Urban Mobility and Air Quality in UAE Cities: A Satellite-Driven Comparative Analysis

Manal Ali Ahmed Alteneiji, Hawra Mohammed S Al Sinan, Mariam Khalid Ali Yousif Alali , Asama Hassan Alhammadi
Department of Computing and Informatics,
University of Sharjah, UAE
Email: {U23103284, U22106610, U23102355, U23200052}@sharjah.ac.ae

***Abstract*—High-resolution human mobility and air-quality datasets offer unprecedented opportunities to analyze anthropogenic pollution drivers at fine spatial scales, yet fusing heterogeneous streams—smartphone mobility traces, ground sensors, and satellite imagery—remains computationally demanding. In this paper, we develop a Python-based, two-phase framework that integrates Google Community Mobility Reports, OpenAQ ground NO_2 measurements, and Sentinel-5P satellite rasters on a unified 5-character geohash grid, enabling scalable spatial aggregation. In Phase I (2020–2022), feature engineering (mobility change rates, log-population density), K-Means clustering, and ordinary least squares regression uncover mobility–pollution regimes and quantify their linear interdependencies. In Phase II (2023–2025), three synthetic mobility scenarios—proportional shifts, spatial perturbations, and population-weight variations—are generated and fed into a Gradient Boosting Regressor (100 estimators, max_depth = 3, learning_rate = 0.1) to capture non-linear mobility impacts on NO_2 concentrations. On held-out data, the Synthetic Mobility scenario achieves CV R^2 = 0.966 and Test R^2 = 0.959, reducing MAE to 1.0 × 10^-5 and RMSE to 1.4 × 10^-5—outperforming alternative strategies. This end-to-end pipeline produces fine-grained risk maps that empower urban planners and public-health officials to target air-quality interventions with minimal computational overhead.**

***Index Terms*—geohash; human mobility; air-quality forecasting; K-Means clustering; gradient boosting; scenario simulation.**

I. Introduction

The rapid proliferation of smartphone-based mobility tracking and dense air-quality sensing has generated unprecedented volumes of georeferenced data, promising new insights into how human movement drives urban pollutant concentrations [1]. Yet these heterogeneous streams—from Google Community Mobility Reports through OpenAQ ground-station NO_2 measurements to Sentinel-5P satellite rasters—arrive in disparate formats and resolutions, complicating efforts to integrate them at fine spatial scales. Natural experiments such as the COVID-19 lockdowns demonstrated the sensitivity of NO_2 levels to abrupt drops in workplace and transit activity [2], [3], revealing clear mobility–pollution couplings. However, most analyses to date have focused on historical correlation within coarse spatial aggregates or have employed linear models that fail to capture localized heterogeneity and non-linear interactions. Moreover, few end-to-end frameworks exist that both explain past pollutant dynamics and forecast future air-quality risk under varying mobility scenarios.

To fill these gaps, we develop a fully reproducible, two-phase Python pipeline that harmonizes multi-source mobility and pollution datasets on a common 5-character geohash grid, then applies clustering, regression, and boosting-based simulation to interpret past trends and project future hotspots. In Phase I (2020–2022), workplace mobility data, ground-station NO_2 readings, and satellite rasters are discretized into approximately 200 geohash zones; feature engineering (e.g., mobility change rates, log-population density) is followed by K-Means clustering to uncover mobility–pollution regimes and ordinary least squares regression to quantify their linear dependencies. In Phase II (2023–2025), three "what-if" scenarios—proportional mobility shifts, spatial perturbations, and population-weighted noise—are generated and fed into a GradientBoostingRegressor (100 trees, depth = 3, learning_rate = 0.1) to capture non-linear mobility impacts on NO_2 concentrations.

The principal contributions of this work are:

1) A unified geohash-based data-fusion framework for high-resolution mobility and air-quality datasets;
2) A two-phase analytical suite combining clustering, OLS regression, and gradient boosting to both explain past pollutant dynamics and forecast future risk; and
3) Fine-grained risk maps under alternative mobility scenarios, enabling urban planners and public-health officials to target interventions where they are most needed.

The remainder of this paper is organized as follows. Section II surveys related work on mobility-driven pollution modeling and spatial forecasting. Section III details our data sources, preprocessing steps, and methodological pipeline. Section IV presents experimental results and scenario analyses. Finally, Section V concludes and outlines avenues for future research.

II. Related Work

Urban air pollution remains a pressing issue shaped by multiple interacting systems. This section synthesizes insights from 14 peer-reviewed studies into three broad thematic areas: (1) Emission Sources and Exposure Pathways, (2) Urban Environment and Mobility Interactions, and (3) Monitoring, Forecasting, and Policy Tools. These categories reflect an integrative approach necessary to understand and mitigate the multifactorial nature of air pollution.

Emission Sources and Exposure Pathways

Pollutant concentrations in urban environments stem from a combination of localized and regional emissions. Dong et al. [4] combined air dispersion models with GIS to highlight vehicle emissions as primary contributors to SO_2 exposure in China. Chauhan et al. [5] quantified black carbon (BC) in Varanasi, identifying fossil fuels as dominant sources, with seasonal modulation by meteorology and biomass burning.

Amritha et al. [6] used satellite and reanalysis data to reveal a global decline in SO_2 emissions during the COVID-19 lockdown, particularly in high coal-use regions, underscoring the anthropogenic nature of major SO_2 hotspots. Obiefuna et al. [7] mapped pollutant variability across land-use types in Nigeria, emphasizing the impact of urban planning on exposure patterns.

Complementing these, a clustering analysis of multi-point in situ data identified distinct pollutant dispersion patterns and emphasized long-range transport effects [8].

Urban Environment and Mobility Interactions

Urban mobility and spatial configuration play pivotal roles in shaping air quality. Ghaffarpasand et al. [9] demonstrated the influence of traffic on $PM_{2.5}$ levels in Kampala, with roadside locations showing a sharper response to mobility restrictions. They employed a regression tree model, achieving an R^2 of 0.53, to capture the relationship between air pollution and mobility trends.

During the COVID-19 lockdown, Brunelli et al. [10] observed temporary improvements in air quality, advocating for structural changes like low-emission transport and congestion charges. From an infrastructural perspective, Mei et al. [11] modeled pollutant retention in street canyons, showing how thermal flows and skyline shapes influence vertical ventilation and accumulation. Similarly, a study on street cleaning activities revealed brief increases in particulate resuspension, but highlighted how proper timing can mitigate exposure [12].

