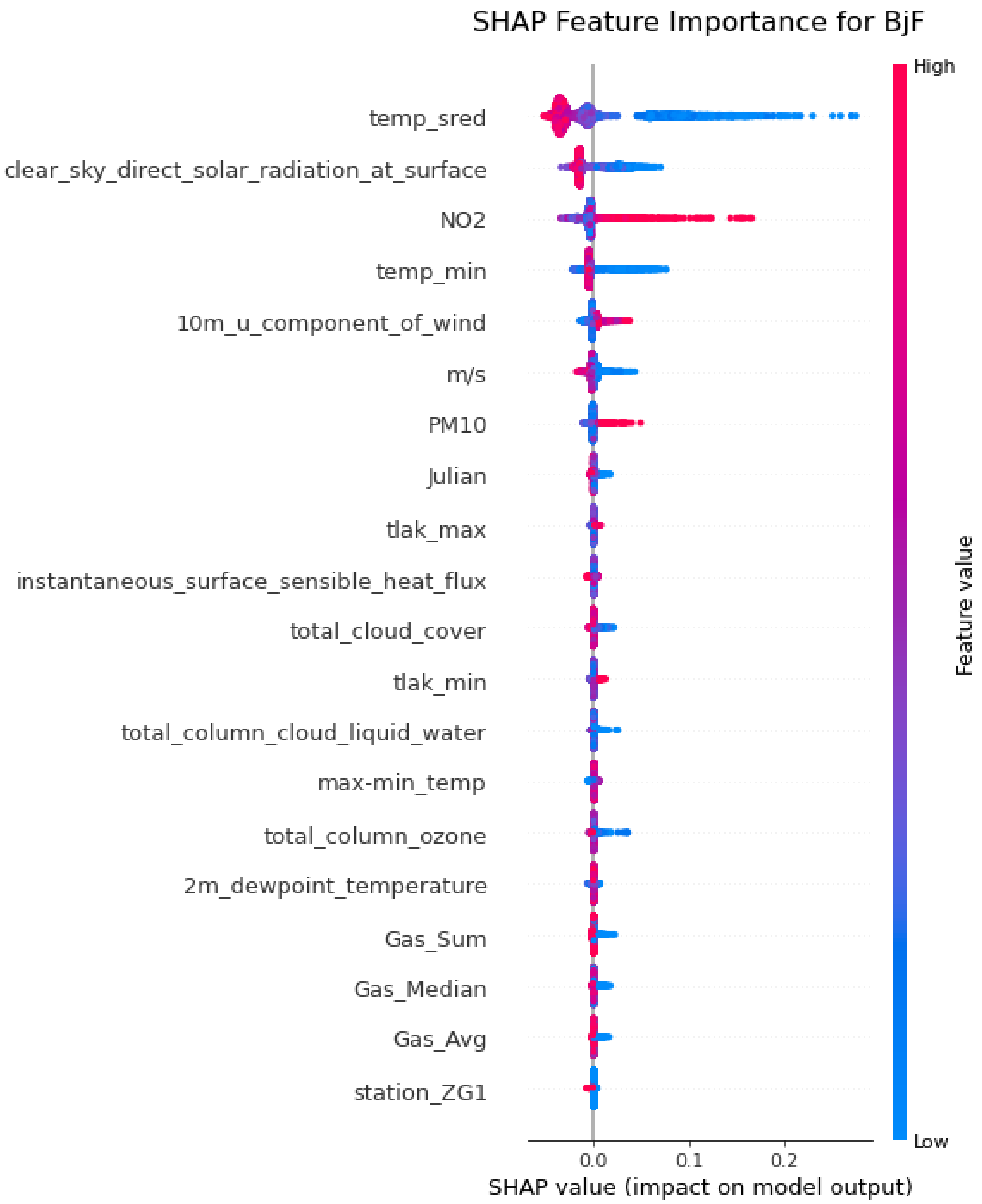


Figure S26. SHAP feature importance plot for BjF

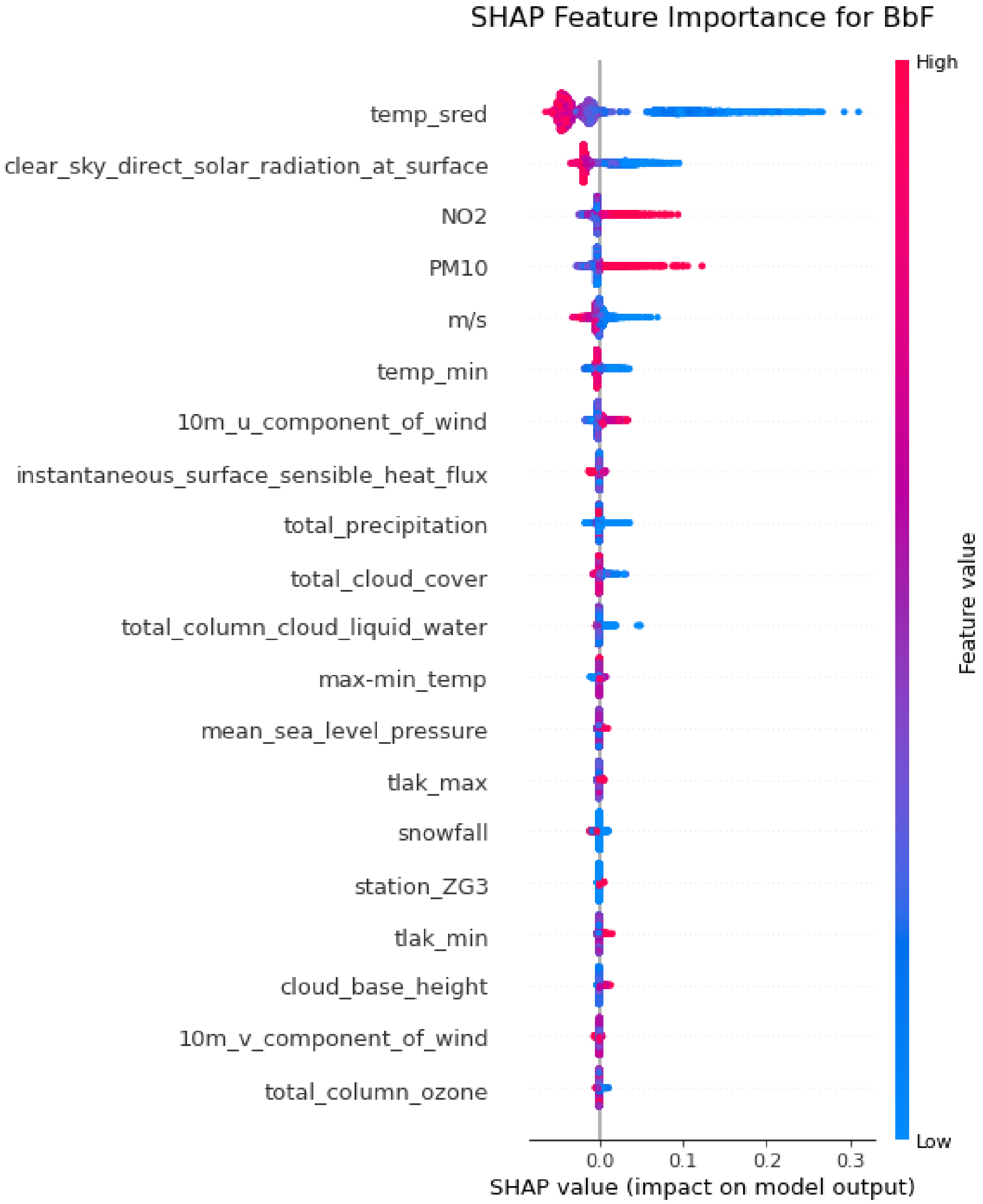


Figure S27. SHAP feature importance plot for BbF

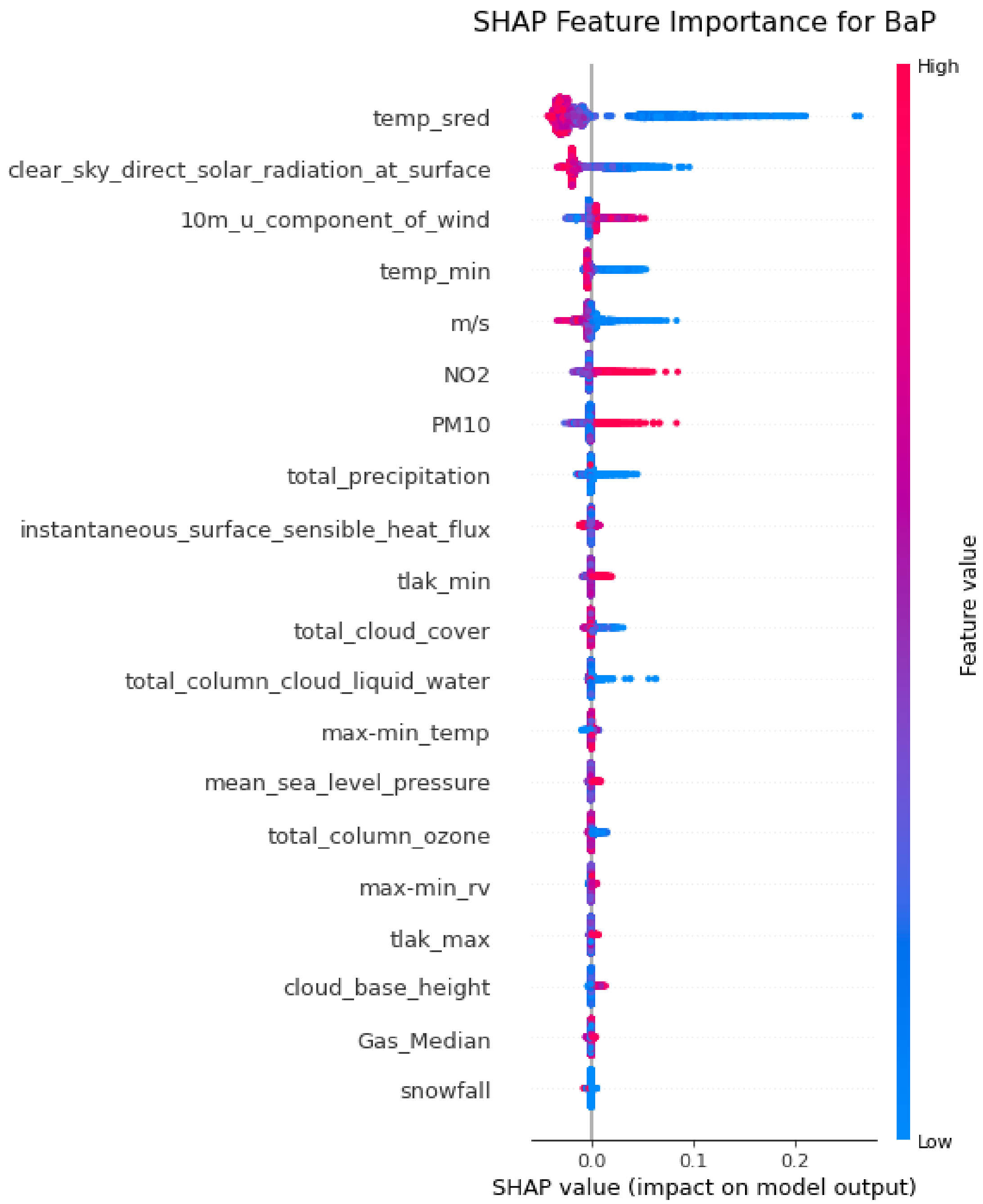


Figure S28. SHAP feature importance plot for BaP

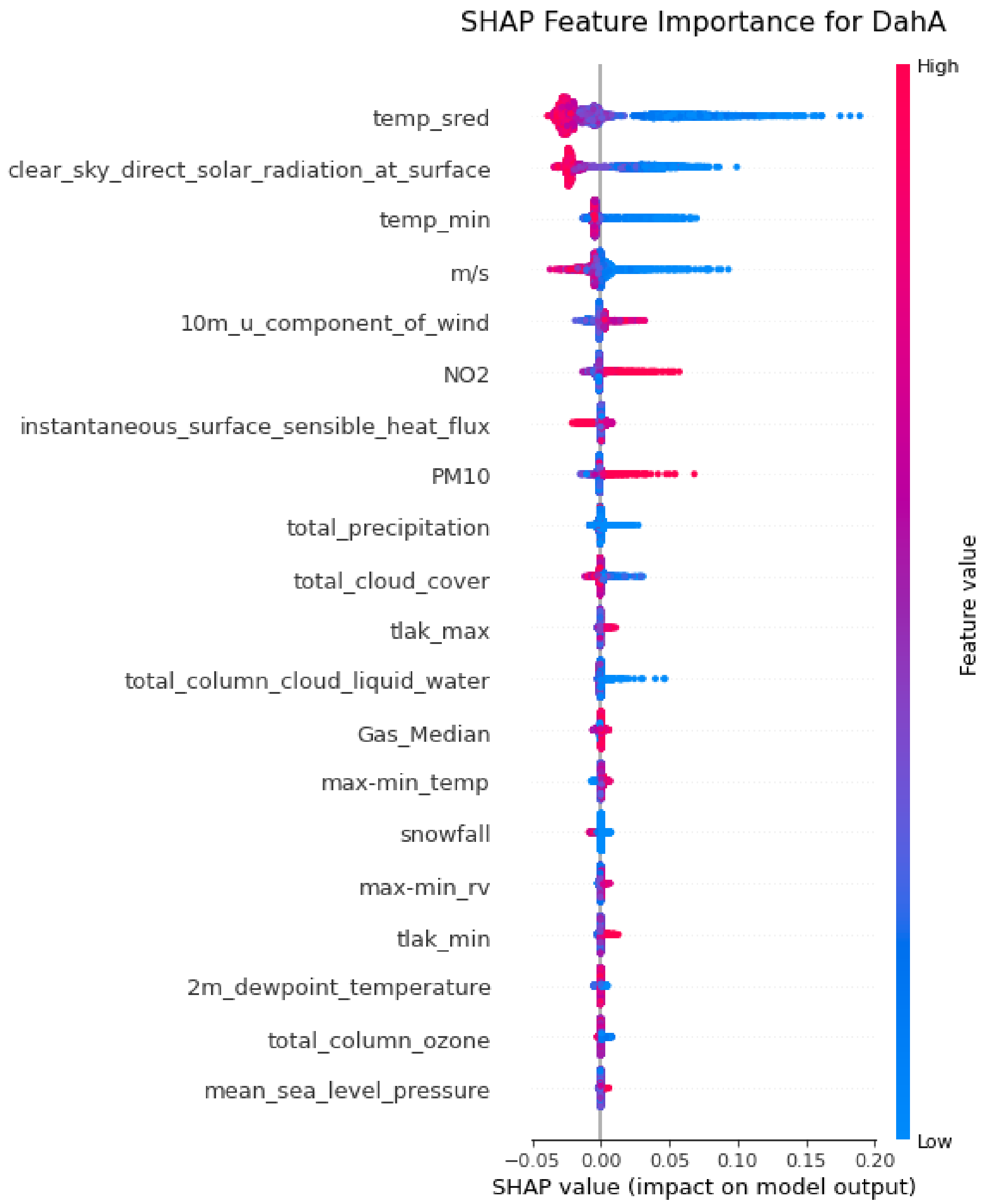


Figure S29. SHAP feature importance plot for DahA

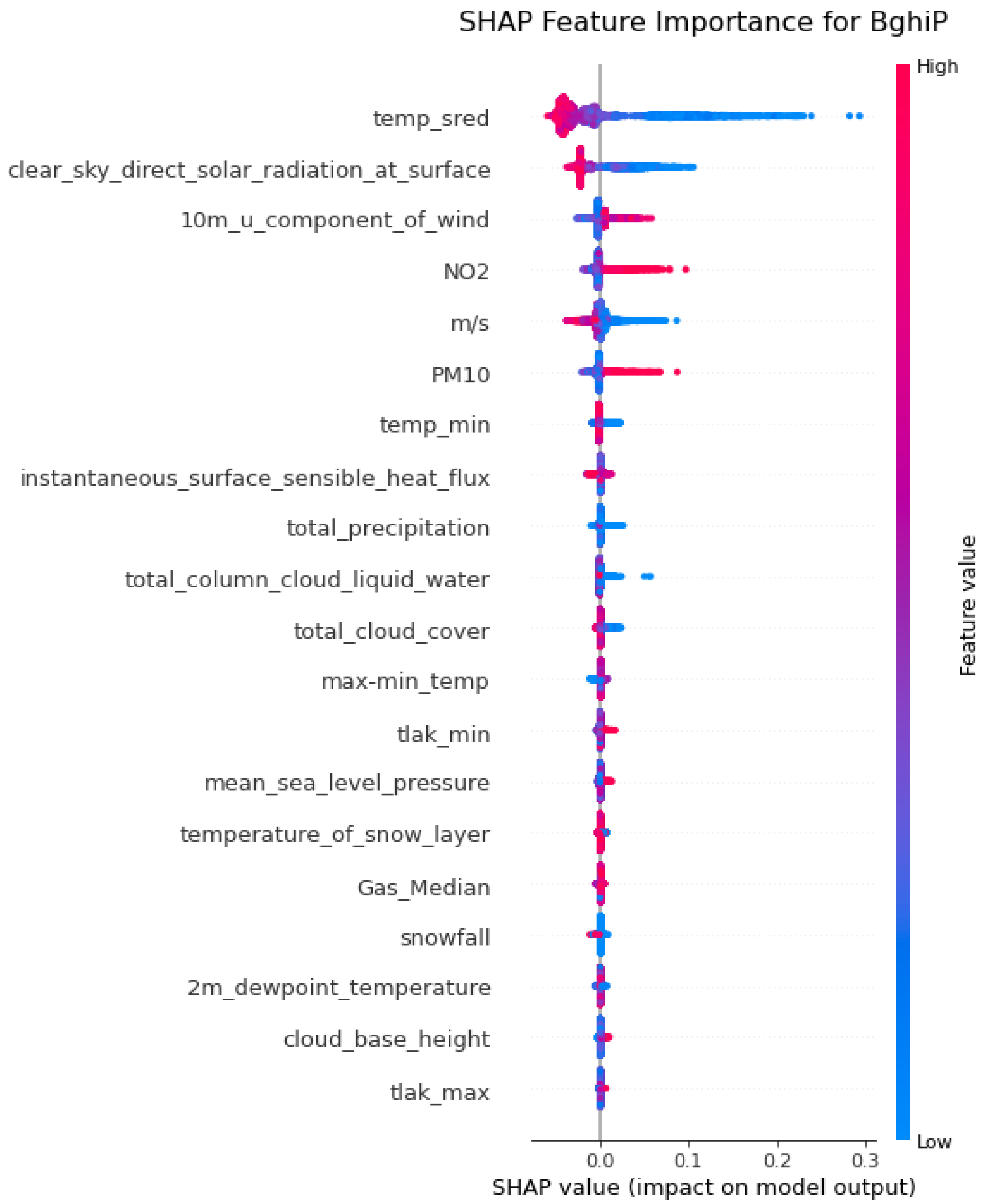


Figure S30. SHAP feature importance plot for BhgiP

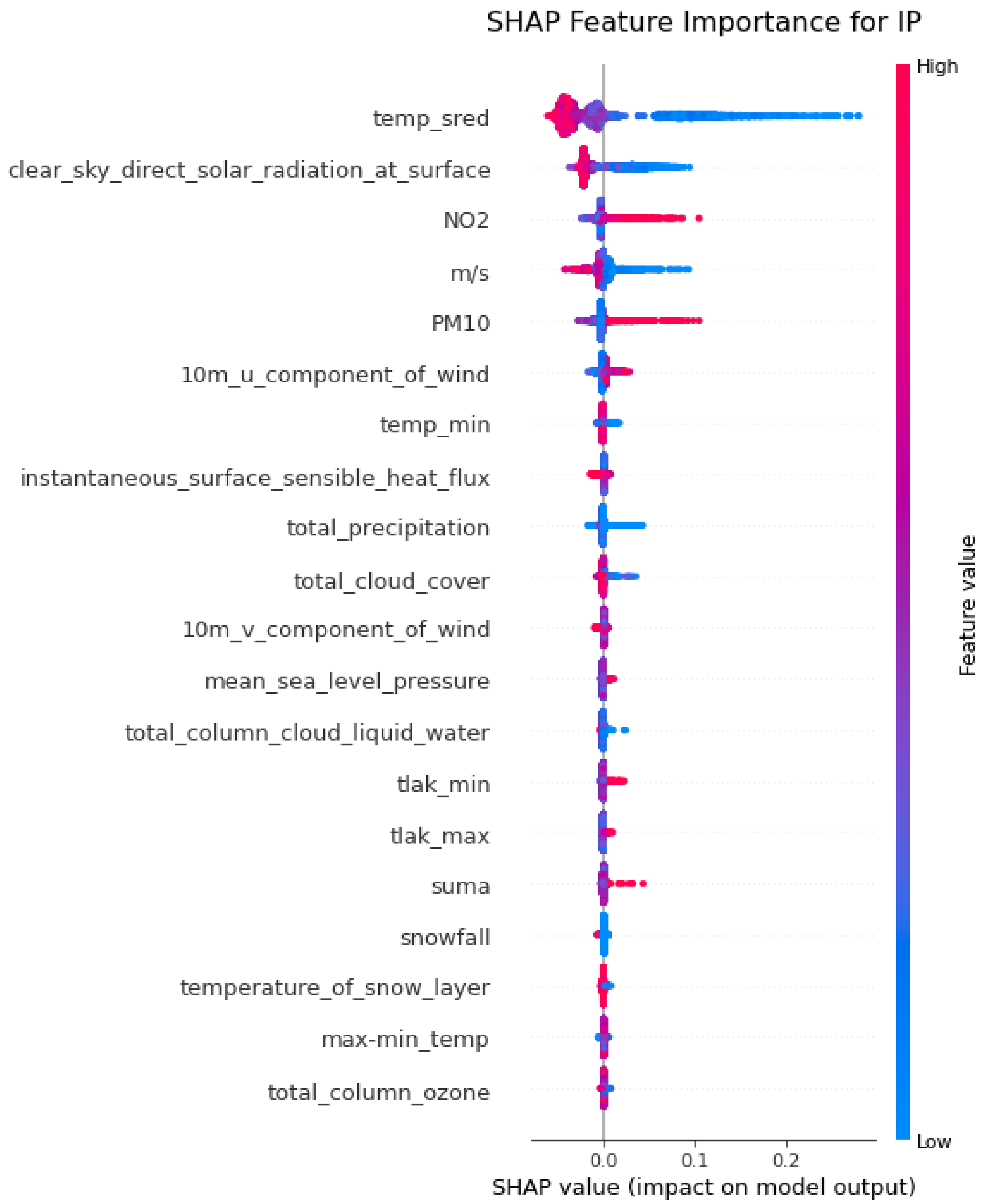


Figure S31. SHAP feature importance plot for IP

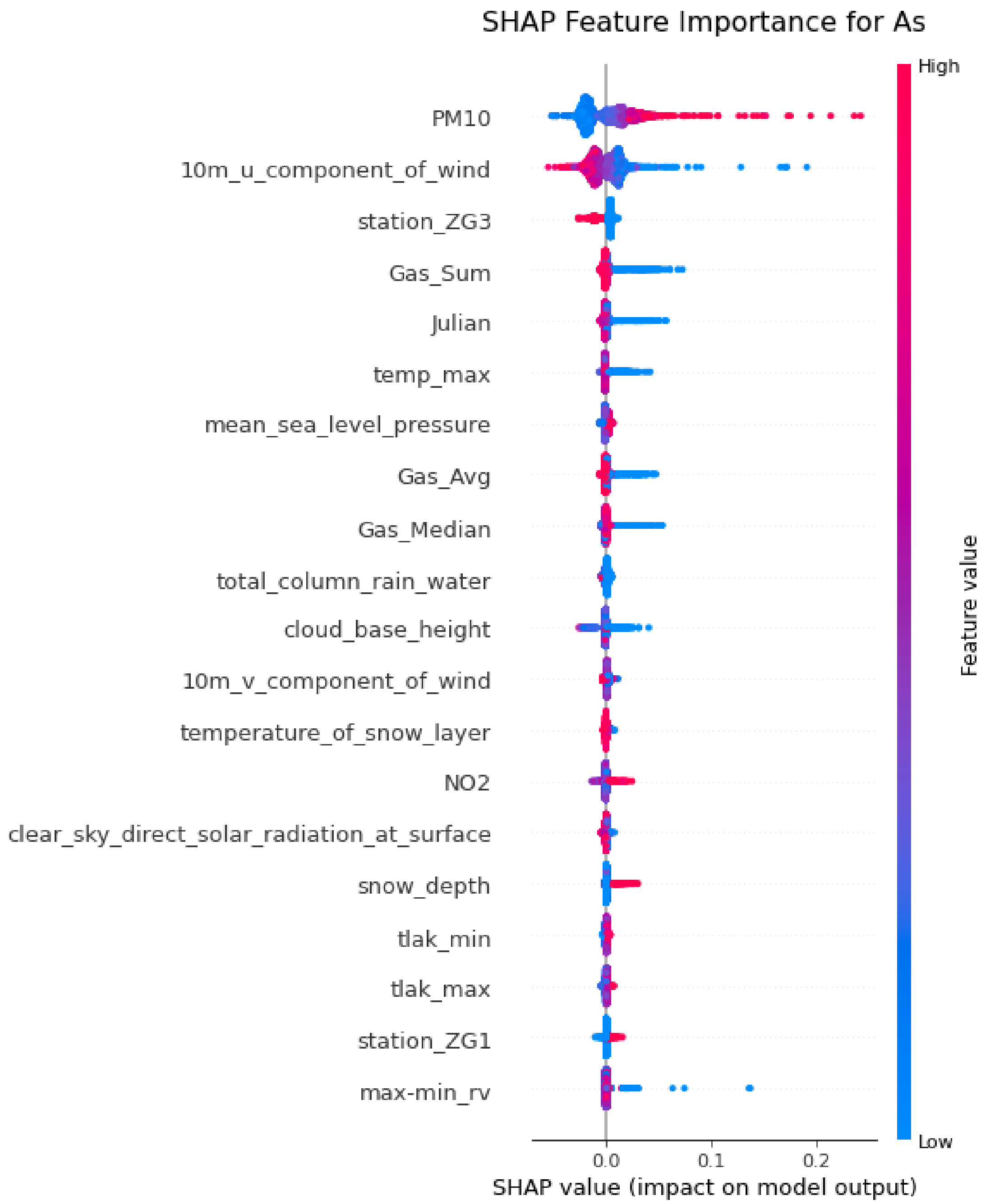


Figure S32. SHAP feature importance plot for As

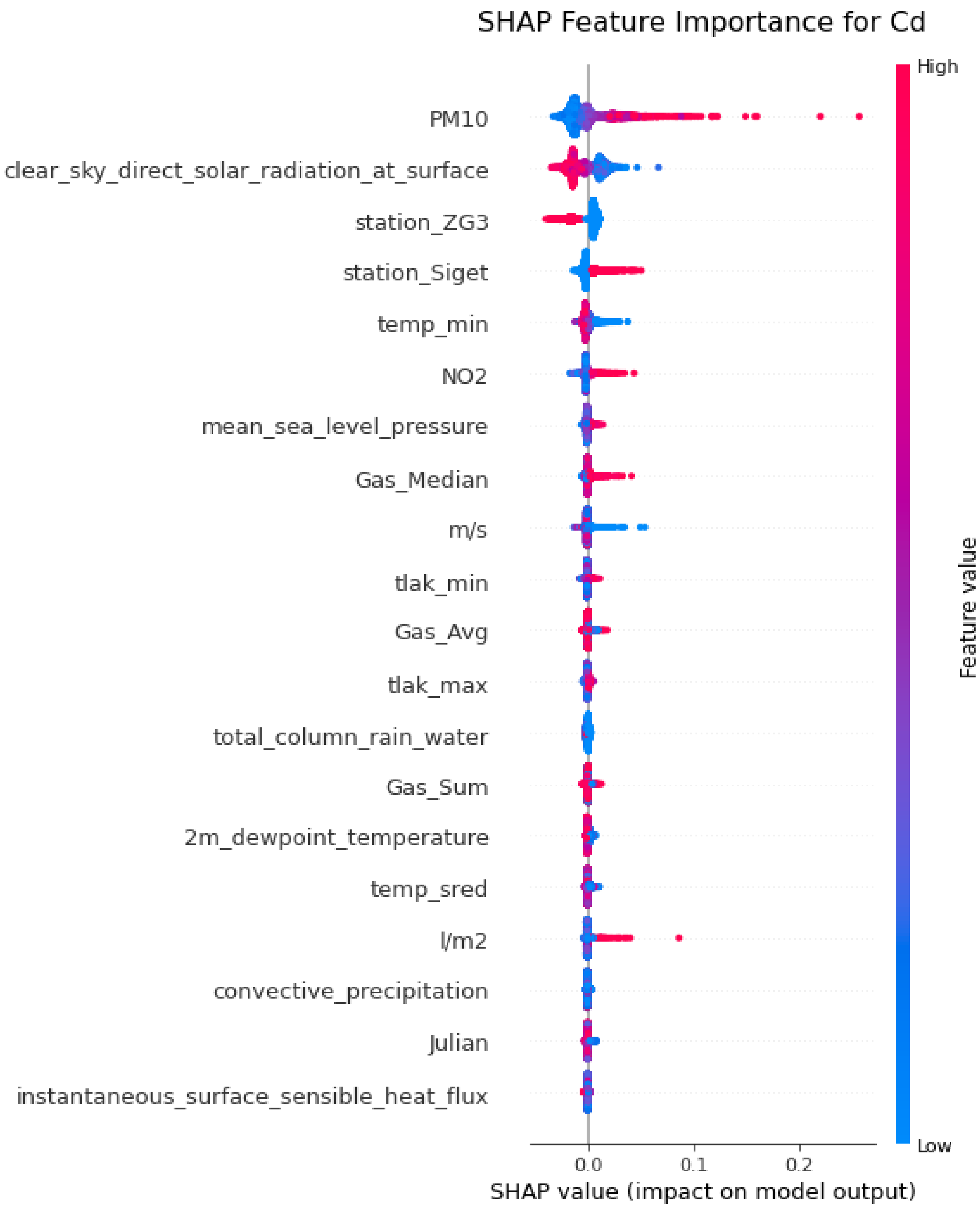


Figure S33. SHAP feature importance plot for Cd

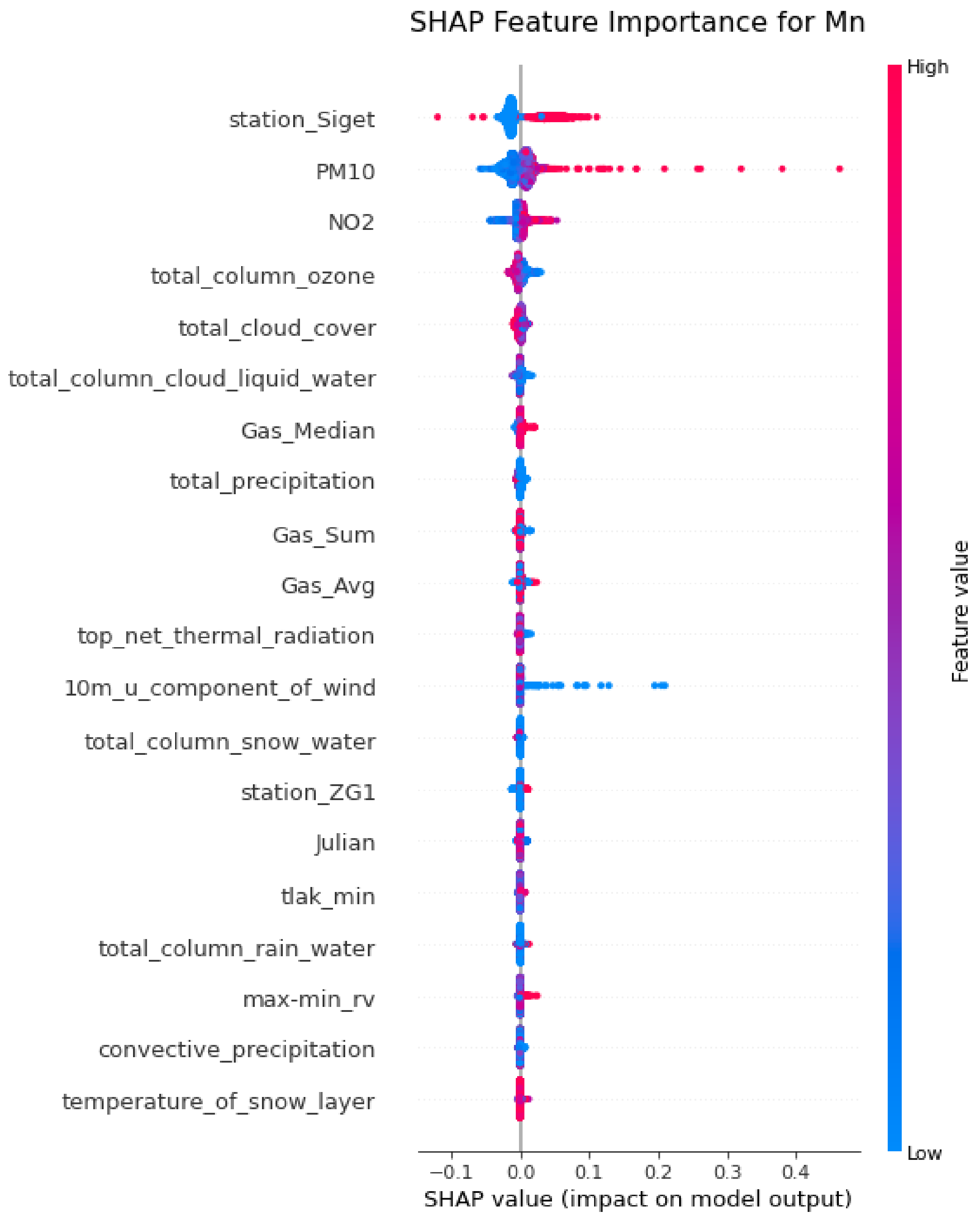


Figure S34. SHAP feature importance plot for Mn

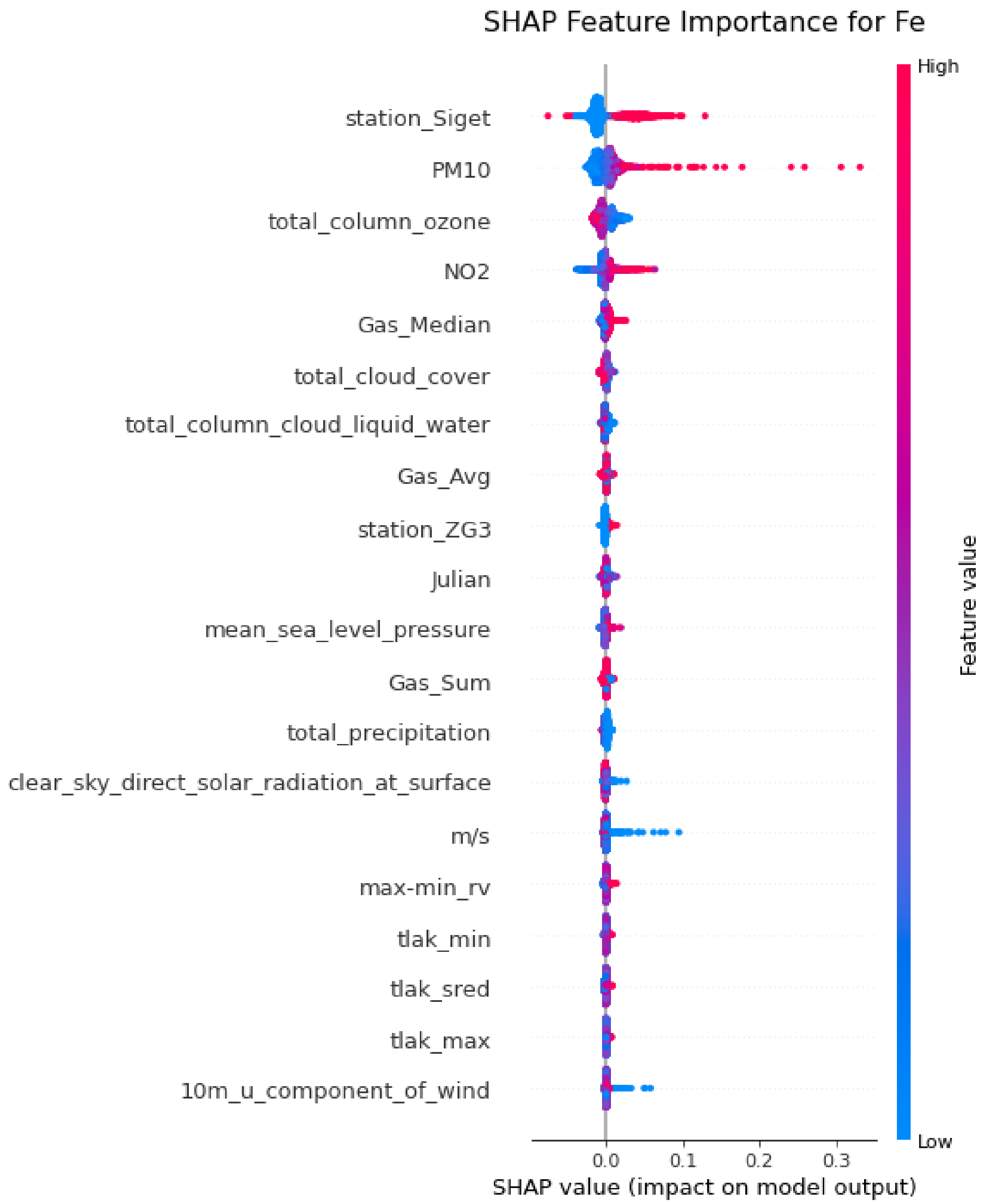


Figure S35. SHAP feature importance plot for Fe

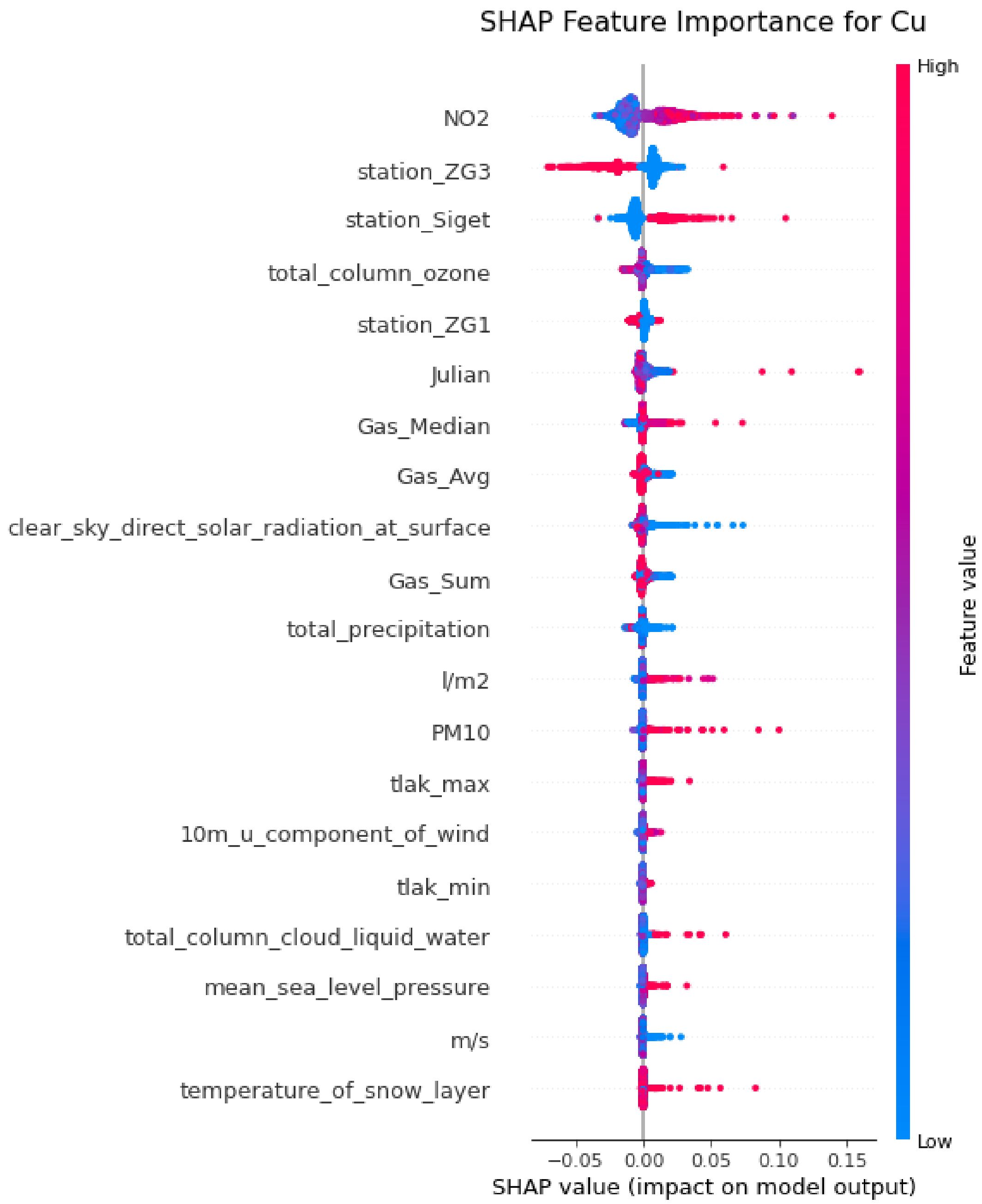


Figure SS36. SHAP feature importance plot for Cu

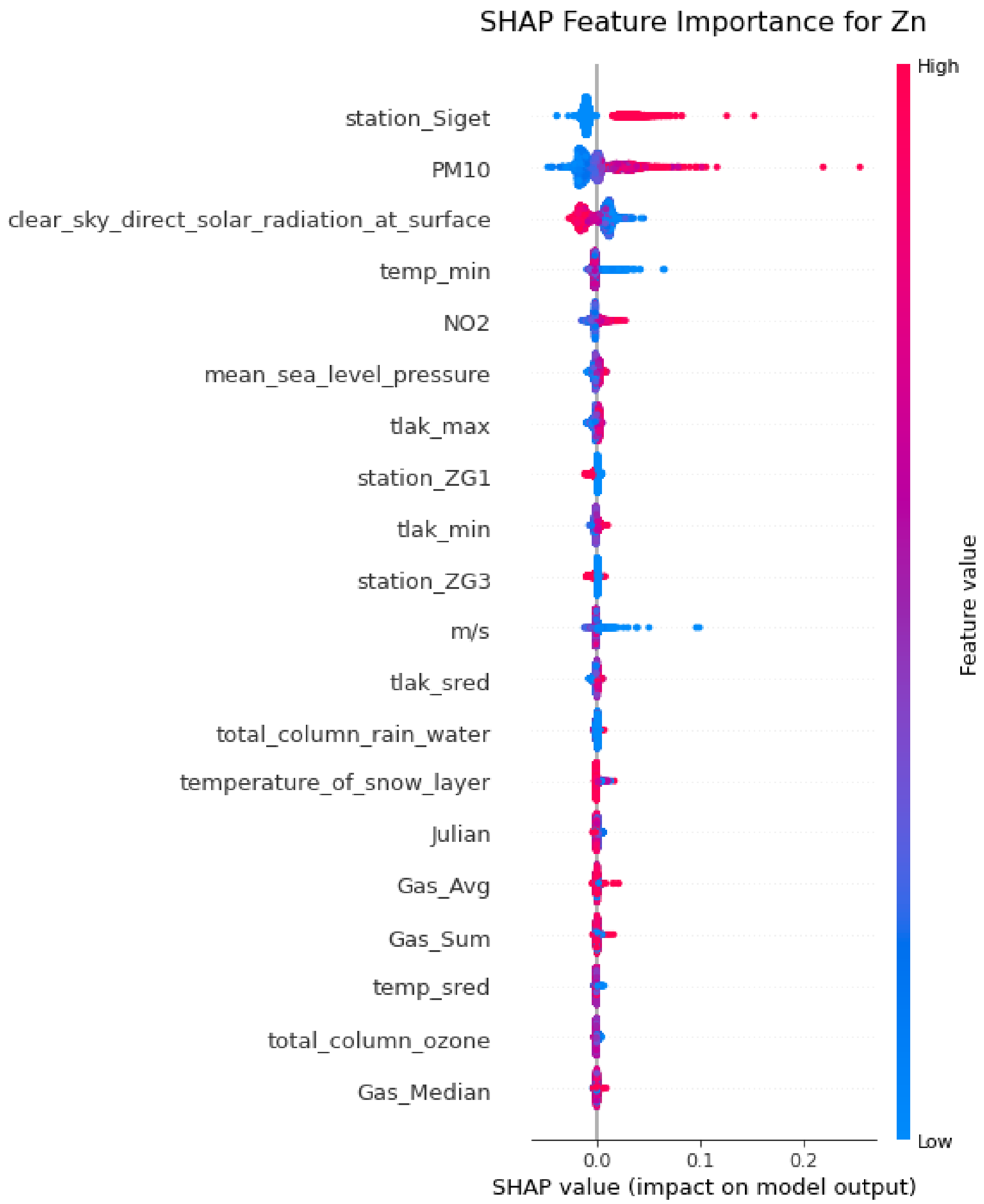


Figure S37. SHAP feature importance plot for Zn

