

Mapping unmonitored urban pollution: an ST-GNN and Sentinel-5p fusion for air quality prediction

Kehkasha Fatema A. Dumasia¹, Saikat Kumar Paul²

¹Research Scholar, Dept. of Architecture and Regional Planning, IIT Kharagpur

²Professor, Dept. of Architecture and Regional Planning, IIT Kharagpur

kehkasha25@kgpian.iitkgp.ac.in

+91-9512540533

skpaul@arp.iitkgp.ac.in

+91-3222-283212

ABSTRACT

This study addresses the critical challenge of monitoring air quality in unmonitored urban sectors of **Ahmedabad, India**. It presents a hybrid **Spatio-Temporal Graph Neural Network (ST-GNN)** framework that fuses **Sentinel-5P** satellite data with urban spatial features and meteorological variables. By integrating **Graph Attention Networks (GAT)** for spatial dependencies with an optimized **XGBoost** regressor, the model achieves high predictive precision. The results demonstrate a R^2 of **0.727** for $PM_{2.5}$ and a R^2 of **0.809** for NO_2 . This framework provides a scalable solution for high-resolution, citywide pollution mapping to support data-driven urban policy.

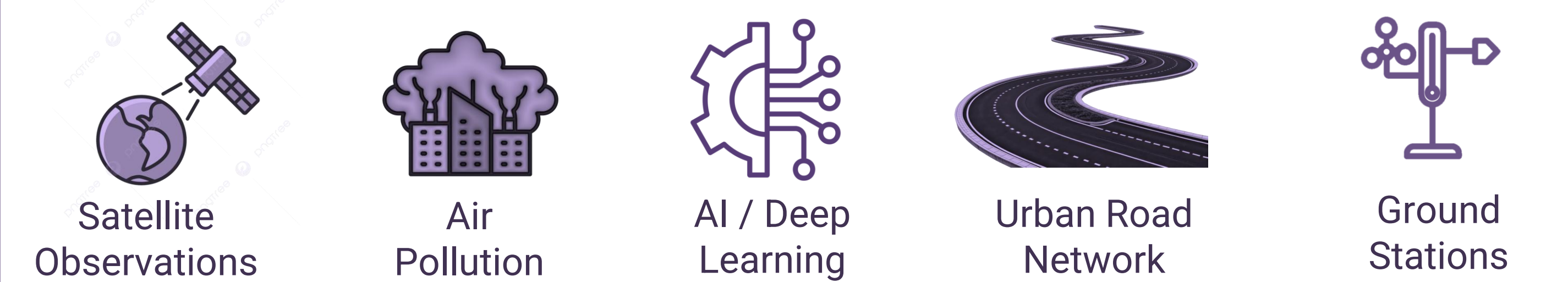
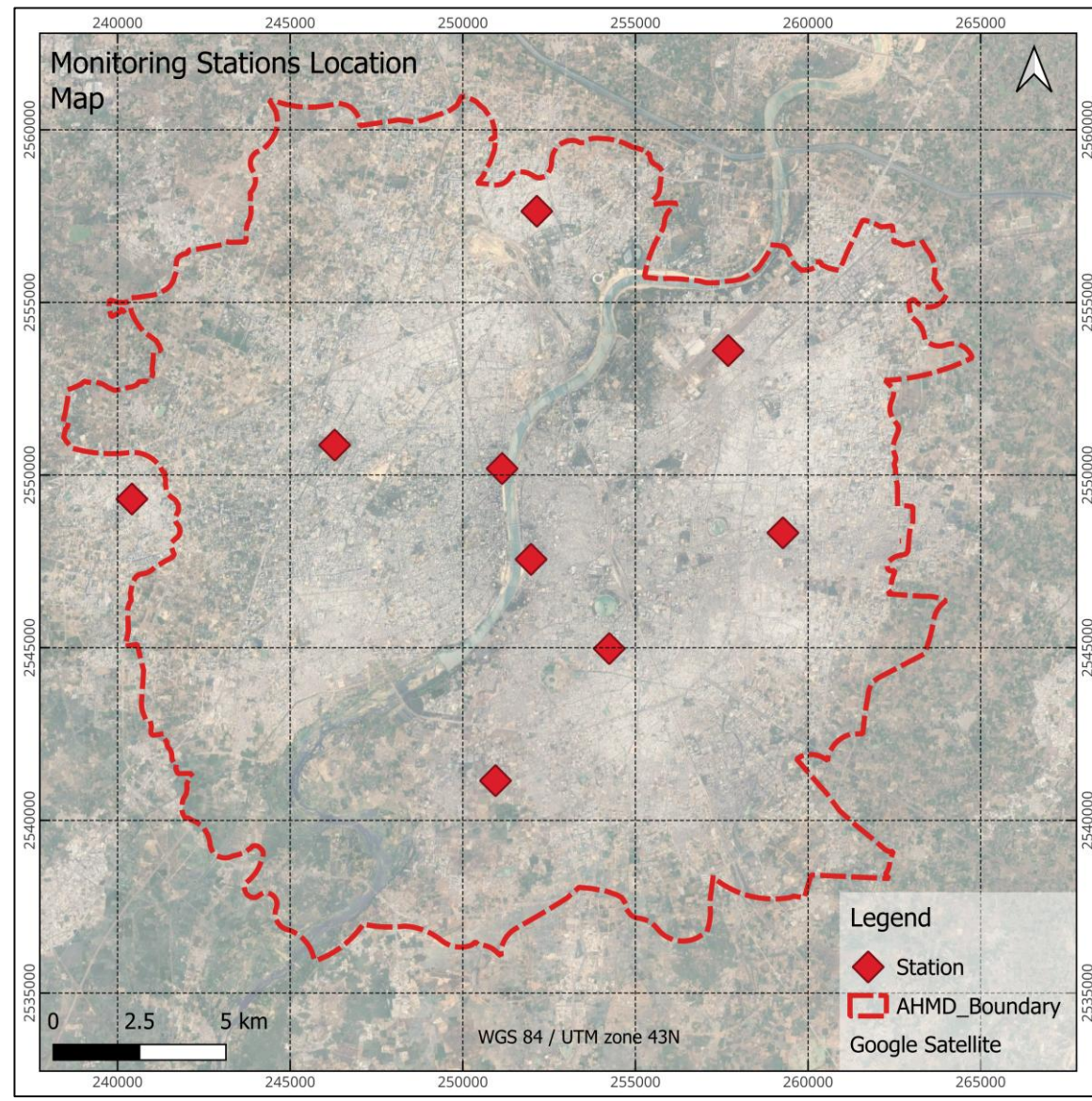
INTRODUCTION & AIM