Traffic-induced noise and pollution exposure modeling by Kephalopoulos et al. [13] demonstrated the benefit of integrating flow dynamics with environmental risk assessment tools.

Monitoring, Forecasting, and Policy Tools

Technological advances are reshaping air pollution monitoring and forecasting. Singh et al. [14] showcased how mobile sensors, combined with machine learning (Random Forest), can produce high-resolution pollution maps for urban interventions. The model achieved a mean absolute error (MAE) of 2.8 $\mu g/m^3$, effectively revealing fine-grained spatial pollution patterns across Philadelphia.

Sadeghi et al. [15] proposed a smart city framework that integrates IoT and citizen sensing for environmental monitoring, although no machine learning techniques were employed in the analytical pipeline. In contrast, Tang et al. [16] developed a hybrid deep learning model to nowcast ozone concentrations across East Asia. Their model achieved a test R^2 of 0.686 and a cross-validation R^2 of –0.459, revealing good in-sample performance but also potential overfitting, likely due to small datasets and target values near machine precision. In a parallel effort [16], the same authors refined the forecasting approach using emission and meteorological data, achieving consistent results ($R^2 \approx 0.68$).

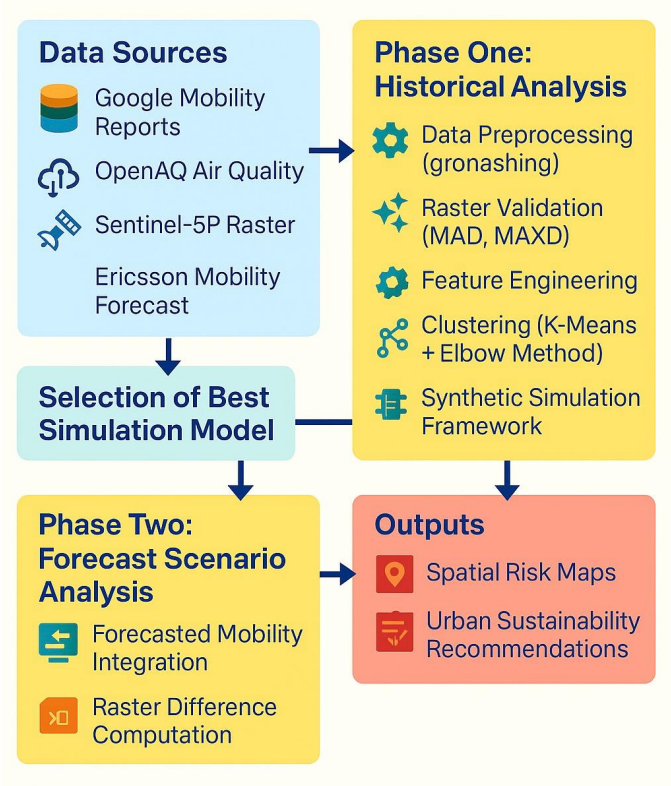

Fig. 1. Methodology Overview

These tools offer critical insights for both short-term air quality advisories and long-term urban environmental planning, supporting data-driven decision making and proactive pollution management. Table I summarizes the studies that incorporated machine learning into their analysis.

III. Methodology

A methodical integration of diverse datasets, including satellite observations, ground sensor data, and forecasted models, is necessary to investigate urban air quality and mobility [17], [18]. The COVID-19 pandemic provided a rare chance to investigate the effects of anthropogenic mobility decrease on pollution, highlighting the urgent need for interdisciplinary methods. This study employs a two-phase methodology: Phase One conducts a historical analysis based on observed data (2020–2022), and Phase Two projects future risk scenarios using forecasted data (2023–2025). Figure 1 illustrates the overall methodology flow adopted in this study.

A. Phase One: Historical Analysis Using Observed Data (2020–2022)

Data Acquisition and Preprocessing: We obtained four main datasets: Google's UAE Mobility Reports (2020–2022), OpenAQ ground sensor air quality measurements, satellite

TABLE I
SUMMARY OF MACHINE LEARNING-BASED STUDIES

Ref	Objective	Methodology	Results	ML Method and Accuracy	Data Type/Source
[5]	Analyze $PM_{2.5}$ patterns and mobility variation during COVID-19	Low-cost sensors + ML regression	10% $PM_{2.5}$ drop observed; mobility linked to roadside concentrations	Regression Tree; R^2 = 0.53	Low-cost air quality sensors + mobility data (Kampala)
[10]	Create pollution maps using mobile sensors and ML	Sensor data + ML mapping	Revealed fine-grained spatial patterns and hotspots	Random Forest; MAE = 2.8 $\mu g/m^3$	Mobile sensors + spatial air quality data (Philadelphia)
[12]	Improve real-time ozone prediction	Hybrid deep learning model	Model showed better short-term accuracy than traditional models	Hybrid DL model; R^2 = 0.686 (test), –0.459 (CV)	Satellite + meteorological + mobility data (East Asia)
[13]	Develop robust forecasting for urban ozone	Hybrid model with meteorology + emissions	Captured complex spatial/temporal ozone variability	Hybrid DL model; $R^2 \approx 0.68$	Meteorological + emissions + air quality data (East Asia)

pollution rasters (NO_2, SO_2, aerosol indices), and preliminary regional mobility forecasts. Air quality and mobility data were cleaned, parsed, and time-standardized to monthly intervals. To ensure consistent EPSG 4326 projection, raster datasets were processed using geospatial libraries. Following spatial pollution monitoring frameworks [19], raster data were clipped to the UAE boundary and preprocessed to mask invalid pixel values. Table II summarizes the datasets utilized in Phase 1, including their key features and corresponding data sources:

1) Feature Engineering and Geohashing: Normalized pollution indices, mobility change rates, and spatial interaction terms were produced by feature engineering. To facilitate effective spatial aggregation and analysis beyond localized urban zones, latitude and longitude coordinates were geohashed at a precision of 5–6 characters (about 1–5 km^2) [23]. This geospatial discretization enhanced computational tractability and enabled scaled spatial joins to be performed throughout the modeling and simulation stages

2) Raster Difference Analysis and Validation: Validation of satellite pollution rasters across years employed pixel-level statistical analysis. The Mean Absolute Difference (MAD) computed quantifies the average magnitude of pixel-level differences between two raster datasets, disregarding directionality. It provides a robust measure of the overall discrepancy between observed and forecasted pollutant concentrations. The MAD is computed as:

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i|$$

Where $X_i$and Y_i are corresponding pixel intensities in different years. Maximum Absolute Difference (MAXD) was also computed to capture extreme spatial changes [24]. In addition to reducing satellite noise, this ensured that the raster layers utilized for analysis caught actual atmospheric fluctuations.