Table S2. Results of feature importance rankings

Feature	As	BaA	BaP	BbF	BghiP	BjF	BkF	Cd	Chry	Cu	DahA	Flu	Fe	IP	Mn	Pb	Pyr	Zn
10m_u_component_of_wind	12.8	2.0	5.8	2.8	5.7	3.2	3.7	0.9	2.5	1.3	3.6	0.7	2.4	2.7	6.4	0.9	1.5	1.0
10m_v_component_of_wind	0.8	0.2	0.3	0.3	0.3	0.3	0.3	1.4	0.4	0.2	0.3	0.2	0.5	0.5	0.4	2.4	0.2	1.1
2m_dewpoint_temperature	0.4	0.5	0.5	0.4	0.5	0.5	0.5	0.6	0.6	0.3	0.7	0.4	0.4	0.2	1.2	0.6	0.5	1.0
Gas_Avg	4.9	0.4	0.1	0.2	0.1	0.9	0.2	2.2	0.5	4.8	0.2	1.3	3.3	0.2	3.0	3.1	1.8	1.3
Gas_Median	3.7	0.3	0.3	0.2	0.4	0.8	0.2	2.4	0.5	4.2	0.5	1.3	2.6	0.4	1.5	2.7	1.1	1.2
Gas_Sum	5.0	0.3	0.1	0.2	0.1	0.8	0.2	1.9	0.4	4.7	0.3	1.3	3.5	0.3	3.3	3.3	1.3	1.7
Julian	5.6	0.4	0.1	0.3	0.2	0.9	0.2	1.3	0.4	6.8	0.3	1.3	3.5	0.3	2.6	1.6	2.1	1.2
NO2	2.3	5.0	4.5	6.1	5.2	11.6	6.4	4.7	7.3	17.6	4.1	3.3	9.3	6.1	8.3	7.6	4.6	4.5
PM10	23.6	5.6	4.5	7.1	4.7	2.8	5.8	29.9	6.5	5.5	4.5	10.3	14.8	6.7	18.4	18.8	9.6	24.5
clear_sky_direct_solar_radiation_at_surface	1.5	5.8	11.1	9.0	12.3	8.2	7.8	12.2	4.1	3.5	17.7	1.3	2.4	12.1	0.7	1.6	1.6	9.7
cloud_base_height	3.9	0.7	0.8	0.5	0.6	0.5	0.5	0.7	0.5	0.9	0.8	0.3	0.9	0.3	0.7	1.0	0.4	0.3
