international conference
Ružička days
TODAY SCIENCE – TOMORROW INDUSTRY
September 18–20, 2024 | Vukovar, Croatia

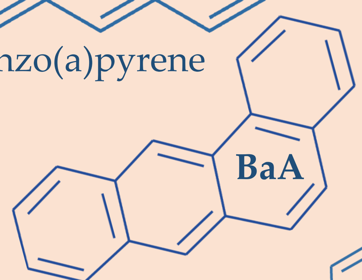
POLYCYCLIC AROMATIC HYDROCARBONS (PAHs) IN PM_{10} AT AN URBAN MEASURING STATION IN ZAGREB: A RANDOM FOREST REGRESSION ANALYSIS

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

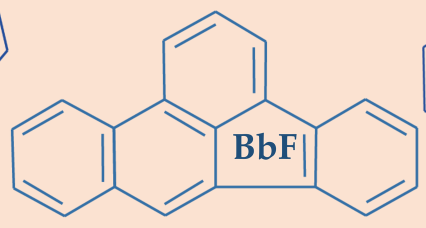
[1]Institute for Medical Research and Occupational Health, Environmental Hygiene Unit, Ksaverska cesta 2, 10000 Zagreb, Croatia
[2]The Lisbon Council, Rue de la Loi 155, 1040 Brussels, Belgium
[3]Centre for Bioanthropology, Institute for Anthropological Research, Gajeva 32, HR-10000 Zagreb, Croatia

PAHs:

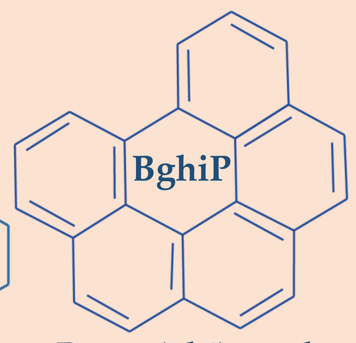
Benzo(a)pyrene



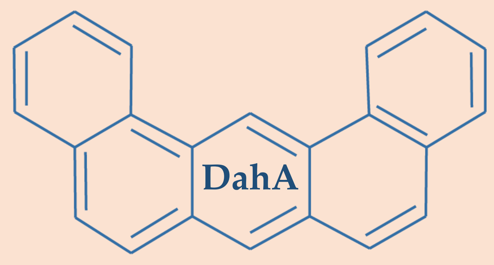
Benzo(a)anthracene



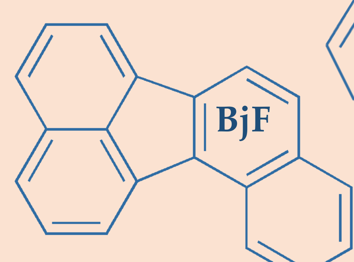
Benzo(b)fluoranthene



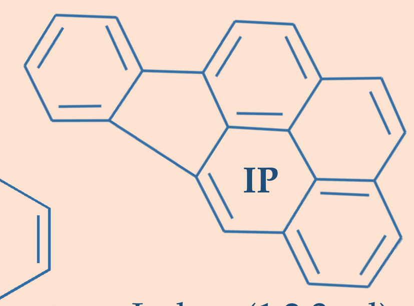
Benzo(ghi)perylene



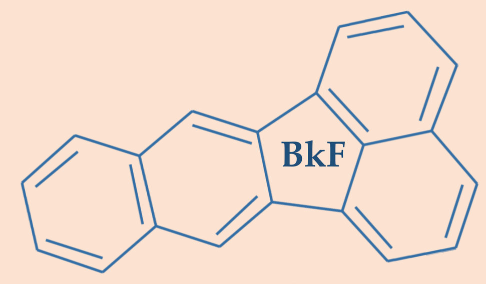
Dibenzo(ah)anthracene



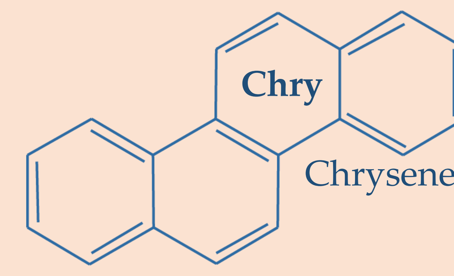
Benzo(j)fluoranthene



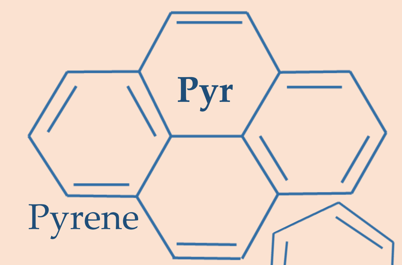
Indeno(1,2,3-cd)pyrene



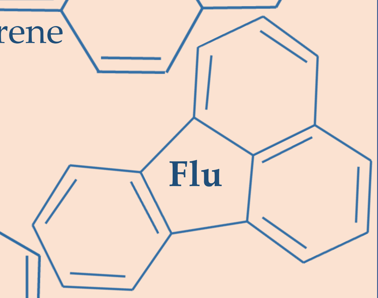
Benzo(k)fluoranthene



Chrysene



Pyrene



Fluoranthene

INTRODUCTION

- Measurements of **polycyclic aromatic hydrocarbons** (PAHs) in PM_{10} during **4-year period (2017–2020)** at urban **air quality** measuring station **Siget** (excluding COVID lockdown period), funded by the City of Zagreb
- Continuously measurement (**24-h samples**) of PAHs in PM_{10}
- Compounds of interest due to **associated health risks** – cancer, respiratory and cardiovascular diseases
- Impact of meteorological parameters, satellite data, traffic and gas consumption data on PAHs levels

MATERIALS AND METHODS

- **Data collection:**
- Air pollution data (PAHs in PM_{10})
- Meteorological parameters
- Satellite data
- Gas consumption data
- Traffic density dana
- **Measurement period: 2017-2020**
- **Data analysis:**
- Principal component analysis (PCA)
- Random forrest regression (RF)
- Permutation importance analysis
- Shapley additive explanations (SHAP)
- Seasonal AutoRegressive Integrated Moving Average (SARIMA)

RESULTS

- PCA results shows clustering of some PAHs (4-ring: BaA, Chry, Pyr, Flu) which indicate strong correlation among these variables. This suggests that these PAHs might share similar sources or are influenced by similar environmental factors (*Figure 1*).
- SHAP was used to determine the most influential features in the Random Forest model for each PAH – identified minimum temperature, average temperature, NO_2 and satellite data as significant factors affecting PAHs concentrations (*Figure 2a*).
- The SARIMA model was applied to forecast the concentrations of various PAHs over time and was able to capture the seasonal patterns and trends in PAHs concentrations for most pollutants. However, the model's accuracy varied depending on the specific PAH being analyzed (*Figure 2b*).

