Modeling the Spatial Determinants of Urban Air Pollutants: A Multi-Method Analysis of CO_2, PM2.5, NOx, and SO_2 in Functional Urban Areas

Zekang Chen[1], Jiahao Zhu[2]
[1]Institute of Computer Science, University of Tartu
[2]School of Economics and Business Administration, University of Tartu

Introduction

Sustainable development has become a dominant global agenda, yet air pollution remains a pressing challenge across European cities. This study investigates the drivers, spatial patterns, and inter-city propagation of four major air pollutants across Europe's Functional Urban Areas (FUAs).

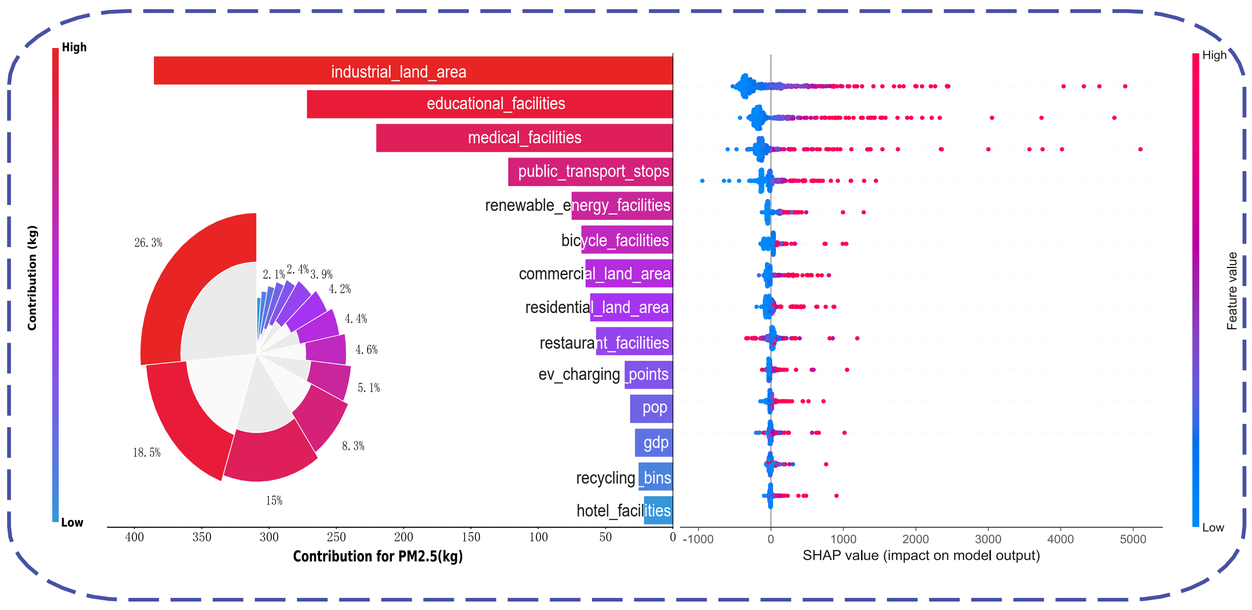
We integrate machine learning, Geographically Weighted Regression (GWR), and Graph Attention Networks (GAT) to identify key emission drivers. These methods also uncover spatial heterogeneity under varying economic and geographic contexts, as well as inter-city propagation structures.

Unit of Analysis: 619 European Functional Urban Areas
Target Variables: CO2, SO2, NOx, PM2.5
Explanatory Variables: A total of 14 spatial and socioeconomic variables are considered, including industrial land area, population, and the number of EV charging stations, public transport stops, and medical facilities, among others.