convective_precipitation	0.4	0.0	0.0	0.1	0.1	0.1	0.1	0.6	0.1	0.5	0.1	0.0	0.4	0.1	0.9	0.3	0.1	0.8
convective_snowfall	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
instantaneous_surface_sensible_heat_flux	0.5	1.0	1.6	1.2	1.9	0.7	1.1	0.8	0.7	0.5	1.6	0.4	0.3	0.9	0.2	0.6	0.5	0.3
rugvica	0.4	0.5	0.6	0.4	0.5	0.3	0.5	0.3	0.3	0.2	0.5	1.9	0.5	0.6	0.4	0.8	1.1	0.2
l/m2	2.0	0.7	0.5	0.5	0.5	0.4	0.5	2.6	0.6	3.4	0.5	0.6	1.0	0.4	0.9	2.0	0.8	1.4
lucko	0.4	0.4	0.4	0.4	0.6	0.3	0.4	0.3	0.3	0.3	0.3	1.0	0.6	0.7	0.4	1.0	0.8	0.3
m/s	0.4	5.2	7.1	4.1	5.3	2.6	3.7	4.1	2.6	1.1	6.9	3.4	3.8	6.3	1.0	0.9	3.8	3.9
max-min_rv	2.6	0.5	0.3	0.2	0.3	0.2	0.3	0.2	0.4	0.4	0.3	0.4	1.4	0.2	1.6	1.0	0.3	0.2
max-min_temp	0.6	0.8	0.8	0.6	0.7	0.5	0.9	0.3	0.6	0.4	0.7	0.7	0.7	0.4	0.5	0.5	0.3	0.3
max-min_p	0.8	0.4	0.2	0.2	0.2	0.3	0.2	0.3	0.3	0.2	0.3	0.4	1.2	0.3	0.7	0.7	0.5	0.3
mean_sea_level_pressure	1.2	0.4	0.6	0.4	0.4	0.3	0.5	1.3	0.3	1.0	0.4	0.4	1.5	0.5	0.5	3.5	0.4	1.9
precipitation_type_encoded	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
rv_max	0.2	0.3	0.2	0.2	0.2	0.1	0.2	0.4	0.2	0.9	0.2	0.3	1.1	0.1	0.6	0.4	0.4	0.4