3) Clustering Analysis and Optimal Cluster Selection: After normalizing all variables, the mobility-pollution geohash features were subjected to K-Means clustering to reveal underlying urban trends. The optimal number of clusters was selected using the Elbow Method, minimizing within-cluster sum of squares and balancing model complexity and interpretability [25]. The formal expression for K-Means clustering is to minimize the variance inside each cluster, which is as follows:

$$\arg\min_{S}\sum_{i=1}^{k}\sum_{X\in S_i}\|X-\mu_i\|^2$$

where μ_i represents the centroid of the cluster S_i.

4) Predictive Modeling with Linear Regression: OLS linear regression was used to estimate the annualized workplace mobility and NO_2 averages for 2020–2022. The fit was assessed using the coefficient of determination (R^2 score). OLS linear regression was used to estimate the annualized workplace mobility and NO_2 averages for 2020–2022. For Phase Two validation, baseline estimates were generated by extrapolating to 2023–2025.

5) Simulation Framework and Boosting Techniques: We created a simulation framework with three synthetic scenarios to evaluate the model's resilience:

- **Synthetic Mobility Scenarios**
  We introduced controlled variations in observed mobility data M_{obs} to simulate fluctuations in urban movement. The synthetic mobility M_{sim} was generated by applying proportional adjustments:
  $$M_{\text{sim}} = M_{\text{obs}} \times (1+\delta)$$
  Here, δ represents synthetic changes of $\pm 10\%$, $\pm 20\%$, and $\pm 30\%$.
- **Spatial Perturbations**
  To simulate real-world disturbances like zoning regulations or new infrastructure, we randomly reassigned the mobility values between nearby geohashes [26]. This examined how sensitive the model was to regional spatial changes.
- **Population Weight Perturbations**
  To replicate demographic fluctuations, we applied normally distributed noise to the population weights P. The perturbed population P' was calculated as:
  $$P' = P \times (1+\gamma)$$
  where $\gamma \sim \mathcal{N}(0,\sigma^2)$ follows a normal distribution, introducing realistic randomness.

We employed boosting approaches that approximated second-order interactions between geohash clusters, mobility, and pollution to improve the simulations [27]. This allowed the framework to capture non-linear urban dynamics, improving its adaptability to complex real-world scenarios. Among the simulated models, the model that performed the best based on evaluation parameters such as mean absolute difference (MAD), standard deviation (σ), and geographic consistency with historical data was chosen for Phase Two forecasting and scenario risk evaluation.

TABLE II
DATASETS USED FOR PHASE ONE (HISTORICAL ANALYSIS, 2020–2022)

Dataset	Features / Variables	Source
Google Mobility Reports	Workplace mobility, Retail mobility, Transit station mobility, Residential mobility trends	[20]
OpenAQ Air Quality Data	NO_2 concentration, PM2.5 concentration, Timestamp, Location (Latitude, Longitude)	[21]
Sentinel-5P Pollution Raster Data	NO_2 Vertical Column Density, SO_2 Density, Aerosol Index, Multiband raster images (2020–2022)	[22]

B. Phase Two: Forecast Scenario Analysis Using Projected Data (2023–2025)

1) Forecast Data Integration and Preprocessing: As standins for human activity, projected Ericsson Mobility statistics (GB/user/month) for 2023–2025 were used. Forecast rasters for NO_2, SO_2, and aerosol indices that were generated from satellites were loaded concurrently. Forecasted pollution rasters were matched with historical geohash zones created in Phase One, per best standards for spatial forecasting integration [17]. Table III summarizes the datasets utilized in Phase Two, including their key features and corresponding data sources:

2) Raster Comparative Evaluation and Change Quantification: Forecasted raster values were compared against historical baselines using MAD and MAXD metrics. Additionally, the standard deviation (σ) was computed to measure the dispersion of forecasted pollution values across geohash zones relative to the mean. It provides insight into the spatial variability and uncertainty associated with future pollution patterns. A higher standard deviation indicates greater variability in predicted pollution intensities. The standard deviation is calculated as:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

This enabled magnitude and uncertainty evaluation of projected pollution changes [24].

3) Visual Analytics and Interpretation: Visual analytics included time series plots, raster heatmaps, cluster overlays, and spatial standard deviation maps. These visualizations emphasized trends and variation among UAE cities [29]. Visual analytics included:

- **Time Series Plots:**
  Mobility and composite pollution indicators are plotted annually (2020–2025) in comparison, with the observed and forecasted years separated.
- **Raster Heatmaps:**
  Visualizations of pixel-level differences between historical and forecasted pollution levels [29].
- **Cluster-Enhanced Maps:**
  K-Means clusters derived in Phase One were overlaid on forecast maps to assess differential urban risk evolution.
- **Standard Deviation Spatial Maps:**
  Zones with the highest forecast variability were identified as potentially unstable environmental risk zones.

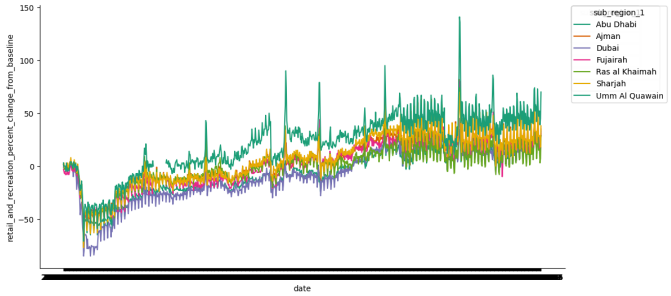

Fig. 2. Time series of retail and recreation mobility changes from 2020 to 2022 across UAE emirates. Abu Dhabi and Dubai show the largest activity rebounds post-2020 lockdowns.

4) Scenario Risk Evaluation: A methodology for evaluating risk was developed by integrating mobility growth rates, forecasted pollution deterioration, and variability measures. High-risk geohashes were flagged as priorities for urban mitigation strategies, aligning with recommendations from recent urban pollution scenario studies [18], [24].