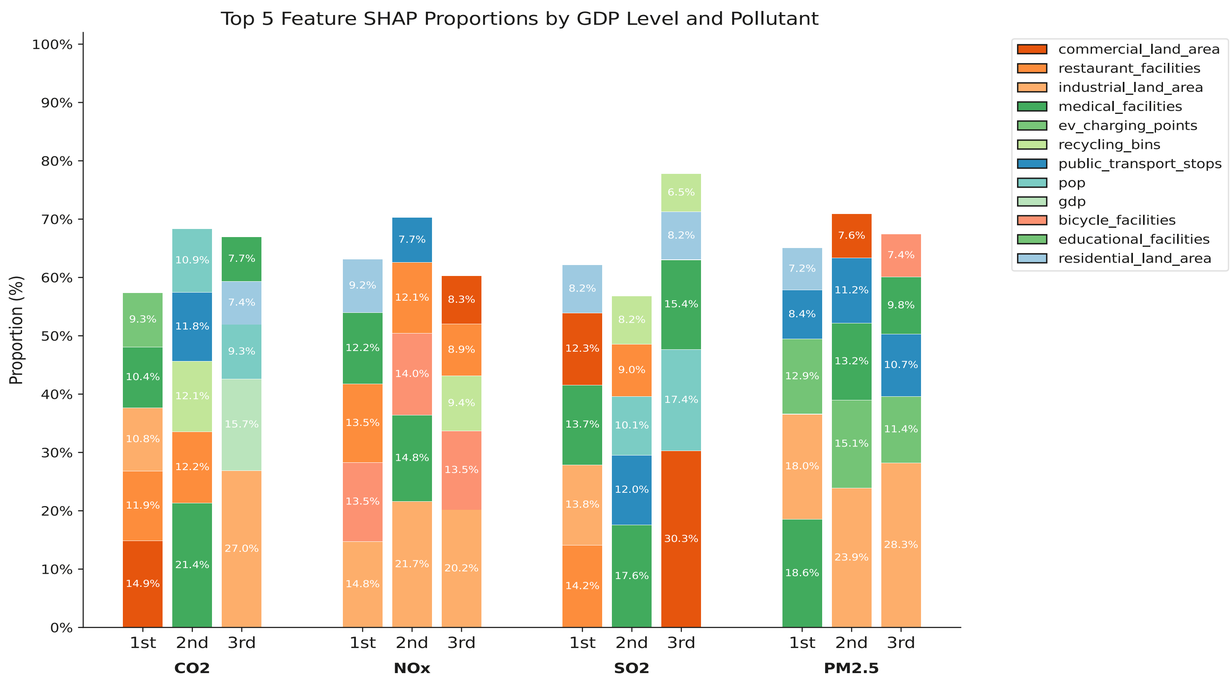
Module 1-Global Pollution Prediction and Feature Attribution

	Models					
	XGBoost	CatBoost	Random Forest	Elastic Net	KNN	Extra Trees
CO2						
R^2	0.3673	0.6543	0.5676	0.5482	0.5333	0.4994
Pearson's r	0.6277	0.8136	0.7540	0.7507	0.7434	0.7211
RMSE	3290019	2432064	2719996	2780269	2825612	2926468
NOx						
R^2	0.9259	0.9571	0.9470	0.9584	0.8979	0.9047
Pearson's r	0.9630	0.9786	0.9745	0.9814	0.9686	0.9641
RMSE	1235	940	1045	925	1450	1401
SO2						
R^2	0.2944	0.3518	0.3163	0.3337	0.1808	0.3349
Pearson's r	0.5618	0.5951	0.5812	0.5826	0.4674	0.5792
RMSE	2339	2242	2302	2273	2520	2271
PM2.5						
R^2	0.8917	0.9517	0.9156	0.9438	0.8642	0.8964
Pearson's r	0.9461	0.9809	0.9573	0.9719	0.9495	0.9664
RMSE	541	362	478	390	606	530

Table 1: Performance comparison of different models on predicting pollutant emissions using R^2, Pearson's r, and RMSE (rounded to nearest integer).