rv_min	0.4	0.2	0.1	0.2	0.2	0.2	0.1	0.2	0.3	0.1	0.2	0.3	0.3	0.2	0.3	1.7	0.1	0.2
rv_sred	0.2	0.3	0.2	0.2	0.2	0.4	0.2	0.2	0.4	0.5	0.2	0.4	0.3	0.2	0.3	0.9	0.5	0.4
snow_density	0.1	0.2	0.4	0.2	0.3	0.3	0.2	0.1	0.3	0.2	0.2	0.3	0.4	0.3	0.6	0.2	0.2	0.3
snow_depth	2.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.3	0.1	0.7	0.6	0.1	0.2	0.2	0.5	0.5
snowfall	0.4	0.4	0.5	0.7	0.5	0.5	0.6	0.2	0.6	0.0	0.7	0.3	0.3	0.4	0.1	0.2	0.3	0.2
snowmelt	1.0	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.1	0.2	0.0	0.2	0.3	0.4
station_IMI	2.3	0.4	0.3	0.3	0.3	0.3	0.1	1.6	0.3	0.4	0.2	0.5	1.2	0.5	0.2	0.1	0.7	2.3
station_Siget	0.5	0.4	0.3	0.4	0.4	0.3	0.3	3.7	0.4	6.3	0.6	1.1	16.4	0.4	21.5	12.9	1.2	12.9
station_ZG1	1.3	0.3	0.2	0.2	0.2	0.3	0.2	0.9	0.2	5.3	0.1	0.2	0.8	0.1	3.6	2.0	0.3	1.4
station_ZG3	2.6	0.2	0.1	0.1	0.1	0.0	0.1	6.3	0.2	10.2	0.1	0.1	2.6	0.0	0.3	1.6	0.1	1.3
sum	0.5	0.6	0.4	0.3	0.3	0.4	0.3	0.3	0.5	0.3	0.5	0.4	0.2	0.7	0.2	0.5	0.6	1.7
surface_net_solar_radiation	0.3	2.4	1.2	1.0	0.6	0.5	0.8	0.4	0.8	0.5	1.0	0.6	0.6	0.6	1.1	0.4	0.5	0.2
surface_net_thermal_radiation	2.2	0.5	0.5	0.4	0.2	0.5	0.3	0.8	0.4	0.2	0.6	0.2	0.7	0.4	0.5	2.1	0.5	2.5