IV. RESULTS AND DISCUSSION

A. Historical Phase Results (2020–2022)

1) Urban Air Quality Overview (2020–2022): An initial analysis of urban air quality and mobility trends in the UAE from 2020 to 2022 was conducted using data aggregated by geohash which is a spatial encoding system that divides geographic coordinates into small, grid-like cells. Each geohash in this study represents an area of approximately 3 km^2, allowing for detailed tracking of local environmental changes. Figure 2 shows changes in retail and recreation mobility across all seven emirates during this period. The sharp decline in early 2020 reflects pandemic lockdowns, with Abu Dhabi and Dubai showing the strongest rebounds in activity levels in the years that followed. To generate mobility scores at the local level:

- Latitude and longitude values were converted to 5-character geohashes using the geohash2.encode() function.
- A composite mobility score was calculated by averaging workplace and transit station changes from Google's mobility dataset.

TABLE III
DATASETS USED FOR PHASE TWO (FORECAST SCENARIO ANALYSIS, 2023–2025)

Dataset	Features / Variables	Source
Ericsson Mobility Forecast	Forecasted GB/user/month data for the Middle East and Africa, including the UAE	[28]
Sentinel-5P Forecasted Pollution Raster Data	Predicted NO_2 Vertical Column Density, SO_2 Density, Aerosol Index (2023–2025)	[22]

- Incomplete or missing data entries were removed to ensure accuracy.
- Data was then grouped by geohash, yielding 205 unique zones with mobility data.

Pollution data that was sourced from OpenAQ and similar platforms, was also aggregated by geohash. This provided average concentrations of key pollutants including NO_2, SO_2, and PM_{10} for each area. Geohash codes were normalized to lowercase across both datasets to allow for clean merging.

After data cleaning, only 7 geohash zones had valid entries in both the pollution and mobility datasets. These overlapping zones were merged into a final filtered_df, which served as the basis for the joined_table used in later visualizations and statistical modeling. As shown in Table IV, a preview of the combined dataset shows the structure of available variables.

TABLE IV
PREVIEW OF THE COMBINED DATASET SHOWING POLLUTION CONCENTRATIONS AND MOBILITY SCORES FOR SELECTED UAE GEOHASH ZONES (2020–2022).

Geohash	NO_2 (µg/m³)	PM_{10} (µg/m³)	SO_2 (µg/m³)	Mobility Score (%)
thqf4	76.8	94.6	12.5	12.10
thqfb	31.1	87.1	11.0	-11.81
thr97	19.2	67.2	10.4	-13.07

2) Geospatial Distribution of Pollutants: To study how air pollution varied across the UAE, georeferenced raster data for NO_2, SO_2, and the aerosol index (AAI) from 2020 to 2022 were visualized. A composite heatmap was created in Figure 3, where brighter colours show higher pollution levels. These maps were produced using Rasterio and Matplotlib, and they highlight areas with more serious air quality risks.

The highest pollution levels, especially for NO_2 and PM_{10}, were observed in Abu Dhabi and Dubai. These areas align with heavy traffic and industrial zones. However, SO_2 hotspots were found to be more scattered and did not closely follow urban mobility trends, suggesting they come from fixed sources like oil, gas, or power facilities. Also, remote and desert areas showed consistently low pollution, serving as a clear baseline for comparison. These spatial patterns were used as the basis for the zone-based clustering and forecasting discussed in later sections.

3) Correlation Between Mobility Scores and Pollutant Concentrations: To explore the relationship between urban activity and air quality, the correlation between mobility scores and NO_2 levels was analysed across geohash-based zones. It was expected that lower mobility such as during lockdowns would

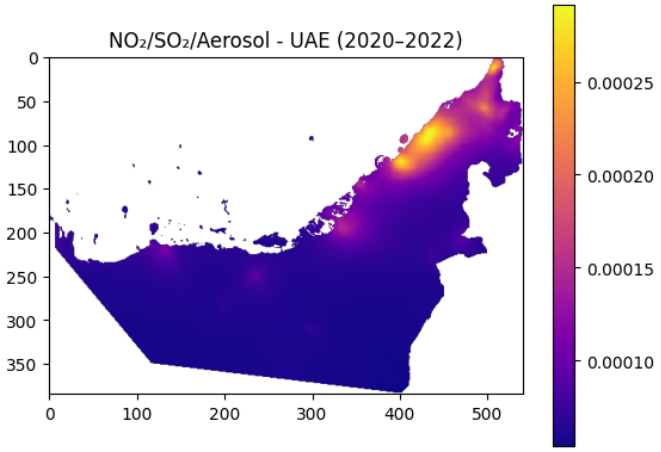

Fig. 3. Geospatial distribution of NO_2, SO_2, and Aerosol concentrations in the UAE (2020–2022). Brighter colors indicate higher pollutant levels.

be linked to lower NO_2 pollution caused by traffic. A summary of the data showed that NO_2 levels ranged from 7.3 µg/m^3 to 76.8 µg/m^3, while mobility scores ranged from –27.5% to +12.1%. According to the WHO, the safe limit for NO_2 is 40 µg/m^3. Only three zones which are 'thqem', 'thqf4', and 'thqf8' went above this level. The relationship is shown in Figure 4, where an overall inverse pattern can be seen: zones with lower mobility often had lower NO_2. This suggests that traffic emissions are a major source of NO_2 pollution in urban areas of the UAE.

However, an exception was observed in zone 'thqf4', which recorded the highest NO_2 level (76.8 µg/m^3) even though its mobility score was +12.1%. This unusual case suggests that non-traffic sources such as industrial activities or construction may be affecting air quality in that area. Identifying such outliers is important for designing targeted and effective environmental policies.

In addition to pairwise scatterplots, a correlation heatmap was generated to summarize the statistical relationships among pollutant levels and mobility scores across the filtered geohash zones (Figure 5). The heatmap reveals several key patterns such as:

- A moderate positive correlation (r = 0.52) is seen between NO_2 and PM_{10}, indicating that these pollutants may often co-occur in high-traffic or industrial zones.
- SO_2 shows a strong correlation (r = 0.84) with mobility scores, suggesting possible confounding factors such as

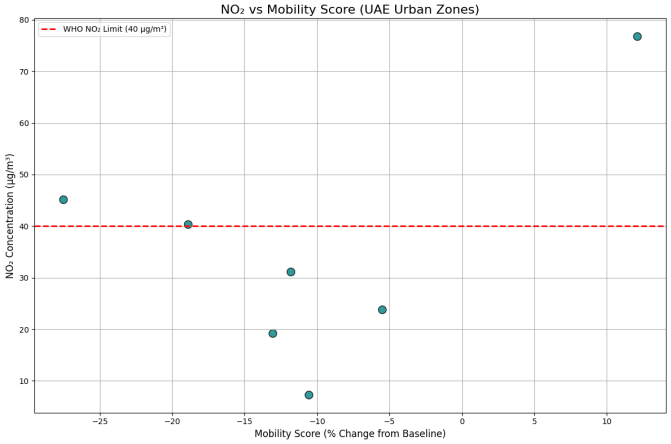

Fig. 4. NO_2 concentrations plotted against mobility score across UAE urban zones. The red dashed line indicates the WHO recommended limit (40 µg/m^3). Zone thqf4 appears as a significant outlier with high NO_2 despite positive mobility.