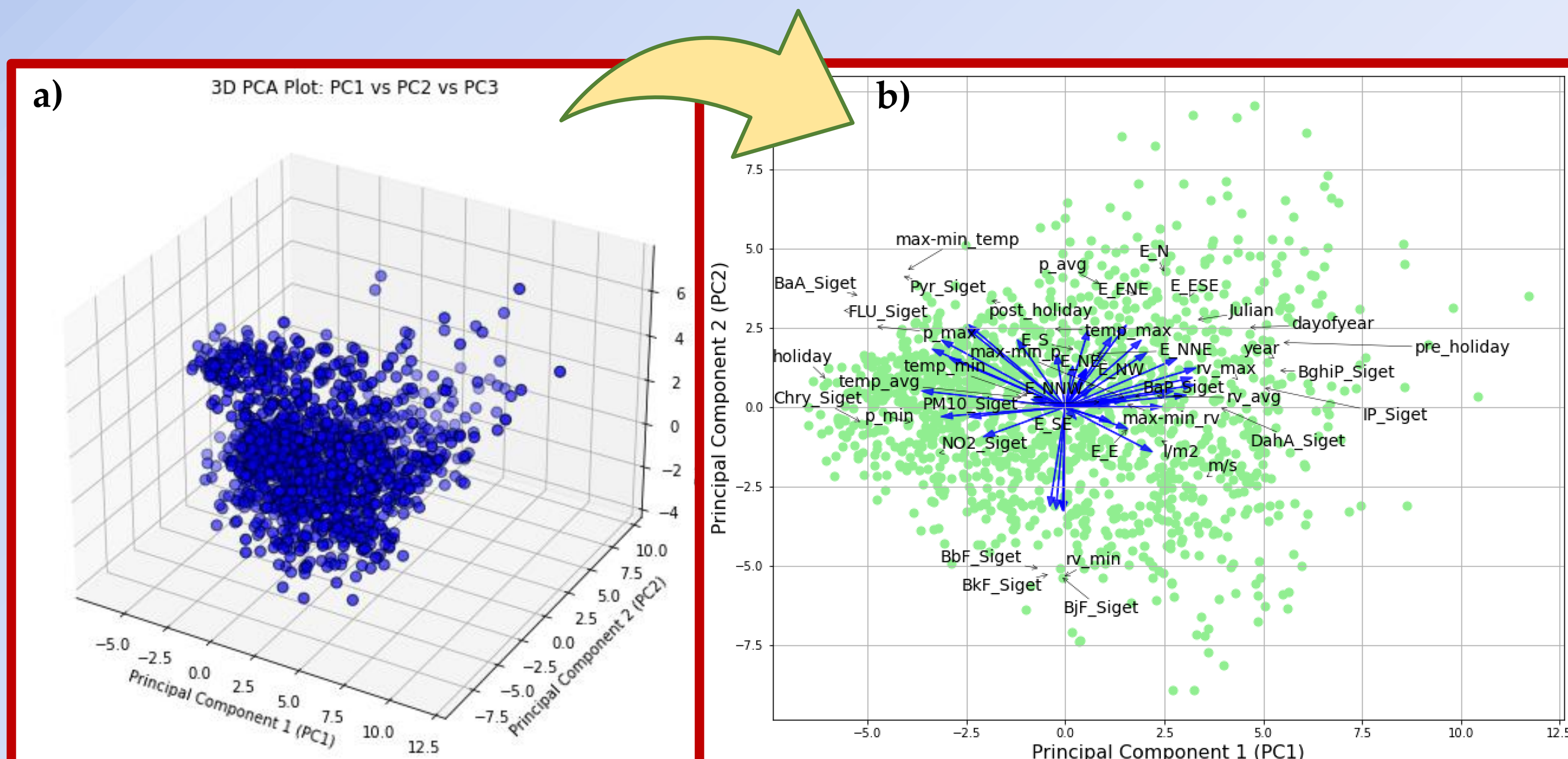


Figure 1. *PCA scatter plot (a) of measurements for three components and PCA bi-plot (b)*