temp_max	3.9	0.4	0.4	0.3	0.4	0.4	0.3	0.3	0.3	0.6	0.3	0.3	0.7	0.3	0.4	1.3	0.3	0.7
temp_min	0.3	25.1	6.0	3.2	2.1	9.3	4.5	5.5	13.7	1.1	10.3	35.7	0.6	1.8	0.5	0.6	33.7	5.7
temp_avg	0.6	24.9	38.7	49.9	45.9	44.2	51.3	1.8	43.8	0.3	30.9	17.0	0.3	46.6	0.3	1.2	15.0	0.6
temperature_of_snow_layer	0.9	0.7	0.3	0.5	0.5	0.6	0.4	0.9	0.7	2.9	0.5	0.8	1.6	0.4	1.3	0.8	1.0	1.3
p_max	0.9	0.6	0.3	0.4	0.4	0.5	0.4	0.5	0.6	1.8	0.7	0.6	0.6	0.3	0.4	1.1	0.5	1.3
p_min	0.5	1.1	1.6	0.6	0.8	0.5	0.6	1.4	0.7	1.0	0.6	0.9	1.1	0.6	0.5	4.8	1.0	1.5
p_avg	0.7	0.3	0.3	0.2	0.3	0.2	0.2	0.4	0.2	0.4	0.2	0.3	0.8	0.2	0.3	1.0	0.5	1.0
top_net_thermal_radiation	0.4	0.4	0.3	0.3	0.3	0.2	0.3	0.5	0.3	0.4	0.4	0.4	0.5	0.2	1.0	1.2	0.3	0.4
total_cloud_cover	0.3	1.9	1.6	1.4	1.1	0.8	1.1	0.4	0.9	0.4	2.1	2.2	2.3	1.4	2.5	1.9	2.6	0.4
total_column_cloud_liquid_water	0.6	0.9	1.2	0.8	0.9	0.6	0.9	0.6	1.1	1.6	1.0	1.1	1.6	0.7	2.0	1.7	1.6	0.6
total_column_ozone	1.0	0.7	0.7	0.5	0.6	1.1	0.6	1.2	1.0	3.5	0.5	0.8	6.2	0.4	4.3	2.4	0.9	0.9
total_column_rain_water	0.8	0.0	0.1	0.1	0.1	0.1	0.1	0.3	0.1	0.2	0.1	0.0	0.4	0.1	1.1	0.5	0.0	0.7
total_column_snow_water	0.4	0.5	0.3	0.1	0.3	0.3	0.2	0.4	0.3	0.5	0.3	1.4	0.5	0.2	0.6	1.0	1.4	0.3
total_precipitation	0.3	4.5	2.8	2.0	1.6	0.5	1.6	0.2	1.5	1.7	2.4	0.5	1.4	2.2	1.3	1.2	0.5	0.8
sveta_helena	0.5	0.3	0.3	0.2	0.2	0.3	0.2	0.8	0.3	0.3	0.4	1.2	0.5	0.3	0.3	1.0	1.2	2.0

Feature description:

10m_u_component_of_wind: Horizontal wind component (east-west) at 10 meters above ground level (m/s).

10m_v_component_of_wind: Horizontal wind component (north-south) at 10 meters above ground level (m/s).

2m_dewpoint_temperature: Dew point temperature at 2 meters above ground level (°C), reflects moisture content.

Gas_Avg / Gas_Median / Gas_Sum: Average, median, or sum of gas consumption (proxy for heating activity).

Julian: Day of the year (1–365), a proxy for seasonality.

NO2: Nitrogen dioxide concentration (µg/m³), indicator of traffic and combustion emissions.

PM10: Particulate matter with diameter ≤10 µm (µg/m³)

clear_sky_direct_solar_radiation_at_surface: Solar radiation at ground level without cloud interference (W/m²)

cloud_base_height: Height of the cloud base above ground level (m), indicator of atmospheric mixing.

convective_precipitation / convective_snowfall :Precipitation or snowfall resulting from convective processes (mm), influencing wet deposition.

instantaneous_surface_sensible_heat_flux: Upward flux of heat from the surface to the atmosphere (W/m²), reflects land-atmosphere energy exchange.

l/m2: Liquid water equivalent precipitation (liters per m²).

lucko / rugvica / sveta helena / sum: Trafiic in main highway access (see fig 1 in paper)

m/s: Wind speed (meters per second).

max-min_rv / max-min_temp / max-min_tlak (p): Daily variability (range) of relative humidity, temperature, and pressure.

mean_sea_level_pressure: Atmospheric pressure at sea level (hPa), used to identify stagnant air conditions.

precipitation_type_encoded: Categorical variable coding precipitation types (e.g., rain, snow, mixed).

rv_max / rv_min / rv_sred: Maximum, minimum, and average daily relative humidity (%).

snow_density / snow_depth / snowfall / snowmelt: Snow physical properties and melting processes, affect pollutant deposition.

station_IMI / station_Siget / station_ZG1 / station_ZG3: Dummy variables representing the monitoring stations (location-based variability).

surface_net_solar_radiation / surface_net_thermal_radiation: Net solar and thermal radiation at the surface (W/m^2).

temp_max / temp_min / temp_avg: Maximum, minimum, and average daily temperature (°C).

temperature_of_snow_layer: Temperature inside the snowpack (°C).

top_net_thermal_radiation: Net thermal radiation at the top of the atmosphere (W/m^2).

total_cloud_cover: Fraction of the sky covered by clouds (0–1).

total_column_cloud_liquid_water / total_column_ozone / total_column_rain_water / total_column_snow_water: Integrated column amounts (kg/m^2) of clouds, ozone, rain, and snow in the atmosphere.

total_precipitation: Total precipitation accumulated over a day (mm).

5.6.2 Supplementary materials with publication 3:

SUPPLEMENTARY MATERIAL

Linking atmospheric and soil contamination: a comparative study of PAHs and metals in PM_{10} and surface soil near urban monitoring stations

Nikolina Račić[1,2], Stanko Ružičić[3], Gordana Pehnec[1], Ivana Jakovljević[1], Zdravka Sever Štrukil[1], Jasmina Rinkovec[1], Silva Žužul[1], Iva Smoljo[1], Mario Lovrić[4,5,2*]

[1]*Institute for Medical Research and Occupational Health, Ksaverska cesta 2, 10000 Zagreb, Croatia*
[2]*The Lisbon Council, Rue de la Loi 155, 1040 Brussels, Belgium*
[3]*Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva 6, 10000 Zagreb, Croatia*
[4]*Institute for Anthropological Research, Gajeva 32, 10000 Zagreb, Croatia*
[5]*Josip Juraj Strossmayer University of Osijek, Faculty of Food Technology Osijek, Franje Kuhača 20, 31000 Osijek, Croatia*
Corresponding author: Mario Lovrić (mario.lovric@inantro.hr)