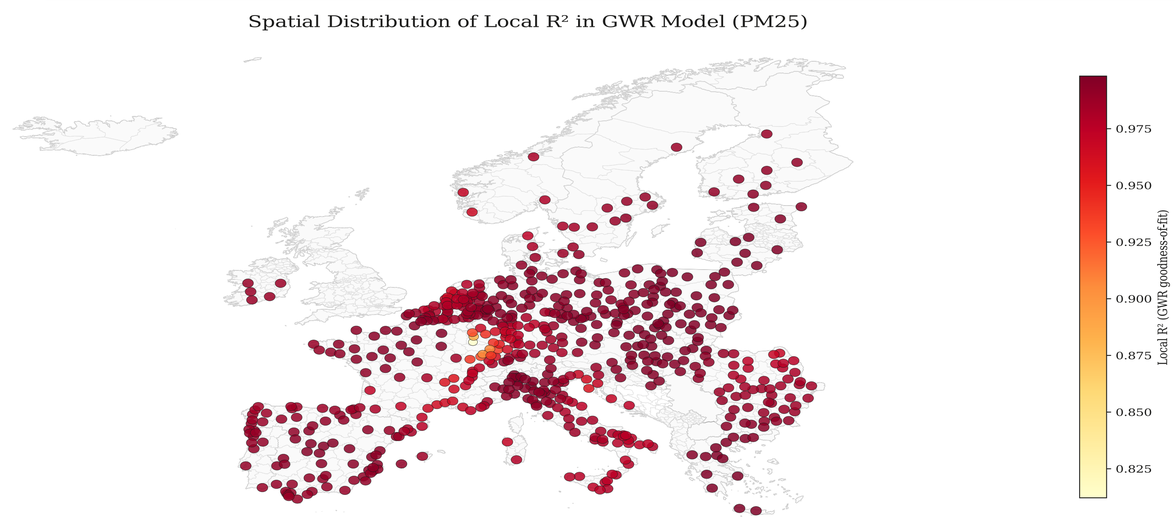
Module 2-Stratified SHAP Analysis by Economic Tier

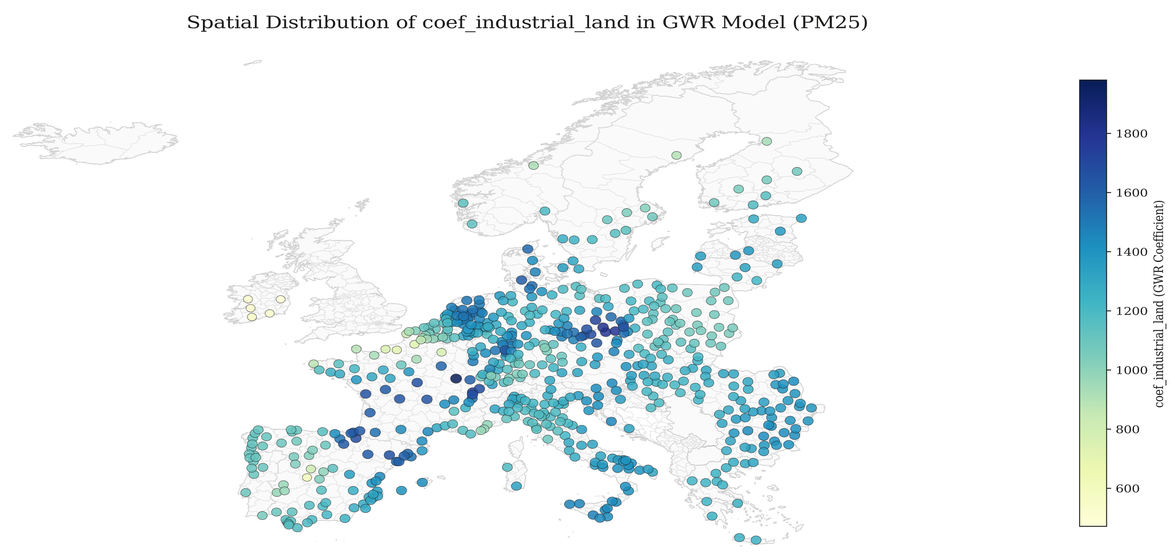


- The importance of industrial land is relatively stable, but its impact is stronger in lower economic tiers.
- Medical facilities are consistently important drivers of $PM_{2.5}$ and SO_2 emissions across all economic tiers.