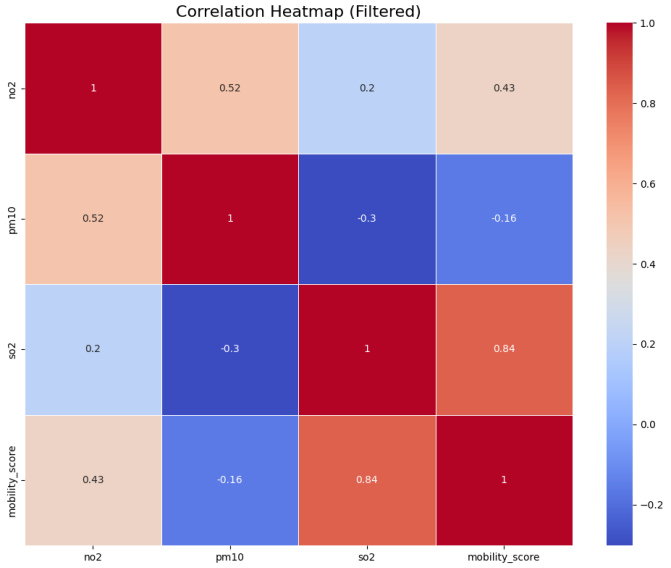

Fig. 5. Correlation heatmap showing relationships among NO_2, PM_{10}, SO_2 concentrations and mobility scores in overlapping geohash zones. Stronger correlations are shown in darker red or blue depending on direction.

industrial zones with both workforce mobility and stationary emissions.

- PM_{10}, however, displays a weak or negative correlation with both SO_2 (r = –0.30) and mobility (r = –0.16), implying it may arise from different sources like dust or construction.

These correlation patterns help distinguish between pollutants driven by mobile sources like NO_2, and those linked to fixed-point emissions like SO_2, guiding future modeling and policy decisions.

4) Critical Zones with Persistent Pollution: While a general inverse relationship between mobility and NO_2 concentrations was observed across urban zones in the UAE, several zones did not follow this trend, suggesting the presence of additional pollution sources unrelated to human movement. Most zones aligned with the expected pattern meaning lower mobility was associated with reduced NO_2 levels. However, three geohash zones exceeded the WHO safety threshold for NO_2 (40 µg/m^3), as listed in Table V. Notably, zone 'thqf4' recorded the highest NO_2 concentration (76.8 µg/m^3) despite a positive mobility score (+12.1%), identifying it as a key environmental outlier.

TABLE V
URBAN ZONES EXCEEDING THE WHO NO_2 LIMIT (40 µG/M^3) WITH CORRESPONDING MOBILITY SCORES.

Geohash	NO_2 (µg/m^3)	Mobility Score (% change)
thqem	40.40	–18.93
thqf4	76.80	+12.10
thqf8	45.15	–27.53

This anomaly was visualized in the NO_2 vs. mobility scatterplot (Figure X in Section 4.1.3), where zone 'thqf4' appears as a clear outlier positioned above the WHO-recommended threshold line. While most zones demonstrated a decrease in NO_2 levels alongside reduced mobility, 'thqf4' deviated from this pattern. This suggests that its elevated pollution levels are likely driven by stationary emission sources such as industrial facilities, large-scale construction sites, and power generation infrastructure.

These findings highlight the importance of considering land use and infrastructure zoning when interpreting mobility, pollution correlations and designing environmental intervention strategies. To pinpoint the most affected areas:

- Zones with NO_2 levels above 40 µg/m^3 were extracted using: high_no2 = joined_table.where('no2', are.above(40))
- Among the seven merged geohash zones, three (thqf4, thqem, and thqf8) exceeded the WHO threshold.
- Of these, only thqf4 had both elevated mobility and extreme NO_2 levels, marking it as the most critical environmental hotspot in the historical dataset.

5) Discussion on Non-Mobility-Related Pollution Sources: While NO_2 levels showed a moderate inverse correlation with mobility scores which is indicating a link to traffic-related emissions, SO_2 displayed no significant correlation, either spatially or statistically. This lack of association suggests that SO_2 pollution is driven by sources unrelated to human movement. Likely contributors include oil refineries and natural gas facilities, industrial manufacturing zones, shipping ports and logistics hubs, and cement production and large-scale construction sites.

This observation was visually supported by the SO_2 raster heatmap shown in Figure A, where SO_2 hotspots appeared in areas that did not coincide with densely populated or high-mobility urban zones. Notably, elevated SO_2 levels were recorded in regions such as Abu Dhabi, Ruwais, and Sharjah's industrial belt, even though minimal changes in mobility were observed in these areas.

Such a spatial disconnect between mobility and SO_2 concentrations aligns with global air quality research, which

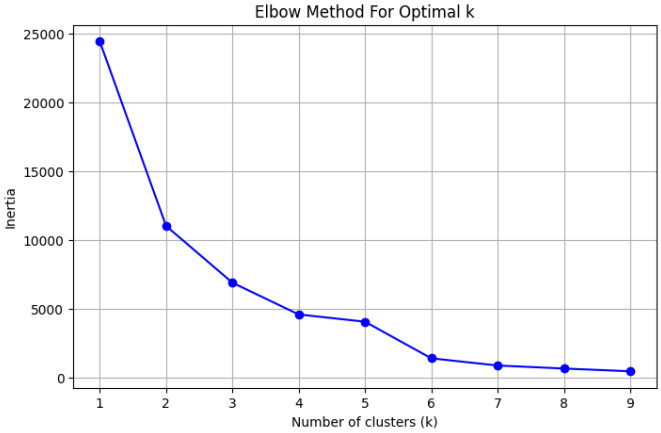

Fig. 6. Elbow Method for Optimal k using mobility scores. The elbow at k = 3 suggests a meaningful separation of urban zones based on movement patterns.