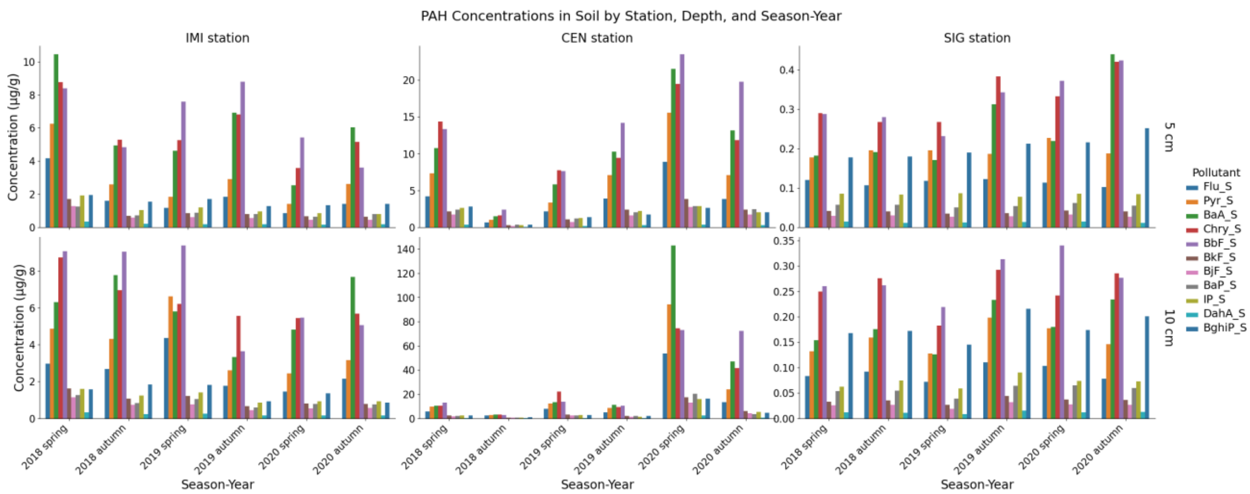


Figure S1. PAH concentrations in soil by season, depth and station

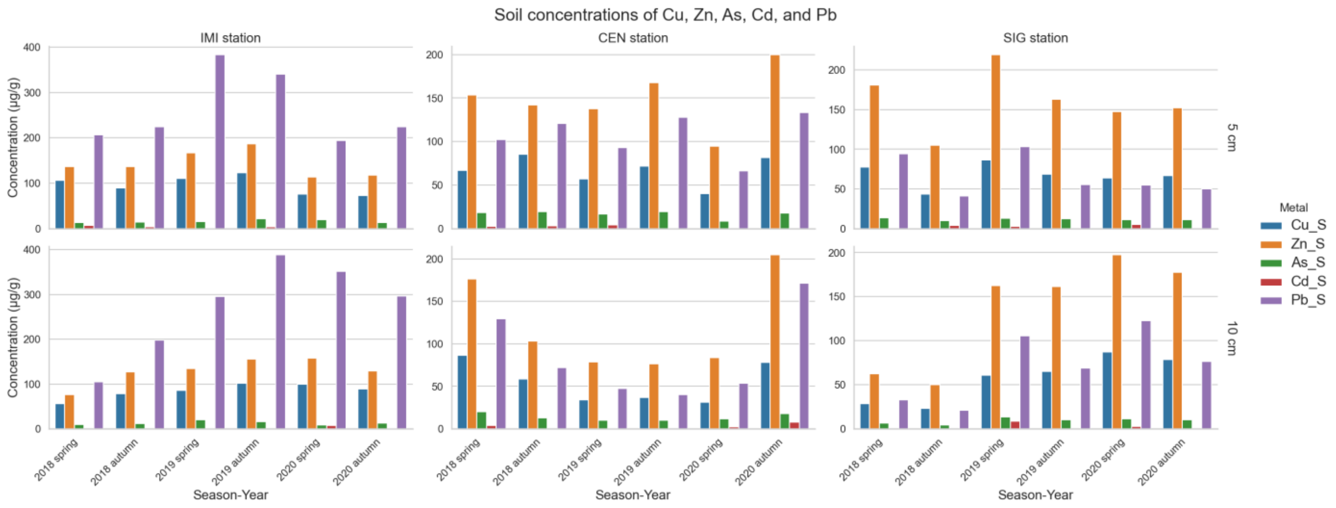


Figure S2. Metals concentration in soil by season, depth and station

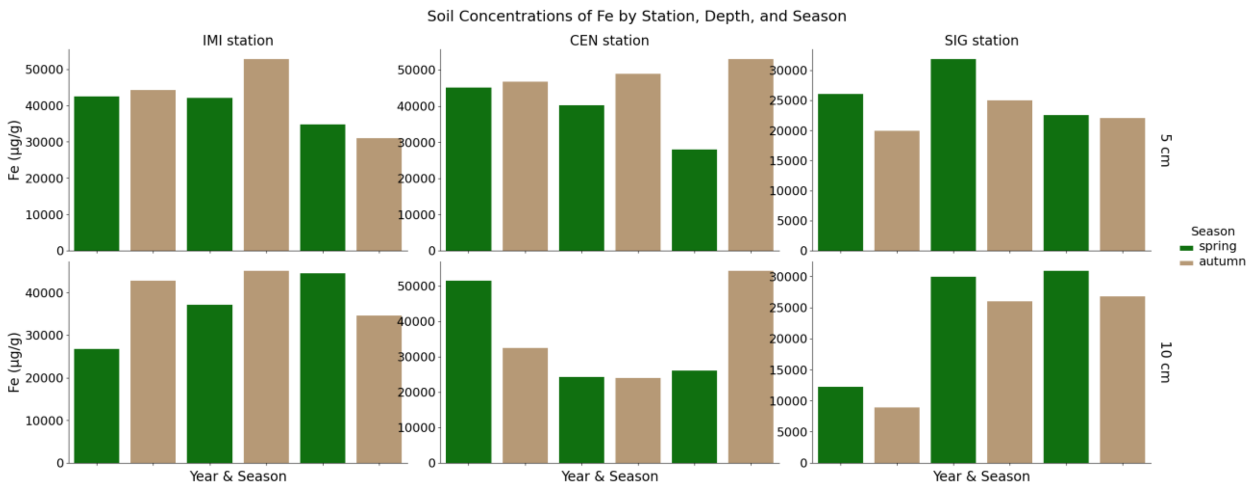


Figure S3. Fe concentration in soil by season, depth and station

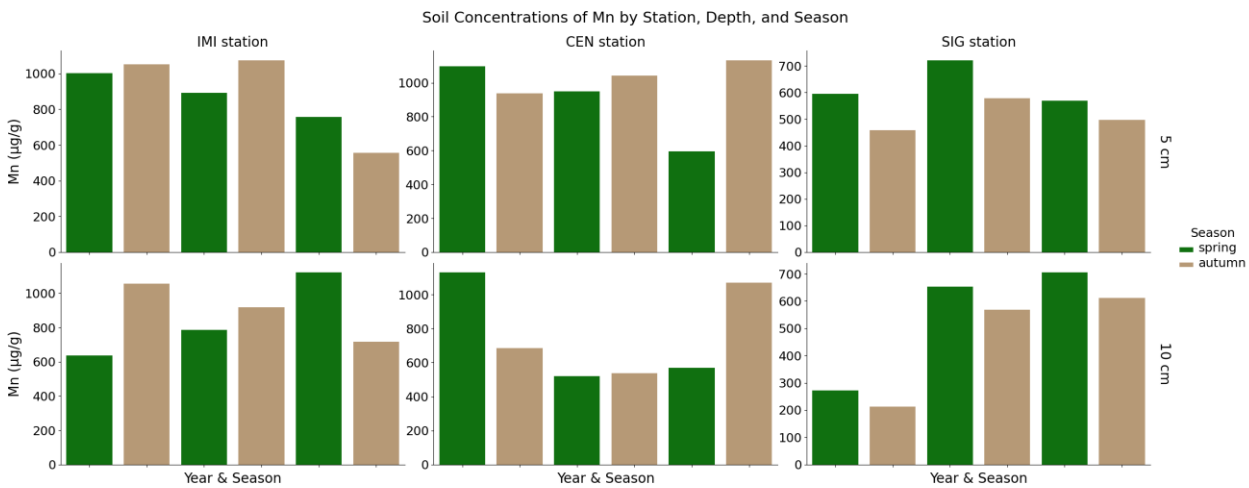


Figure S4. Mn concentration in soil by season, depth and station

Table S1 Correlation analysis for air (PM_{10}) and soil (for different depth) pollutants

pollutant	spearman_r, 5 cm	p_value, 5 cm	spearman_r, 10 cm	p_value, 10 cm
BaA	-0.200	0.704	-0.200	0.704
BaP	-0.486	0.329	-0.314	0.544
BbF	-0.314	0.544	-0.314	0.544
BghiP	-0.600	0.208	-0.314	0.544
BjF	-0.600	0.208	-0.429	0.397
BkF	-0.600	0.208	-0.429	0.397
Flu	-0.714	0.111	-0.657	0.156
Pyr	-0.543	0.266	-0.429	0.397
DahA	-0.714	0.111	-0.314	0.544
IP	-0.600	0.208	-0.429	0.397
Chry	-0.657	0.156	-0.657	0.156
Mn	0.771	0.072	0.943	0.005
Pb	0.943	0.005	0.829	0.042
Zn	-0.257	0.623	-0.257	0.623
Cu	0.714	0.111	0.600	0.208
As	-0.771	0.072	-0.714	0.111
Fe	0.771	0.072	0.886	0.019
Cd	0.883	0.020	0.000	1.000

Table S2. Correlation analysis for air (different time windows) and soil pollutants

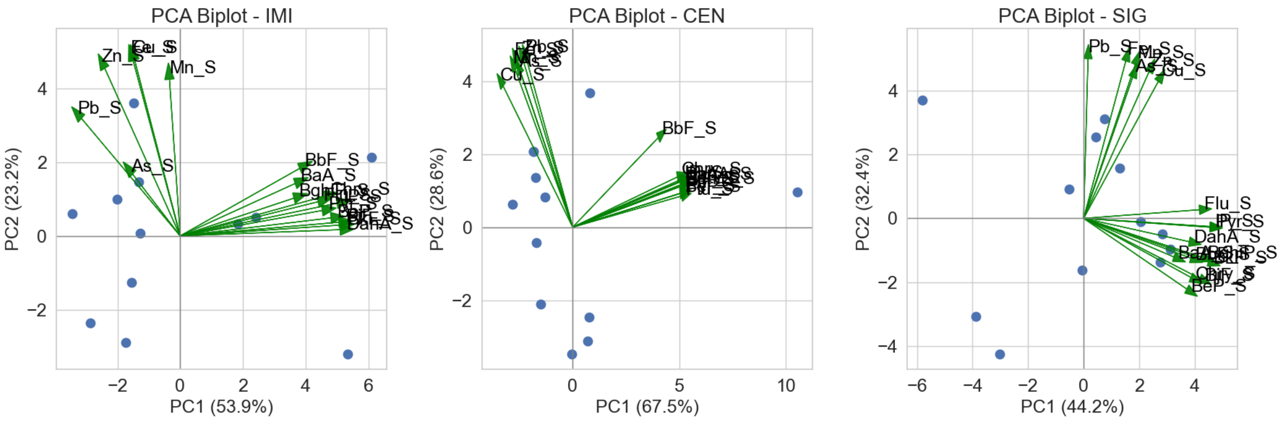
pollutant	months	spearman_r, 5 cm	p_value, 5 cm	spearman_r, 10 cm	p_value, 10 cm
As	1	-0.235	0.463	-0.210	0.513
BaA	1	0.063	0.846	0.196	0.542
BaP	1	0.021	0.948	-0.056	0.863
BbF	1	0.056	0.863	-0.021	0.948
BghiP	1	-0.140	0.665	-0.007	0.983
BjF	1	-0.186	0.564	-0.189	0.557
BkF	1	0.028	0.931	-0.077	0.812
Cd	1	0.167	0.667	0.073	0.877
Chry	1	0.035	0.914	0.056	0.863
Cu	1	0.622	0.031	0.308	0.331
DahA	1	-0.175	0.587	-0.371	0.236
Fe	1	-0.028	0.931	0.308	0.331
Flu	1	0.091	0.779	0.399	0.199
IP	1	-0.063	0.846	-0.112	0.729
Mn	1	-0.182	0.572	0.322	0.308
Pb	1	0.825	0.001	0.727	0.007
Pyr	1	0.175	0.587	0.329	0.297
Zn	1	-0.168	0.601	0.217	0.499
As	3	-0.298	0.347	-0.224	0.484
BaA	3	0.538	0.071	0.559	0.059

BaP	3	0.559	0.059	0.734	0.007
BbF	3	0.371	0.236	0.601	0.039
BghiP	3	0.490	0.106	0.503	0.095
BjF	3	0.165	0.609	0.280	0.379
BkF	3	0.357	0.255	0.566	0.055
Cd	3	0.159	0.683	0.109	0.816
Chry	3	0.650	0.022	0.706	0.010
Cu	3	0.531	0.075	0.273	0.391
DahA	3	0.147	0.649	0.280	0.379
Fe	3	-0.371	0.236	0.189	0.557
Flu	3	0.573	0.051	0.629	0.028
IP	3	0.503	0.095	0.706	0.010
Mn	3	-0.392	0.208	0.217	0.499
Pb	3	0.713	0.009	0.573	0.051
Pyr	3	0.490	0.106	0.566	0.055
Zn	3	-0.413	0.182	0.070	0.829
As	6	-0.336	0.285	-0.217	0.499
BaA	6	0.175	0.587	0.427	0.167
BaP	6	0.434	0.159	0.671	0.017
BbF	6	0.357	0.255	0.699	0.011
BghiP	6	0.385	0.217	0.601	0.039
BjF	6	0.312	0.324	0.510	0.090
BkF	6	0.392	0.208	0.615	0.033
Cd	6	-0.084	0.831	0.182	0.696
Chry	6	0.392	0.208	0.559	0.059
Cu	6	0.455	0.138	0.273	0.391
DahA	6	0.273	0.391	0.650	0.022
Fe	6	-0.455	0.138	0.168	0.602
Flu	6	0.524	0.080	0.720	0.008
IP	6	0.308	0.331	0.538	0.071
Mn	6	-0.392	0.208	0.259	0.417
Pb	6	0.713	0.009	0.615	0.033
Pyr	6	0.406	0.191	0.601	0.039
Zn	6	-0.371	0.235	0.140	0.665