Rapid urbanization and limited monitoring infrastructure create major gaps in understanding urban air pollution. While ground stations provide high-fidelity data, their physical footprint is limited, leaving large urban sectors "blind" to pollution spikes. Recent advancements in remote sensing and graph-based machine learning offer a pathway to fill these gaps by treating monitoring stations as nodes in a connected urban graph.

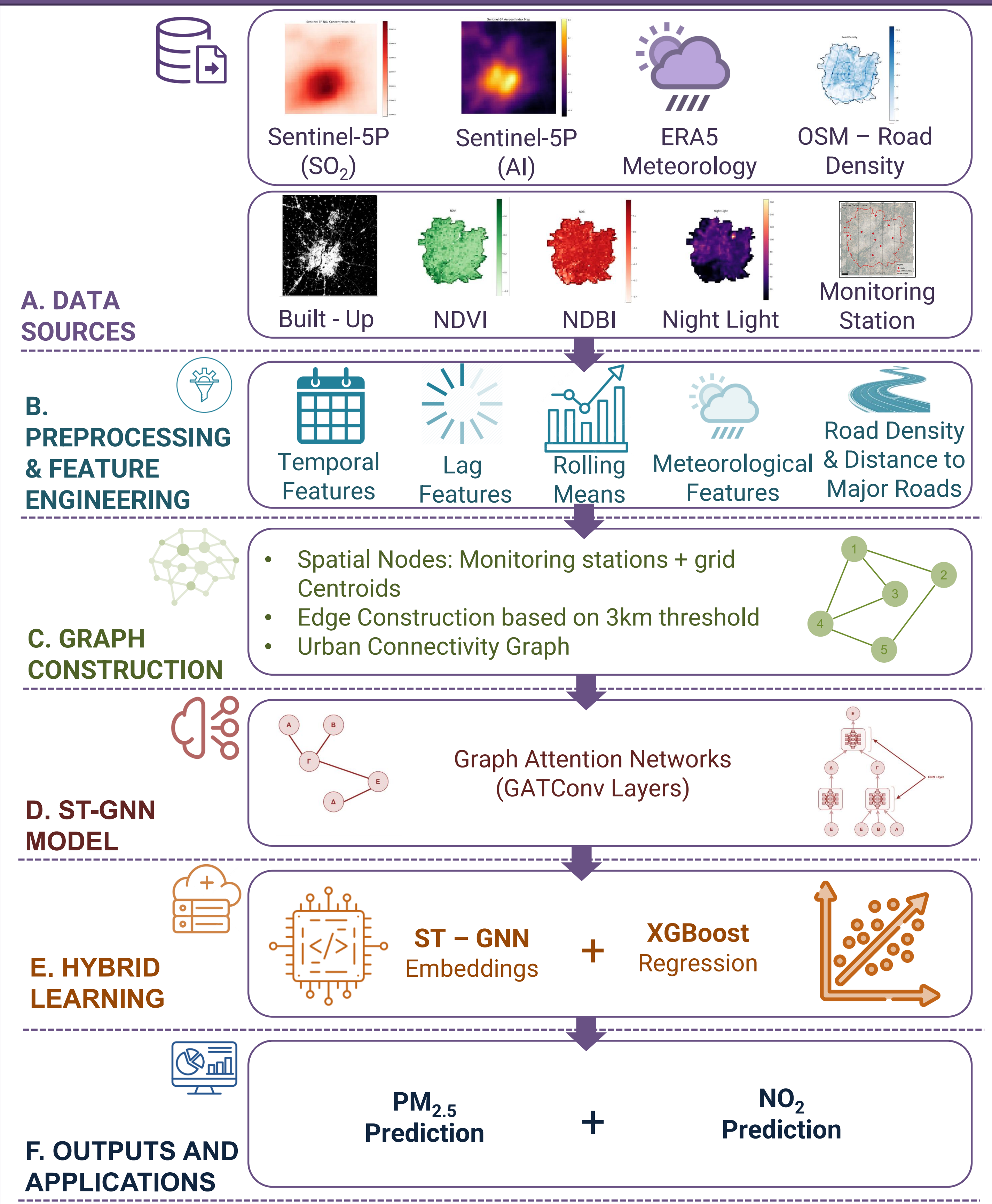
AIM:

To develop a hybrid Spatio-Temporal Graph Neural Network (ST-GNN) framework to:

- Integrate multi-source data including Sentinel-5P (NO_2 , Aerosol Index), ERA5 meteorology, and OpenStreetMap (OSM) road networks.
- Predict high-resolution city wide concentrations of $PM_{2.5}$ and NO_2 across Ahmedabad, India.
- Utilize SHAP (SHapley Additive exPlanations) to interpret the influence of meteorological and traffic-related features on urban pollution.



METHODOLOGY & FRAMEWORK



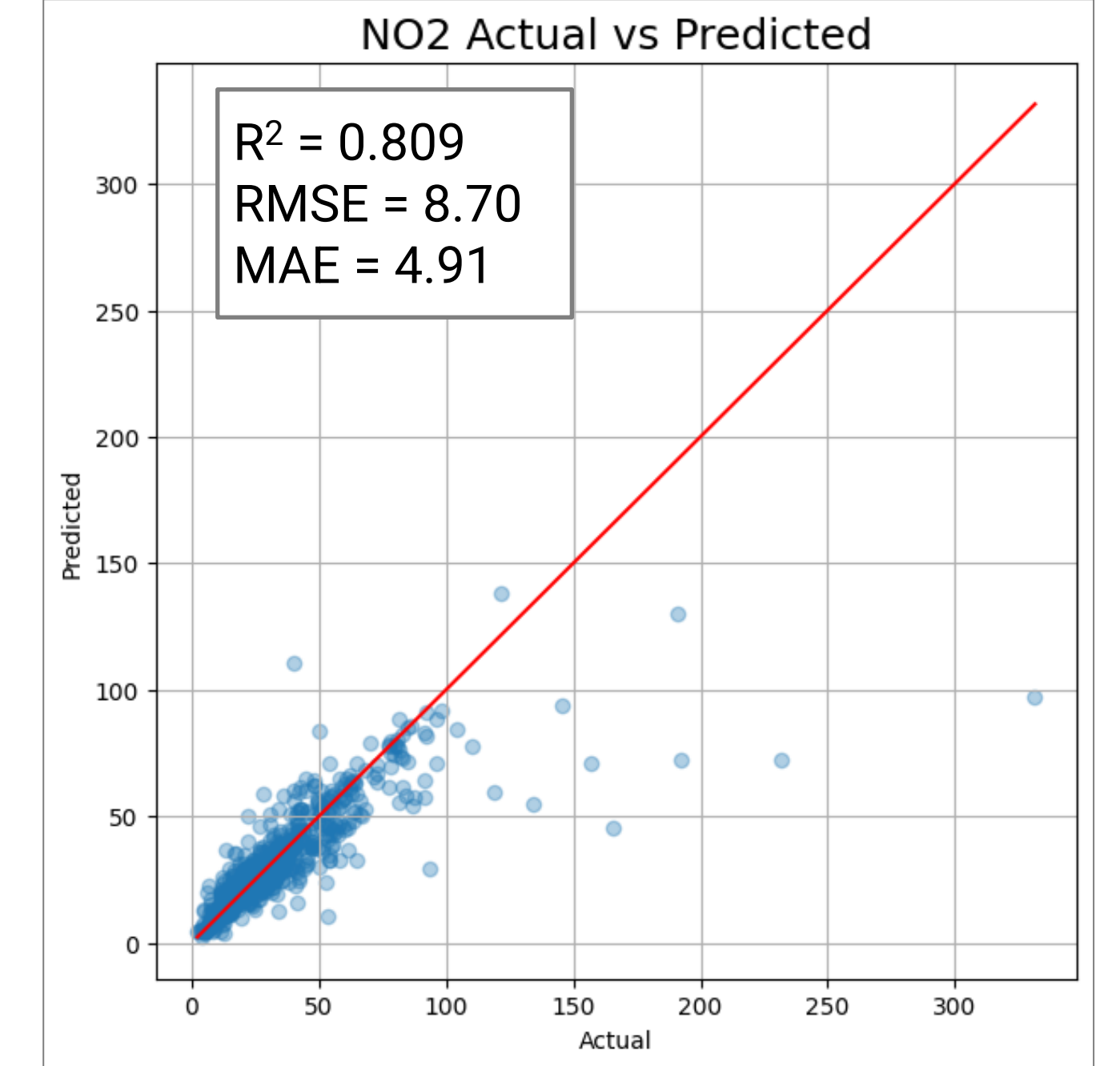
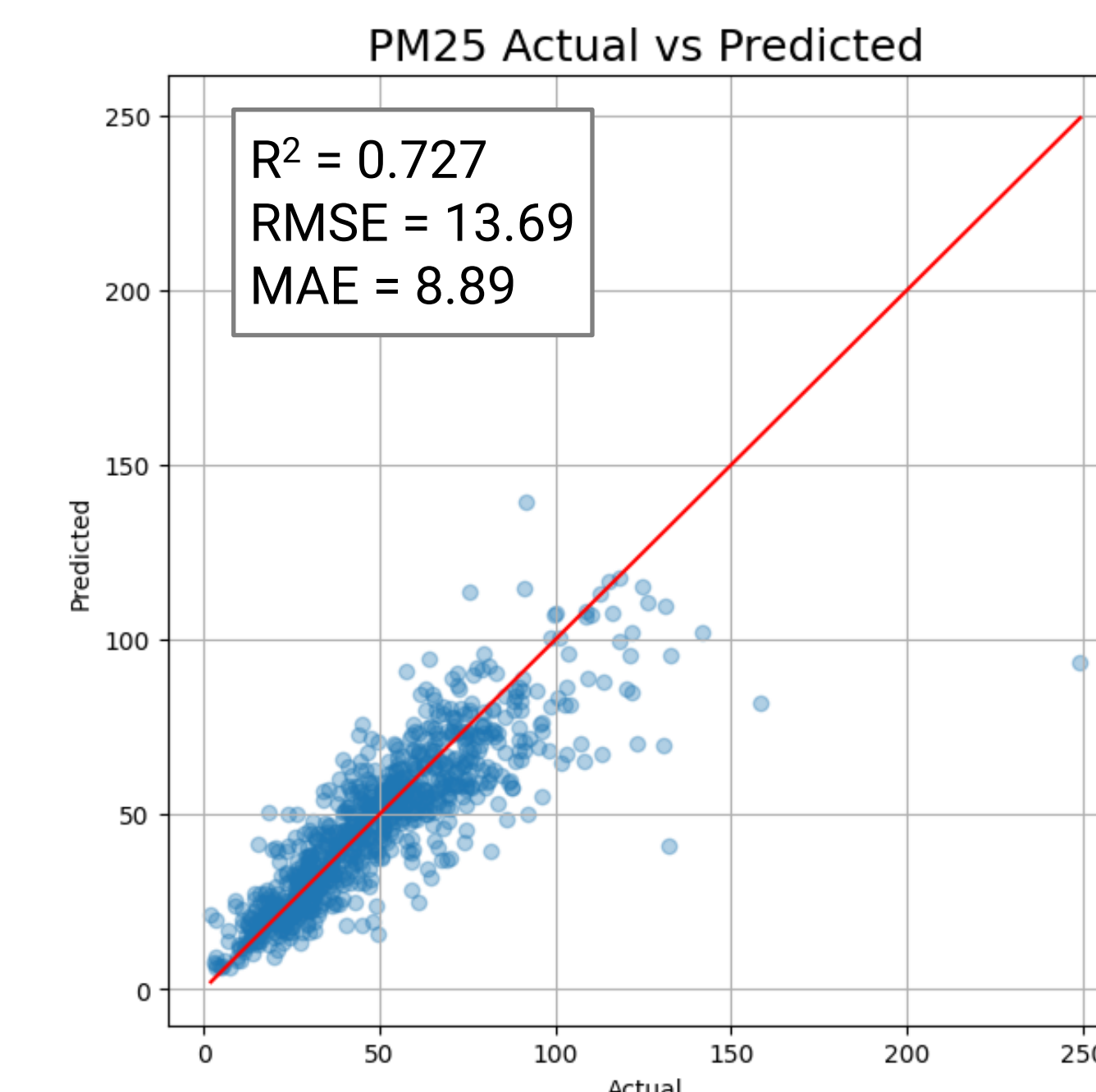
RESULTS & DISCUSSION