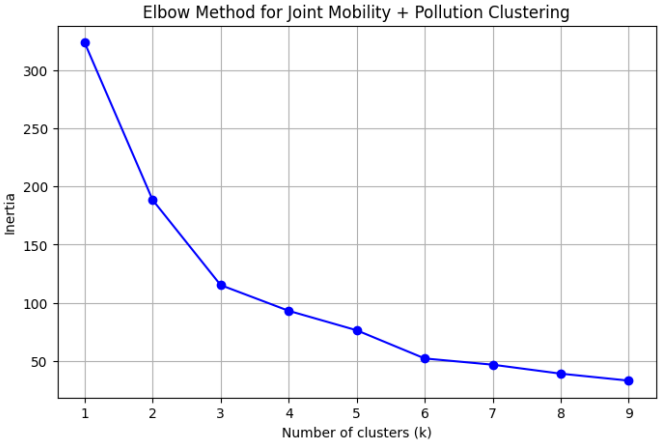

Fig. 7. Elbow Method for Joint Mobility + Pollution Clustering. The inflection at k = 3 reinforces the choice of three clusters for integrated analysis.

consistently identifies fossil-fuel combustion at fixed industrial locations, particularly in power generation and heavy industry, as the dominant source of SO_2 emissions [30]. To address this challenge, urban air quality strategies should be broadened beyond mobility-centered interventions. A comprehensive mitigation framework should include:

- Stricter emissions regulations and caps for SO_2-producing industries;
- Real-time satellite + ground monitoring integration (e.g., TROPOMI, Sentinel-5P);
- Zoning policies to create industrial buffers between high-output facilities and residential areas;
- Mandatory pollution disclosure and audits from large-scale industrial operators.

Without addressing these non-mobile sources, efforts focused solely on transportation reform for example, electrification, modal shifts, will fall short of delivering long-term public health and environmental benefits.

B. Zone-Based Clustering

1) K-Means Clustering of Zones Based on Joint Pollution and Mobility:: To better understand urban zone behaviour, K-Means clustering was applied using mobility score data from 2020 to 2022. This approach grouped zones with similar historical movement patterns. Before determining the number of clusters, the Elbow Method was applied to measure model inertia for k values ranging from 1 to 9. The initial elbow plot (Figure 6) focused on mobility scores alone and showed a distinct inflection at k = 3, indicating three optimal clusters.

To validate this result under combined analysis, the Elbow Method was repeated using both mobility and pollution variables (NO_2, SO_2, PM_{10}). The updated plot (Figure 7) also pointed to k = 3, confirming that three clusters balance complexity and interpretability when accounting for both behavioral and environmental data.

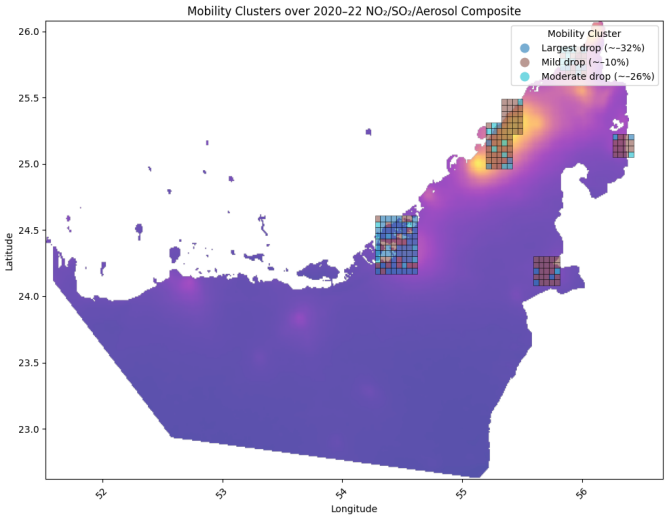

Fig. 8. Spatial distribution of mobility clusters (k = 3) overlaid on NO_2/SO_2/Aerosol Composite (2020–2022). Zones are color-coded by mobility cluster label.

Once k = 3 was selected, each geohash zone was assigned a cluster label. To improve interpretability, the clusters were described as:

- Cluster 0: Mild mobility drop (˜–10%)
- Cluster 1: Largest mobility drop (˜–32%)
- Cluster 2: Moderate mobility drop (˜–26%)

Each geohash zone was then spatially mapped using decoded polygons from geohash2.decode_exactly() and visualized using a GeoDataFrame. The clusters were overlaid on a composite pollution raster (NO_2, SO_2, and aerosol index), as shown in Figure 8, offering a spatial view of how mobility behaviors align with pollution intensity across the UAE.

2) Hotspot Detection and Zone Typologies:: To better contextualize the clustering results from Section 4.B.1, geohash zones were further classified into four environmental–behavioral typologies, based on their combined mobility trends and average pollution levels (primarily NO_2 and PM_{10}).

These typologies were generated by cross-referencing the K-Means mobility clusters with corresponding pollutant concentrations.

- **Type A: High Mobility – Low Pollution**
  Examples: Central downtown zones with efficient public transport such as metro corridors in Dubai.
  Implication: These areas demonstrate environmental efficiency and are ideal candidates for further green mobility investments.
- **Type B: Low Mobility – High Pollution**
  Examples: Industrial belts, oil refineries, and ports with limited residential movement
  Implication: Pollution is likely driven by stationary sources, requiring regulatory oversight and fixed-source monitoring.
- **Type C: High Mobility – High Pollution**
  Examples: Traffic corridors, commercial hubs, or highway-dense city centers
  Implication: Emissions are likely tied to vehicular traffic, suggesting a need for policies like transport electrification and congestion pricing.
- **Type D: Low Mobility – Low Pollution**
  Examples: Suburban or fringe zones with low population density
  Implication: These areas are generally low-risk but may serve as future urban expansion zones or ecological buffers.

These categories were created by analyzing cluster labels, average pollution scores, and mobility change distributions.

To support this classification:

- Figure 9 shows the elbow method used to determine the optimal number of clusters based on population-weighted absolute mobility drop.
- Figure 10 visualizes spatial groupings of geohash zones colored by mobility drop severity, helping identify zones with high social impact.
- Figure 11 presents a scatterplot of zones clustered jointly by mobility and pollution metrics, showing three main groups: Low–Low, High–High, and Mid–Mid.
- Figure 12 maps these joint clusters spatially across the UAE, overlaid on the NO_2/SO_2/aerosol composite raster.