Table S3. Correlation analysis for air and soil pollutants for different seasons and depths (for time windows of 6 months for air (PM_{10}) pollutant concentrations)

season	pollutant	spearman_r, 5 cm	p_value, 5 cm	spearman_r, 10 cm	p_value, 10 cm
autumn	As	-0.783	0.013	-0.444	0.232
autumn	BaA	-0.500	0.170	-0.467	0.205
autumn	BaP	-0.550	0.125	-0.433	0.244
autumn	BbF	-0.500	0.170	-0.467	0.205
autumn	BghiP	-0.500	0.170	-0.483	0.187
autumn	BjF	-0.650	0.058	-0.594	0.092
autumn	BkF	-0.636	0.066	-0.533	0.139

autumn	Cd	0.872	0.054	-1.000	
autumn	Chry	-0.500	0.170	-0.467	0.205
autumn	Cu	0.750	0.020	0.427	0.252
autumn	DahA	-0.427	0.252	-0.417	0.265
autumn	Fe	0.517	0.154	0.633	0.067
autumn	Flu	-0.483	0.187	-0.267	0.488
autumn	IP	-0.700	0.036	-0.650	0.058
autumn	Mn	0.517	0.154	0.650	0.058
autumn	Pb	0.900	0.001	0.767	0.016
autumn	Pyr	-0.517	0.154	-0.400	0.286
autumn	Zn	-0.017	0.966	-0.100	0.798
spring	As	-0.368	0.330	-0.450	0.224
spring	BaA	-0.483	0.187	-0.333	0.381
spring	BaP	-0.667	0.050	-0.550	0.125
spring	BbF	-0.650	0.058	-0.483	0.187
spring	BghiP	-0.733	0.025	-0.517	0.154
spring	BjF	-0.700	0.036	-0.550	0.125
spring	BkF	-0.667	0.050	-0.550	0.125
spring	Cd	0.143	0.760	0.595	0.159
spring	Chry	-0.633	0.067	-0.517	0.154
spring	Cu	0.300	0.433	0.083	0.831
spring	DahA	-0.552	0.123	-0.418	0.262
spring	Fe	0.550	0.125	0.333	0.381
spring	Flu	-0.217	0.576	-0.200	0.606
spring	IP	-0.561	0.116	-0.467	0.205
spring	Mn	0.533	0.139	0.250	0.516
spring	Pb	0.650	0.058	0.467	0.205
spring	Pyr	-0.483	0.187	-0.467	0.205
spring	Zn	-0.567	0.112	-0.200	0.606



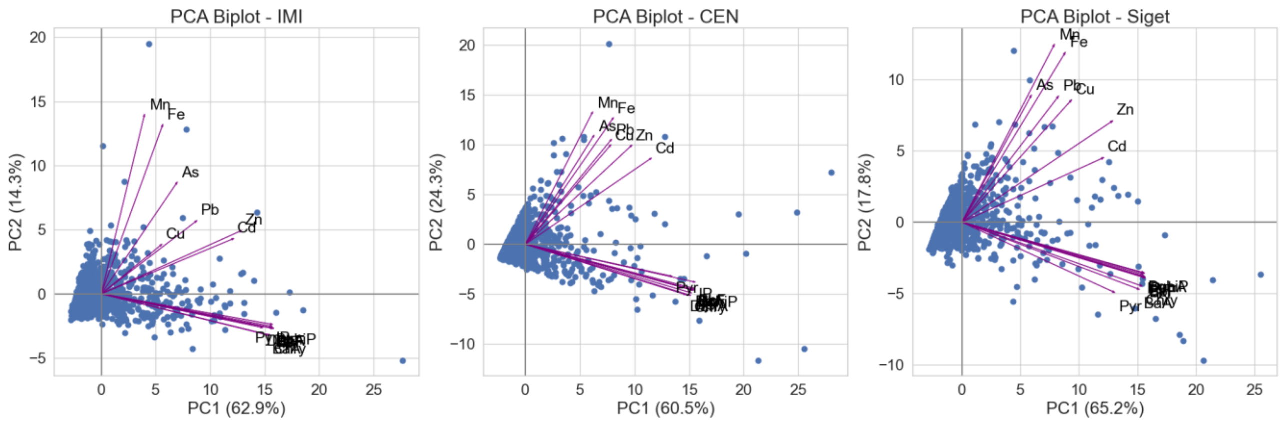


Figure S5. PCA for the soil pollutants (a) and air (PM_{10}) pollutants (b) for different monitoring stations

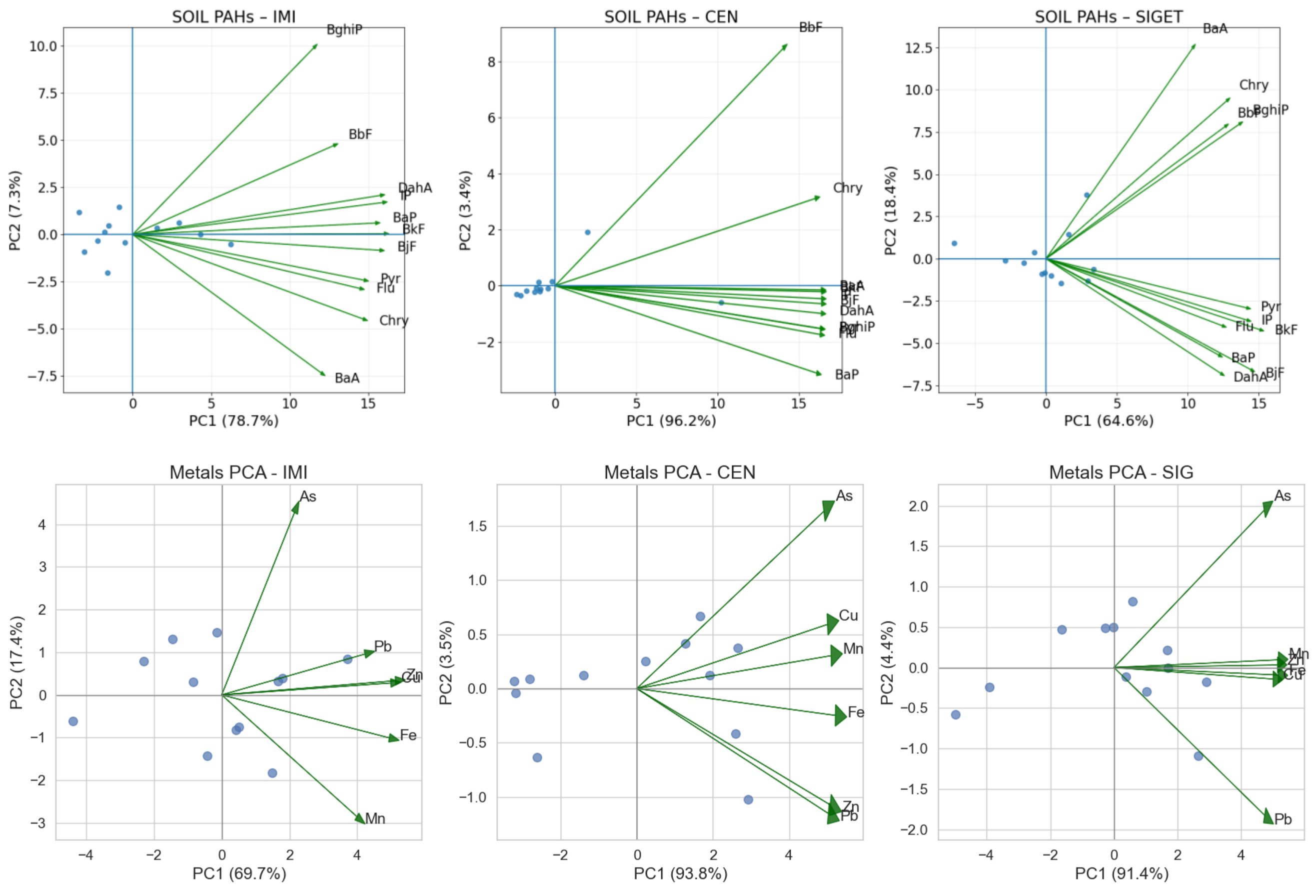


Figure S6. PCA for PAHs and metals soil samples for different monitoring stations

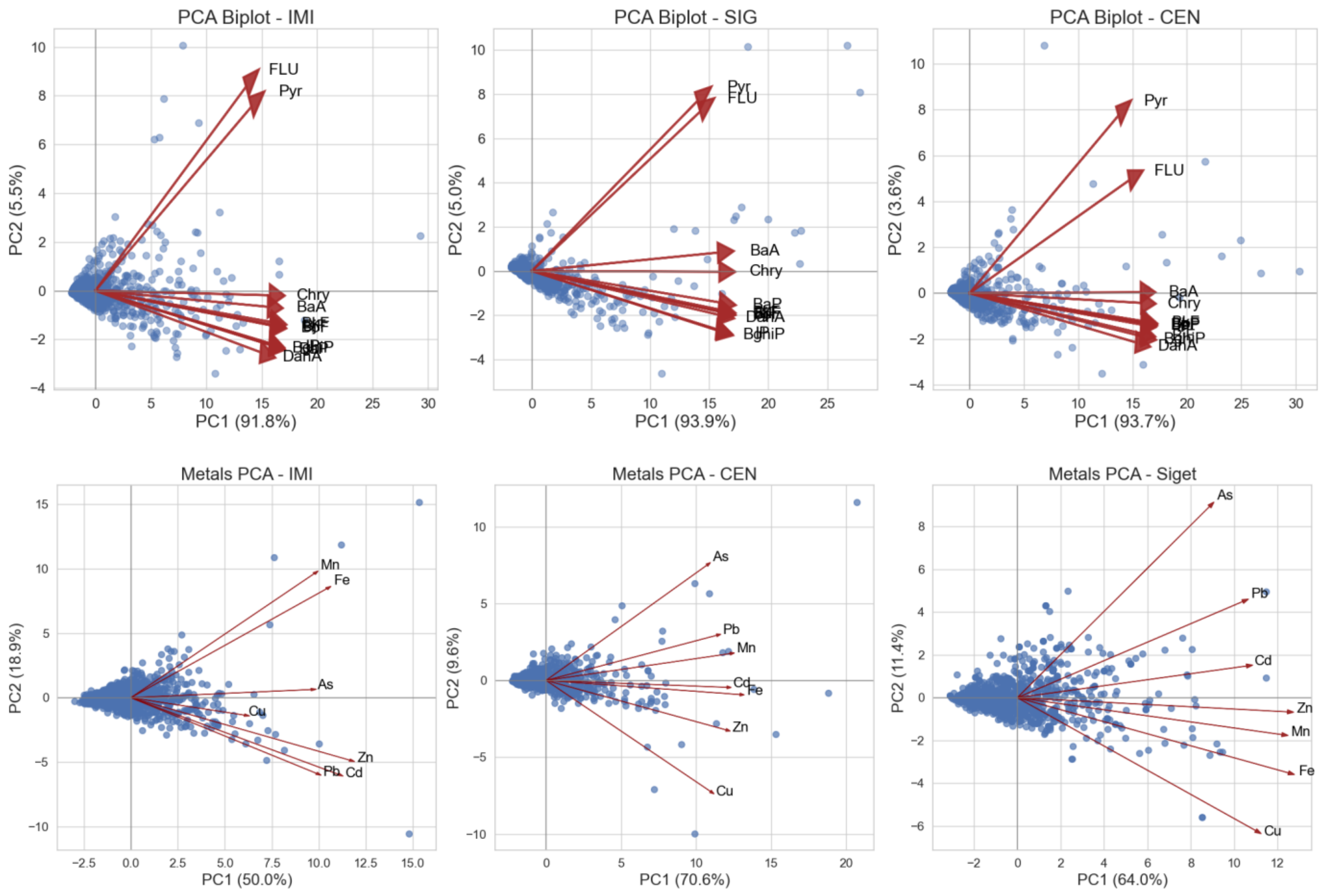


Figure S7. PCA for PAHs and metals air samples for different monitoring stations

6 BIOGRAPHY OF THE AUTHOR

Nikolina Račić was born on the 6th of December 1993 in Banja Luka, Bosnia and Herzegovina. She finished V. gimnazija high school in Zagreb. In 2012, she started studying Geology at the Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb. She completed her undergraduate studies in 2015, earning a bachelor's degree in geological engineering *summa cum laude*. That same year, she began her graduate studies in Geology of Mineral Resources and Geophysical Explorations, which she completed in 2017 by defending her master's thesis "Geomathematical analysis of pyrolysis data from surface samples of Upper Jurassic sediments in Lika and Dalmatia area". Since 2021, she has been employed at the Institute for Medical Research and Occupational Health (IMROH) in Zagreb, within the Division of Environmental Hygiene. Since April 2025 she has also been working as a researcher for the Lisbon Council, contributing to the writing and implementation of EU-funded project proposals. She is coauthor and has been working on several research projects, including Horizon NextAIRE, Horizon EDIAQI and EnvironPollutHealth. She is a recipient of the Croatian national L'Oréal-UNESCO For Women in Science fellowship.

She is an author of 17 scientific papers.