1. MODEL PERFORMANCE

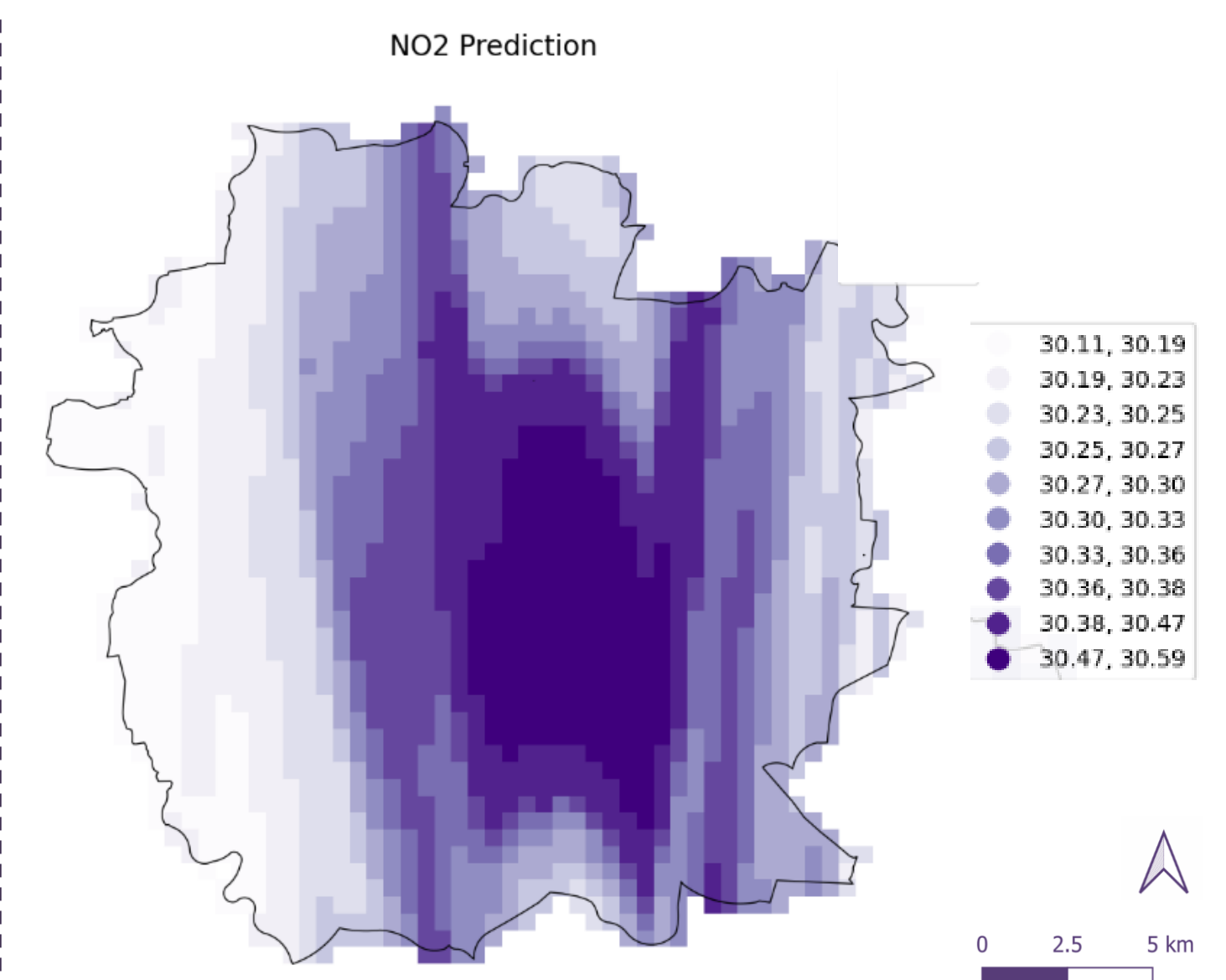
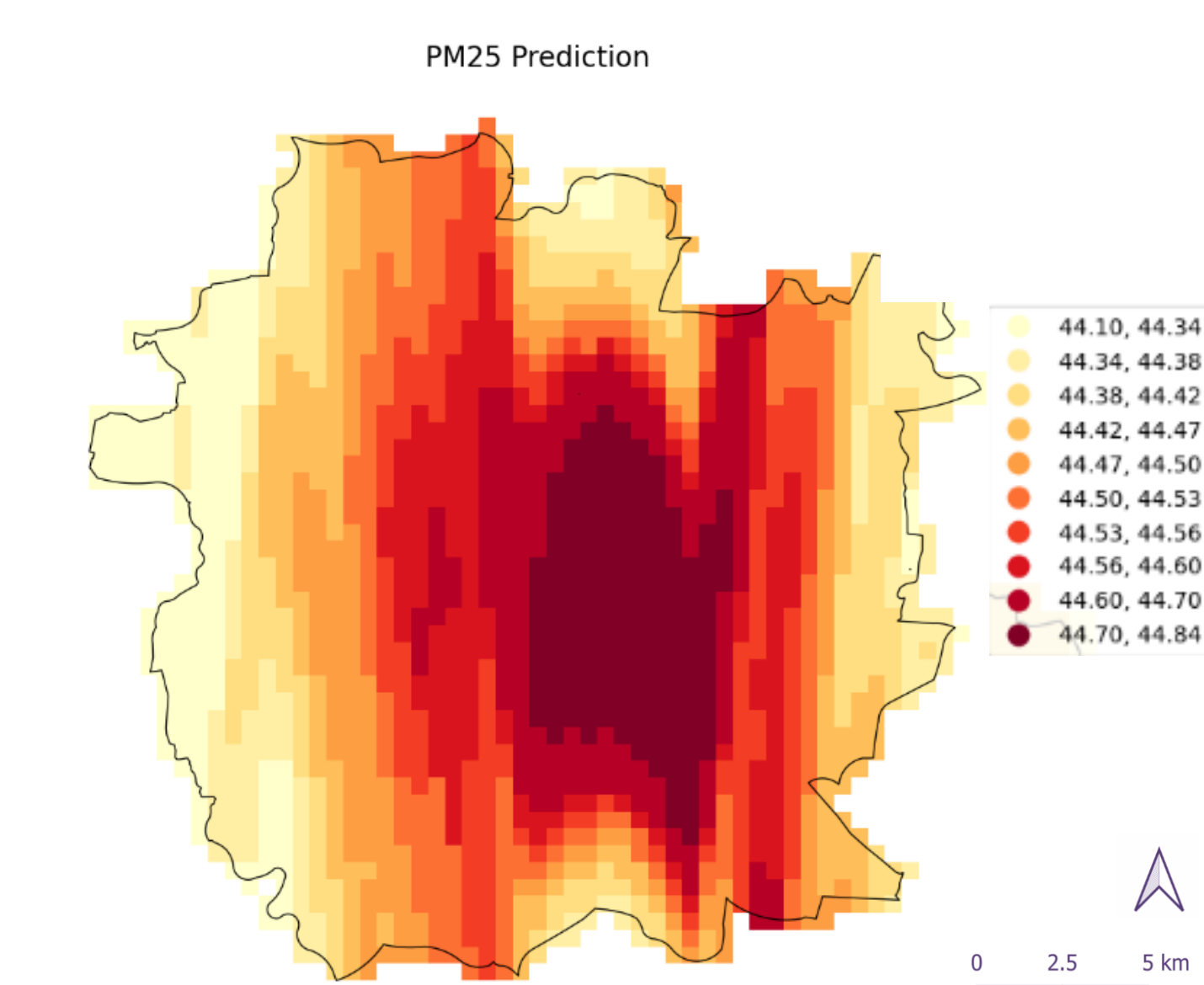
Pollutant	R^2	RMSE	MAE
$PM_{2.5}$	0.727	13.69	8.89
NO_2	0.809	8.70	4.91

Improved NO_2 model achieved strong predictive capability after optimised XGBoost refinement