Figures E through U illustrate the process of refining zone typologies by analyzing both behavioral (mobility) and environmental (pollution) characteristics. Figure E presents the Elbow Method applied to population-weighted absolute mobility drop, where a clear inflection at k = 3 supports the use of three clusters to differentiate zones by severity of mobility reduction. Figure R maps these clusters across the UAE, showing how zones with high population impact are spatially concentrated in key urban centers such as Dubai and Sharjah. Figure T provides a scatterplot of zones jointly clustered by mobility change and average pollution levels, identifying three main groupings: Low–Low, High–High, and Mid–Mid, which directly support the Type A–D typology. Finally, Figure U spatially visualizes these joint clusters on a pollution raster,

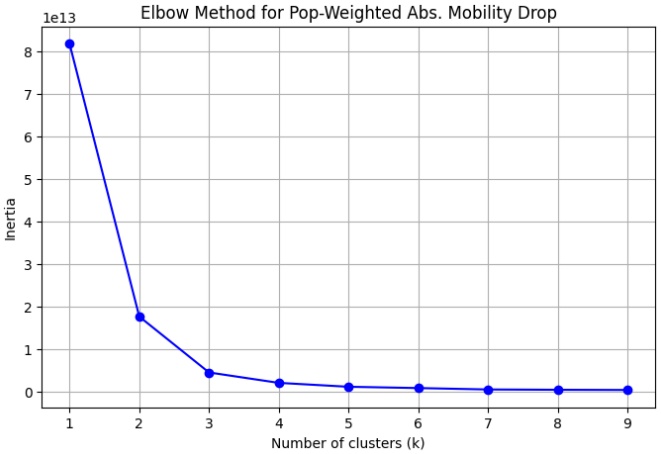

Fig. 9. Elbow method to identify optimal k for clustering zones based on population-weighted absolute mobility drop. Suggests that 3 clusters yield stable segmentation.

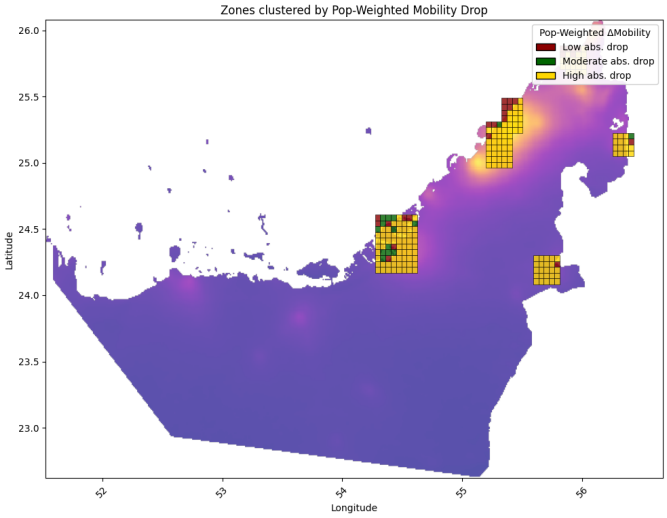

Fig. 10. Spatial clustering of geohash zones by population-weighted absolute mobility drop. Darker zones reflect higher impact reductions.

revealing how environmental stress and behavioral patterns intersect geographically across UAE regions.

C. Pollution Modelling

1) Predictive Modelling of NO_2 Concentrations: To examine how mobility patterns and urban characteristics affect air quality, supervised machine learning models were developed to predict NO_2 concentrations at the geohash level. The primary objective was to understand how variables such as mobility score, population density, and urban zone classification contribute to spatial and temporal pollution variation across the UAE. Two models were used for this task, a Linear Regression model, which served as a baseline for capturing general trends, and a Random Forest Regressor, selected for its ability to model nonlinear and complex interactions between features. Both models were trained using historical data from 2020 to

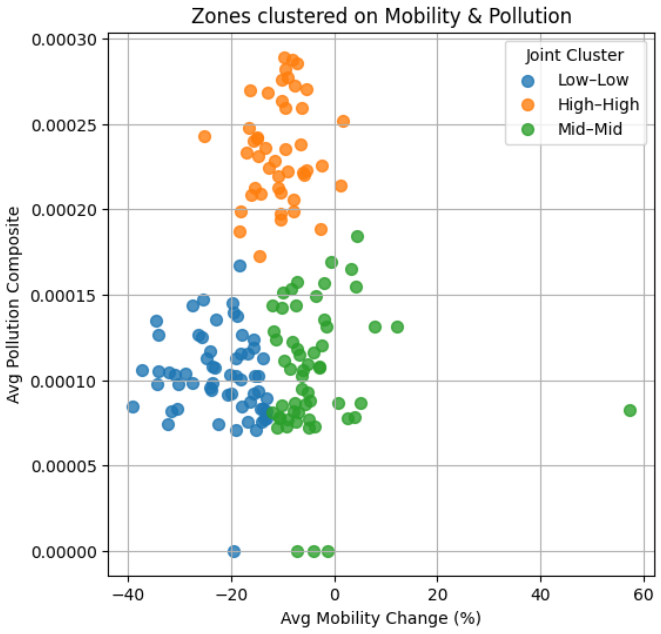

Fig. 11. Scatterplot of zones clustered by both mobility change and average pollution composite. Clusters represent joint behavioral-environmental profiles.

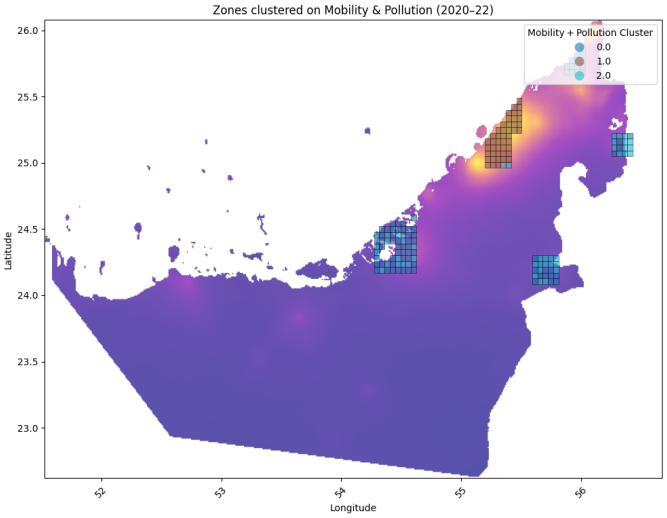

Fig. 12. Geographic distribution of joint mobility–pollution clusters (2020–2022) overlaid on composite pollution raster.