ORCID: https://orcid.org/0009-0007-1468-5564

6.1 Scientific publications and activities

Original scientific publications:

- **Račić, Nikolina**; Petrić, Valentino; Pehnec, Gordana; Jakovljević, Ivana; Lovrić Štefiček, Marija Jelena; Gajski, Goran; Mureddu, Francesco; Lovrić, Mario. Indoor Air Filtration System Performance: Evidence from a Two-Week Office Study Within the EDIAQI Project // Atmosphere, 17 (2026), 4; 393, 17. doi: 10.3390/atmos17040393

- Petrić, Valentino; Strbad, Dejan; **Račić, Nikolina**; Hussein, Tareq; Kecorius, Simonas; Mureddu, Francesco; Lovrić, Mario. Hybrid digital twin framework for real-time indoor air Quality monitoring and filtration optimization // Atmosphere, 17 (2026), 2; 184, 22. doi: 10.3390/atmos17020184

- Petrić, Valentino; Škvarč, Goran; Markulin, Tihomir; **Račić, Nikolina**; Matanović, Hana; Mureddu, Francesco; Burridge, Henry; Pehnec, Gordana; Lovrić, Mario. Seasonal and regional variations in CO2 concentrations: A large-scale sensor-based study from Croatian schools using machine learning // Atmosphere, 17 (2026), 1; 106, 18. doi: 10.3390/atmos17010106

- **Račić, Nikolina**; Ružičić, Stanko; Petrić, Valentino; Terzić, Teo; Antunović, Mario; Škaro, Ivan; Pehnec, Gordana; Bešlić, Ivan; Jakovljević, Ivana; Sever Štrukil, Zdravka et al. Assessment of contributors to airborne PAHs and heavy metals in PM10 using temporal, spatial, traffic and heating data in explainable machine learning models // Atmospheric environment. X, 29 (2026), 100413, 12. doi: 10.1016/j.aeaoa.2026.100413

- **Račić, Nikolina**; Petrić, Valentino; Mureddu, Francesco; Portin, Harri; Niemi, Jarkko V.; Hussein, Tareq; Lovrić, Mario. A Proxy Model for Traffic Related Air Pollution Indicators Based on Traffic Count // Atmosphere, 16 (2025), 5; 538, 14. doi: 10.3390/atmos16050538

- **Račić, Nikolina**; Pehnec, Gordana; Jakovljević, Ivana; Sever Štrukil, Zdravka; Mureddu, Francesco; Forsmann, Michael; Lovrić, Mario Machine learning analysis of drivers of differences in PAH content between PM1 and PM10 in Zagreb, Croatia // Atmospheric Pollution Research, 16 (2025), 7; 102541, 11. doi: 10.1016/j.apr.2025.102541

- Jakovljević, Ivana; Horvat, Tajana; Pehnec, Gordana; **Račić, Nikolina**; Dvoršćak, Marija; Jagić, Karla; Klinčić, Darija. Relationship between polycyclic aromatic hydrocarbons measured in dust collected inside cars and from ambient air // Environmental toxicology and chemistry, 44 (2025), 11; vgaf183, 17. doi: https://doi.org/10.1093/etojnl/vgaf183

- Petrić, Valentino; **Račić, Nikolina**; Hrga, Ivana; Grgec, Danijel; Marić, Marko; Krivohlavek, Adela; Anić, Zvonimir; Lovrić, Mario; Jergović, Matijana Assessment of Sensor Data from an Air Quality Monitoring Network—The Need for Machine Learning-Based Recalibration and Its Relevance in Health Impact Analysis of Local Pollution Events // Atmosphere, 16 (2025), 12; 1358, 23. doi: 10.3390/atmos16121358

- **Račić, Nikolina**; Ružičić, Stanko; Pehnec, Gordana; Jakovljević, Ivana; Sever Štrukil, Zdravka; Rinkovec, Jasmina; Žužul, Silva; Smoljo, Iva; Zgorelec, Željka; Lovrić, Mario Linking Atmospheric and Soil Contamination: A Comparative Study of PAHs and Metals in PM10 and Surface Soil near Urban Monitoring Stations // Toxics, 13 (2025), 10; 866, 20. doi: 10.3390/toxics13100866

- Lovrić, Mario; Petrić, Valentino; Strbad, Dejan; Terzić, Teo; Frka, Sanja; Kušan, Ana Cvitešić; Fermoso, Jose; Düsing, Sebastian; Alas, Honey Dawn; Ladavac, Mila Dobrić et al. Indoor and ambient air pollution dataset using a multi-instrument approach and total event monitoring // Scientific data, 12 (2025), 1; 1584, 8. doi: 10.1038/s41597-025-05847-3

- **Račić, Nikolina**; Frka, Sanja; Cvitešić Kušan, Ana; Petrić, Valentino; Mureddu, Francesco; Lovrić, Mario. Characterizing Indoor Black Carbon Dynamics in a Residential Environment: The Role of Human Activity and Ventilation Behavior // Toxics, 13 (2025), 7; 536-549. doi: https://doi.org/10.3390/toxics13070536

- **Račić, Nikolina**; Terzić, Ivana; Karlović, Nina; Bošnjaković, Anja; Terzić, Teo; Jakovljević, Ivana; Pehnec, Gordana; Horvat, Tajana; Gajski, Goran; Gerić, Marko et

al. Volatile Organic Compounds (VOCs) and Polycyclic Aromatic Hydrocarbons (PAHs) in Indoor Environments: A Review and Analysis of Measured Concentrations in Europe // Indoor air-international journal of indoor air quality and climate, 2025 (2025), 1; 2, 19. doi: https://doi.org/10.1155/ina/5945455

- Jakovljević, Ivana; Sever Štrukil, Zdravka; Pehnec, Gordana; Horvat, Tajana; Sanković, Mandica; Šumanovac, Antun; Davila, Silvije; **Račić, Nikolina**; Gajski, Goran Ambient air pollution and carcinogenic activity at three different urban locations // Ecotoxicology and environmental safety, 289 (2025), 1-16. doi: 10.1016/j.ecoenv.2025.117704

- Lovrić, Mario; Gajski, Goran; Fernández-Agüera, Jessica; Pöhlker, Mira; Gursch, Heimo; The EDIAQI Consortium; Borg, Alex; Switters, Jon; Mureddu, Francesco. Evidence-driven indoor air quality improvement: An innovative and interdisciplinary approach to improving indoor air quality // Biofactors, / (2024), /; /-/. doi: 10.1002/biof.2126

- Lovrić, Mario; **Račić, Nikolina**; Pehnec, Gordana; Horvat, Tajana; Lovrić Štefiček, Marija Jelena; Jakovljević, Ivana. Indoor Polycyclic Aromatic Hydrocarbons—Relationship to Ambient Air, Risk Estimation, and Source Apportionment Based on Household Measurements // Atmosphere, 15 (2024), 1525; 1-17. doi: 10.3390/atmos15121525

- **Račić, Nikolina**; Ružičić, Stanko; Terzić, Teo; Pehnec, Gordana; Jakovljević, Ivana; Sever Štrukil, Zdravka; Žužul, Silva; Rinkovec, Jasmina; Lovrić, Mario Analyzing the relationship between gas consumption and airborne pollutants: case study of Zagreb, Croatia // Air Quality Atmosphere and Health, 18 (2024), -; 507-519. doi: 10.1007/s11869-024-01655-7

- Kecorius, Simonas; Madueño, Leizel; Lovric, Mario; **Racic, Nikolina**; Schwarz, Maximilian; Cyrys, Josef; Casquero-Vera, Juan Andrés; Alados-Arboledas, Lucas; Conil, Sébastien; Sciare, Jean et al. Atmospheric new particle formation identifier using longitudinal global particle number size distribution data // Scientific data, 11 (2024), 1; 1239, 10. doi: 10.1038/s41597-024-04079-1

- **Račić, Nikolina**; Malvić, Tomislav. Relation between air and soil pollution based on statistical analysis and interpolation of Nickel (Ni) and Lead (Pb): Case study of Zagreb, Croatia // Mining of mineral deposits, 17 (2023), 2; 112-120. doi: 10.33271/mining17.02.112

6.2 Conference abstracts

- Petrić, Valentino; Strbad, Dejan; **Račić, Nikolina**; Mureddu, Francesco; Lovrić, Mario Hybrid digital twin framework for real-time indoor air quality monitoring and filtration optimization // International conference and 14th Croatian scientific and professional meeting AIR PROTECTION 2025 - Book of Abstracts / Pehnec, Gordana (ur.). Zagreb: Hrvatsko udruženje za zaštitu zraka (HUZZ) / Croatian Air Pollution Prevention Association (CAPPA), 2025. str. 77-77

- **Račić, Nikolina**; Lovrić, Mario; Pehnec, Gordana; Jakovljević, Ivana; Sever Štrukil, Zdravka; Žužul, Silva; Rinkovec, Jasmina
Using explainable machine learning to understand what affects pahs and heavy metals in PM10 particles // International conference and 14th Croatian scientific and professional meeting AIR PROTECTION 2025 - Book of Abstracts / Pehnec, Gordana (ur.).
Zagreb: Hrvatsko udruženje za zaštitu zraka (HUZZ) / Croatian Air Pollution Prevention Association (CAPPA), 2025. str. 26-26

- **Račić, Nikolina**; Lovrić, Mario; Pehnec, Gordana; Jakovljević, Ivana; Sever Štrukil, Zdravka; Smoljo, Iva; Žužul, Silva; Rinkovec, Jasmina
From airborne particles to soil deposits: understanding the distribution of PAHs and heavy metals // 29th Croatian Meeting of Chemists and Chemical Engineers and 7th Symposium Vladimir Prelog: Book of abstracts / Vrsalović Presečki, Ana (ur.).
Split: Croatian Chemical Society, 2025. str. 265-265

- **Račić, Nikolina**; Lovrić, Mario; Terzić, Teo; Jakovljević, Ivana; Rinkovec, Jasmina; Žužul, Silva; Sever Štrukil, Zdravka; Pehnec, Gordana
Analyzing air quality: a source apportionment study in Zagreb, Croatia // XV Meeting Of Young Chemical Engineers: Book Of Abstracts / XV. susret mladih kemijskih inženjera: knjiga sažetaka / Ujević Andrijić, Željka; Vidak, Andrej (ur.).
Zagreb: Hrvatsko društvo kemijskih inženjera i tehnologa (HDKI), 2024. str. 66-66

- **Račić, Nikolina**. Strojno učenje u istraživanju okoliša: analiza potrošnje plina i onečišćenja zraka u Zagrebu // 5. Znanstvena škola o okolišu "Od klimatskih promjena do primjene AI u istraživanju okoliša" - knjiga sažetaka / Fiket, Željka; Furdek Turk, Martina; Ivanić, Maja (ur.).
Zagreb: Institut Ruđer Bošković, 2024. str. 19-19

- **Račić, Nikolina**; Pehnec, Gordana; Jakovljević, Ivana; Sever Štrukil, Zdravka; Lovrić, Mario. Polycyclic aromatic hydrocarbons (PAHs) in PM10 at an urban measuring station in Zagreb: a random forest regression analysys // 20th International Conference Ružička Days "Today Science – Tomorrow Industry": Book of Abstracts / Jukić, Ante; Ocelić Bulatović, Vesna; Kučić Grgić, Dajana (ur.).
Zagreb, Osijek: Hrvatsko društvo kemijskih inženjera i tehnologa; Sveučilište Josipa Jurja Strosssmayera u Osijeku, Prehrambeno-tehnološki fakultet Osijek, 2024. str. 123-123

- **Račić, Nikolina**; Lovrić, Mario; Forsmann, Michael; Jakovljević, Ivana; Sever Štrukil, Zdravka; Pehnec, Gordana. Understanding the drivers of differences in PAH compositions of PM1 and PM10 – a machine learning study in Zagreb, Croatia // 9th International Symposium Ultrafine Particles - Air Quality and Climate: Book of abstract and program / Saathoff, Harald; Leisner, Thomas; Ziegahn, Karl-Friedrich et al. (ur.).
Karlsruhe: Karlsruhe Institute of Technology (KIT), 2024. str. 38-38

- **Račić, Nikolina**; Terzić, Teo; Pehnec, Gordana; Jakovljević, Ivana; Sever Štrukil, Zdravka; Žužul, Silva; Rinkovec, Jasmina; Lovrić, Mario
The relationship between gas consumption and air pollution in Zagreb // International conference and 13th Croatian scientific and professional meeting air protection 2023 - Proceedings / Pehnec, Gordana (ur.). Zagreb: Croatian Air Pollution Prevention

Association, 2024. str. 63-69

- **Račić, Nikolina**; Terzić, Teo; Bošnjaković, Anja; Lovrić, Mario; Pehnec, Gordana; Jakovljević, Ivana; Sever Štrukil, Zdravka; Žužul, Silva; Rinkovec, Jasmina Analyzing the relationship between gas consumption and airborne pollutants: A case study of Zagreb, Croatia // International conference and 13th Croatian scientific and professional meeting "Air Protection 2023": Book of Abstracts / Pehnec, Gordana (ur.). Zagreb: Hrvatsko udruženje za zaštitu zraka (HUZZ), 2023. str. 29-29