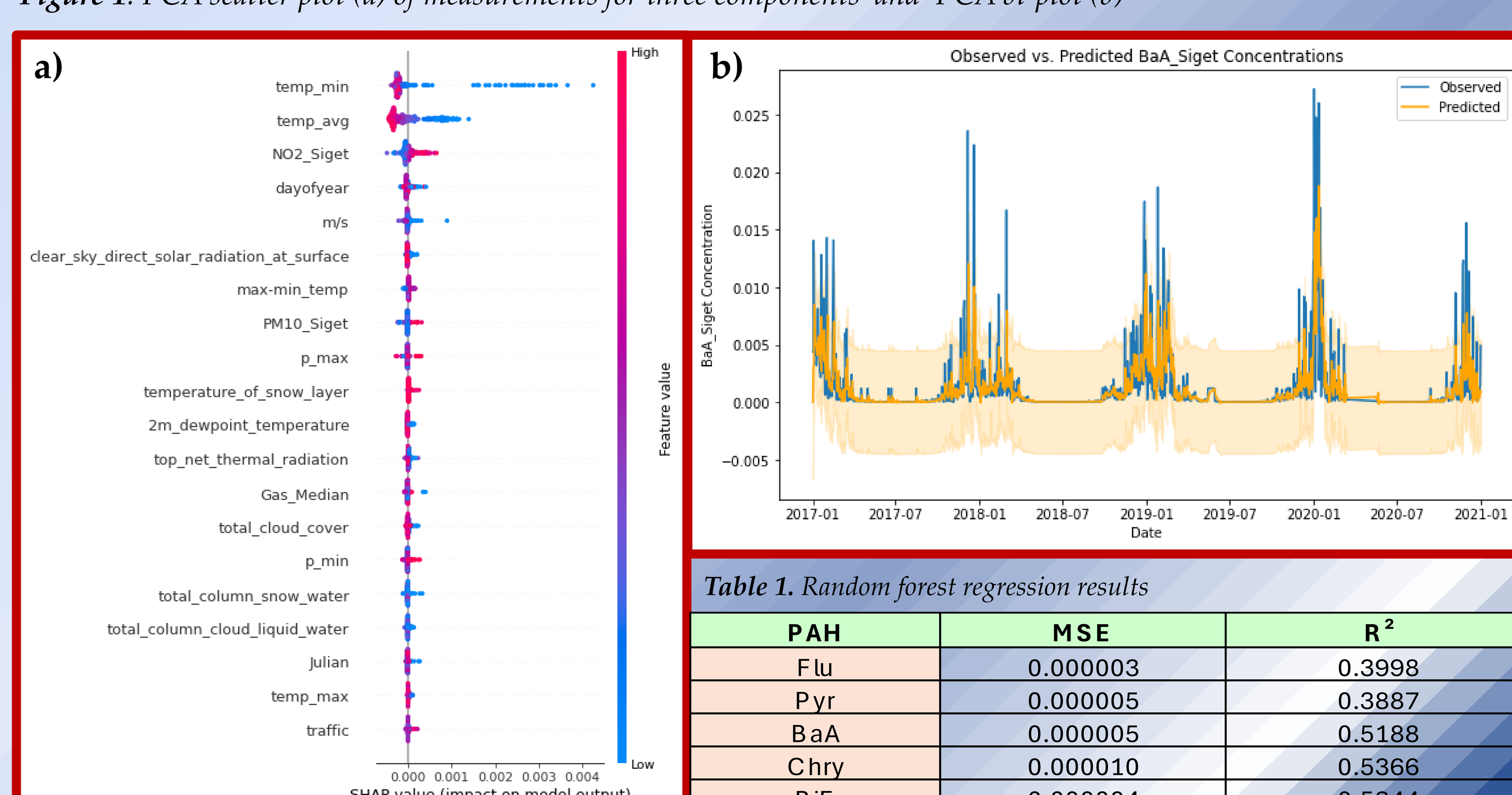


Figure 2. *SHAP results with the most influential features in the Random Forest model (a) and SARIMA time series forcasting for BaA (b)*

Table 1. *Random forest regression results*

PAH	MSE	R^2
Flu	0.000003	0.3998
Pyr	0.000005	0.3887
BaA	0.000005	0.5188
Chry	0.000010	0.5366
BjF	0.000004	0.5344
BbF	0.000009	0.5525
BkF	0.000001	0.5447
BaP	0.000010	0.5262
DahA	0.000000	0.5087
BghiP	0.000006	0.5265
IP	0.000005	0.5400

CONCLUSIONS

PCA effectively reduced the dataset's dimensionality, revealing that temperature, pressure, and wind speed are the primary environmental factors driving variations in PAHs concentrations. The clustering of certain PAHs in the biplot suggests that these pollutants are likely influenced by similar sources or environmental conditions. The RF model identified key predictors of PAH concentrations. Temperature-related variables and NO_2 emerged as the most influential features. This aligns with the PCA findings, where these variables also showed strong contributions to the principal components. SHAP results shows that minimum and average temperature have the most significant impact on the model's predictions across multiple PAHs, confirming the importance of meteorological factors in determining PAHs levels. The combination of PCA, RF, SHAP and SARIMA analyses offers a comprehensive understanding of the factors influencing PAHs concentrations. The findings suggest that both meteorological conditions and specific pollutants like NO_2 play a crucial role in forecasting PAHs levels.

EnvironPollutHealth
Funded by the European Union NextGenerationEU
ReC-IMI
European Union European Structural and Investment Funds

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 – Next Generation EU (Program Contract of 8 December 2023, Class: 643-02/23-01/00016, Reg. no. 533-03-23-0006)