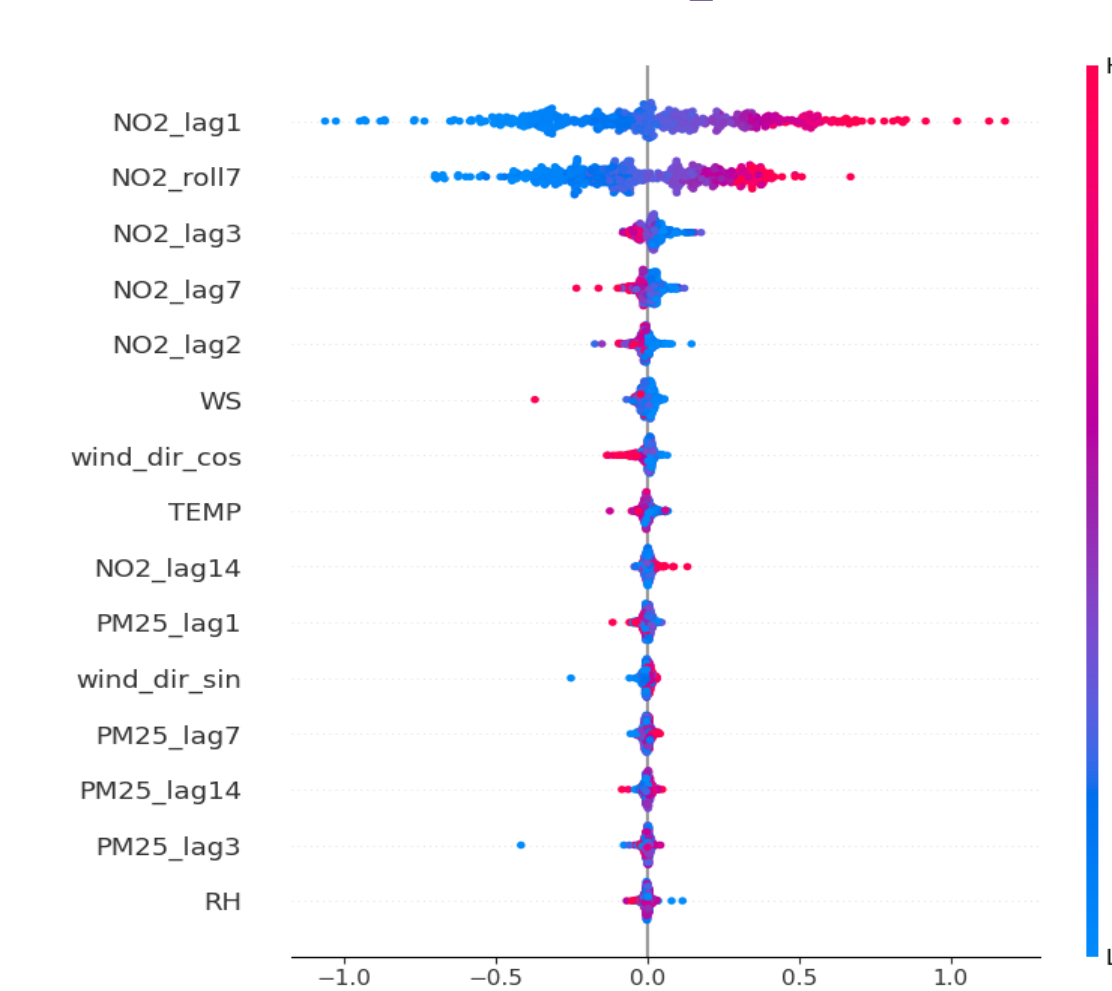
2. ACTUAL VS PREDICTED SCATTER PLOTS



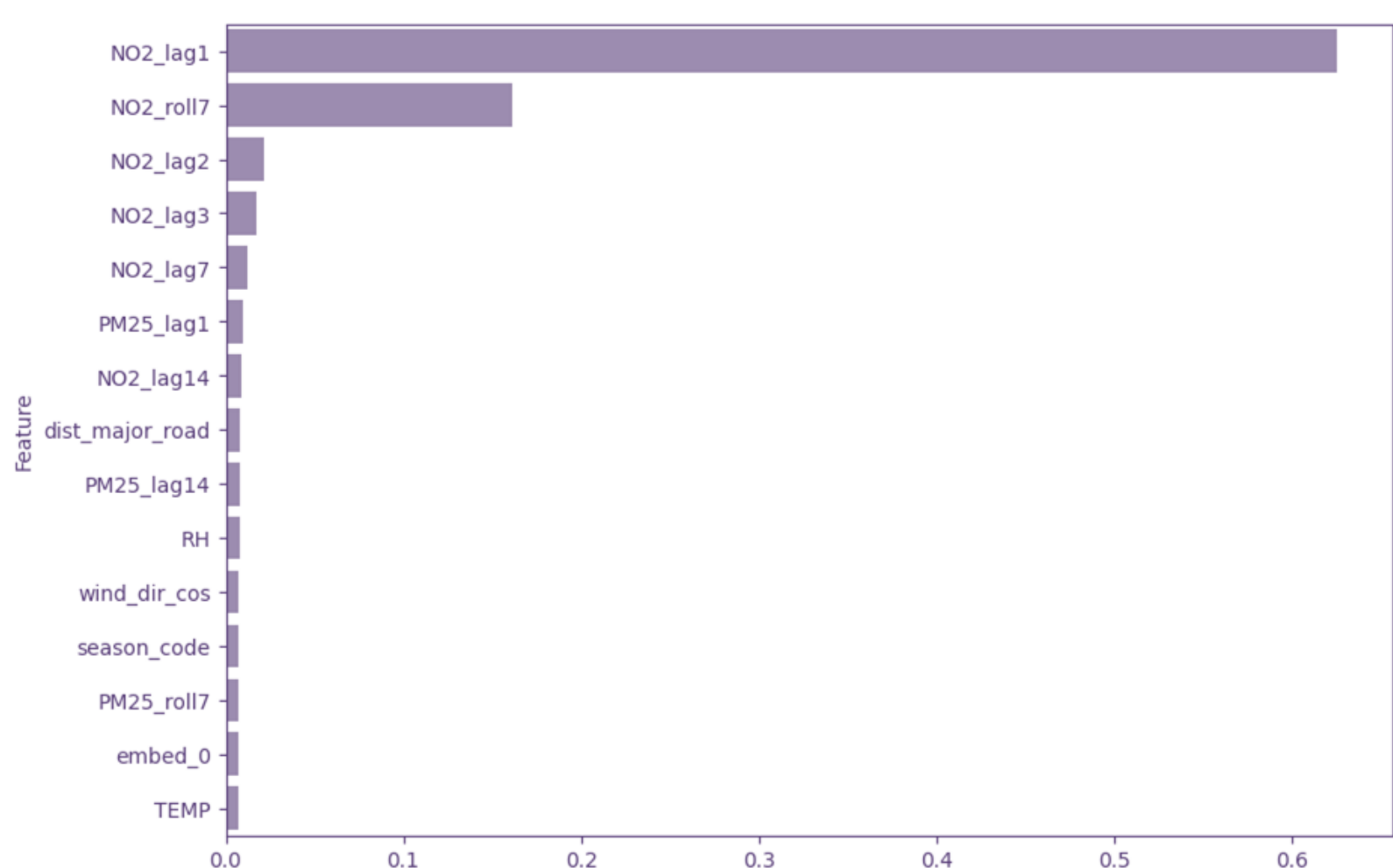
3. CITYWIDE PREDICTION MAPS (AHMEDABAD)



4. SHAP VALUE (NO_2 MODEL)



5. TOP FEATURE IMPORTANCE (XGBOOST)



Monitoring Station data, Road density and meteorological variables were dominant predictors.

CONCLUSIONS

- ✓ Hybrid ST-GNN + XGBoost successfully improved urban pollution prediction accuracy.
- 📡 Sentinel-5P and urban spatial datasets effectively enhanced unmonitored area estimation.
- 🎯 NO_2 prediction achieved high performance ($R^2 = 0.809$).
- 🌐 Graph-based spatial learning captured urban connectivity and pollution propagation patterns.
- 📊 The framework demonstrates strong potential for smart-city environmental monitoring applications.

FUTURE WORK / REFERENCES

- 🕒 Real-time pollution forecasting using streaming data and ST-GNN
- 📡 Integration with IoT sensor networks for dense observations
- 🏙️ Transfer learning for generalisation across cities and regions
- 📱 Mobile and web dashboards for public and policy support
- 🌐 Multi-pollutant forecasting (O_3 , CO , PM_{10} , SO_2 , etc.)
- Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. ICLR.
- ESA (2023). Sentinel-5P Data Product User Manual. European Space Agency.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. KDD.
- Li, L., et al. (2020). Spatio-temporal air quality prediction using deep learning: A survey. Environmental Pollution, 267.
- Zhang, Y., et al. (2021). Satellite remote sensing for urban air quality mapping: Progress and challenges. Remote Sens., 13(2).