Module 3-Spatial Heterogeneity Analysis of Pollutants

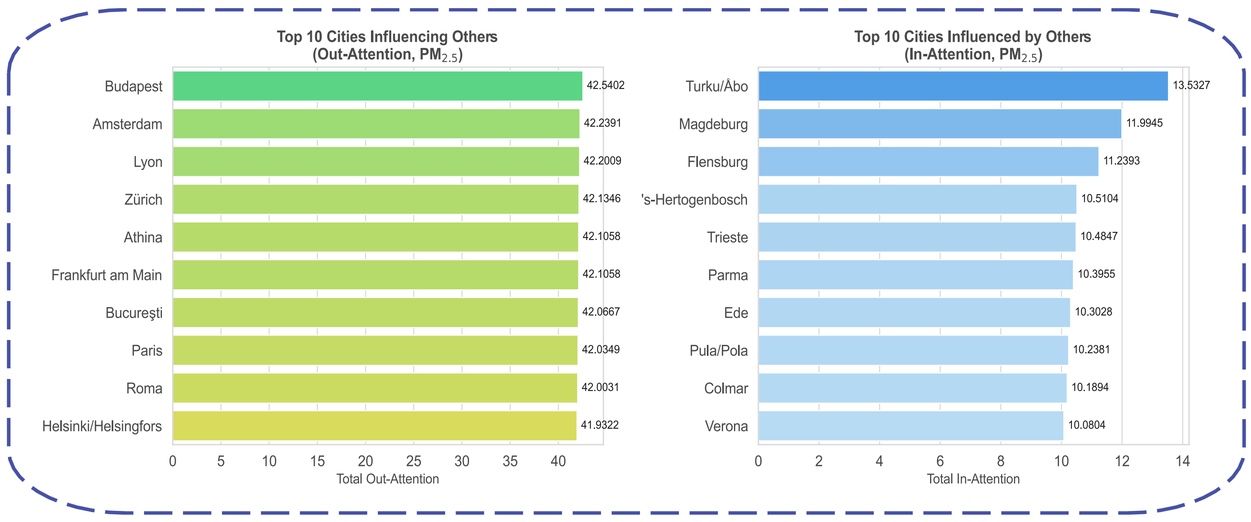
We applied a GWR model to assess how emission driver influence varies across space. By allowing each city to estimate its own set of local coefficients, GWR captures the geographic heterogeneity and spatial non-stationarity in pollutant generation mechanisms.
The analysis focuses on NOx and PM2.5, with five explanatory variables: industrial land area, and the number of public transport stops, restaurants, bicycle lanes, and medical facilities. Local R^2, residuals, and variable coefficients were visualized using graduated color point maps.





Module 4-Urban Emission Influence Analysis via Graph Attention

We construct a GAT based on geographic proximity to model spatial dependencies and pollutant transmission pathways among cities. The GAT framework utilizes attention weights to identify city pairs with significant emission influence, quantifying both the outgoing impact and susceptibility of each city within the network.
In addition, we compute a range of graph centrality measures, including PageRank, betweenness, and closeness centrality. Out-degree and In-degree are used to analyze and visualize the roles of cities in the pollution network.



Conclusion

This study integrates machine learning, SHAP, Geographically Weighted Regression, and Graph Neural Networks to analyze the spatial patterns and driving factors of air pollution in European cities.
Our results reveal that emission drivers vary across economic levels and geographic regions. We further construct an inter-city influence network, identifying varying levels of impact and cross-city propagation.
These insights provide support for geographically targeted and network-aware air quality management strategies aligned with sustainable urban development goals.

References
[1] C. Sun, Y. Liu, L. Zhang, and X. Wang, "Quantifying contributions of geographical features to urban GDP outputs via interpretable machine learning," Sustainable Cities and Society, vol. 121, p. 105647, 2025.
[2] L. Li, S. Sun, L. Zhong, J. Han, and X. Qian, "Novel spatiotemporal nonlinear regression approach for unveiling the impact of urban spatial morphology on carbon emissions," Sustainable Cities and Society, vol. 125, Art. no. 106381, 2025.