2022. Feature inputs included aggregated mobility scores from Google mobility reports, population density from geospatial datasets, and urban zone types labeled as industrial, residential, or mixed-use.

Model performance was evaluated using the R^2 score (coefficient of determination). The Linear Regression model captured overall trends but struggled in areas with irregular pollution behavior, likely due to its inability to model nonlinear effects. In contrast, the Random Forest model achieved R^2 scores above 0.7, demonstrating strong predictive accuracy and the capability to handle nonlinear relationships and variable interactions, especially important in urban environments influenced by both mobile and stationary emission sources. In addition to spatial prediction, time-series forecasting was conducted using a linear trend model to project average mobility and NO_2 levels through 2025. The forecast, shown in Figure 13, suggests that mobility will continue to rise post-pandemic, while NO_2 levels are expected to remain relatively stable if no significant interventions are made.

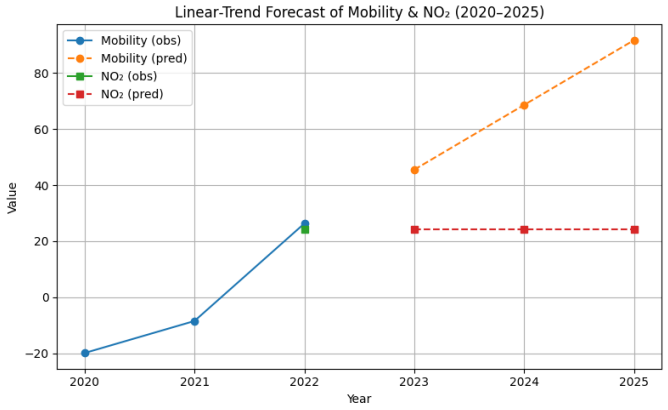

Fig. 13. Linear-trend forecast of average UAE mobility and NO_2 levels from 2020 to 2025. Solid lines represent observed values (2020–2022), while dashed lines show projections (2023–2025). Mobility is expected to increase, whereas NO_2 levels remain flat under a linear assumption.

2) Feature Importance and Interpretability:: To better understand which variables influenced NO_2 predictions, a feature importance analysis was performed using the trained Gradient Boosting Regressor (GBR). This method helped identify which input features contributed most to the model's performance across spatial zones. In the initial model, five features were used: mobility_score, pop_density, and their squared and interaction terms. As shown in Figure 14, population density alone accounted for over 80% of the predictive power. This dominance suggests that densely populated areas may retain more air pollutants due to reduced ventilation, infrastructure density, or lack of green spaces, regardless of temporary changes in human mobility.

To enhance generalizability and interpretability, a refined model was developed. This version included additional engineered features such as 'pop_density_log' in order to correct for skew, 'mobility_abs' to remove directionality bias, and 'mobility_density_ratio' to capture per-capita activity. As illustrated in Figure 15, the updated model distributed importance more evenly across variables. While 'pop_density_log' and 'pop_density' remained strong predictors, new engineered features like 'mobility_density_ratio' gained significance. This shift indicates that more nuanced spatial and behavioral patterns such as activity intensity relative to population which can provide deeper insight into pollution dynamics. Despite strong test performance with an R^2 of 0.686 and a very low mean absolute error (MAE $\approx 2.5 \times 10^{-5}$), the model's 5-fold cross-validation R^2 was negative (–0.459). This discrepancy highlights a potential overfitting issue and reflects challenges

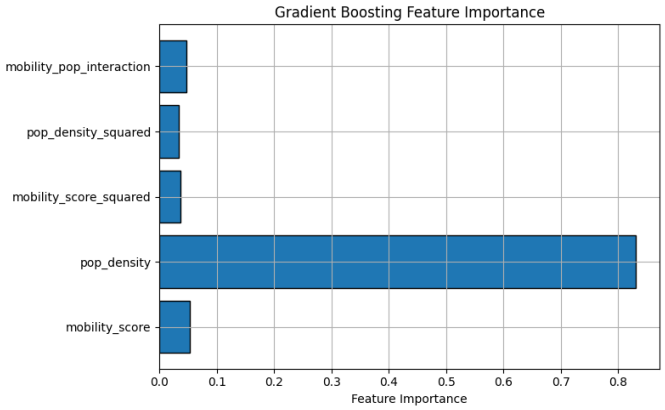

Fig. 14. Feature importance from the initial Gradient Boosting model. pop_density was the dominant predictor of NO_2, while other variables had minor influence.

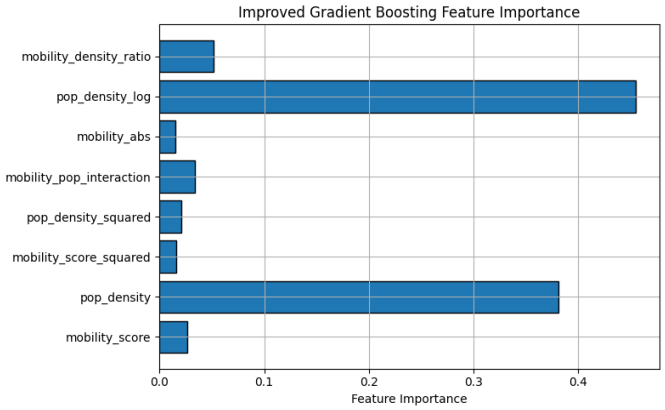

Fig. 15. Feature importance from the improved Gradient Boosting model. Log-transformed density and per-capita mobility enhanced interpretability and balance across features.

such as the small dataset size, very low target values (near machine precision), and high variability across geohash zones. These findings suggest that while the model performs well on the training data, its generalizability remains limited unless trained on a more diverse or larger dataset.

3) Implications for Urban Air Quality Management:: The modeling insights from Sections 4.C.1 and 4.C.2 offer actionable implications for environmental governance. Most notably, the feature importance analysis previously confirmed that population density is a dominant predictor of NO_2 concentrations, reinforcing the role of urban structure and land use in shaping pollution outcomes. To further explore the relationship between mobility and pollution over time, a joint analysis of workplace activity and NO_2 concentrations from 2020 to 2022 was conducted. As shown in Figure 16, seasonal trends indicate that sharp reductions in mobility, specifically during early 2020, were often accompanied by dips in NO_2, highlighting the sensitivity of urban air quality to behavioral and policy-driven shifts in activity.

Looking forward, as shown in Figure 17, Ericsson's urban mobility forecast through 2025 projects a continued rise in

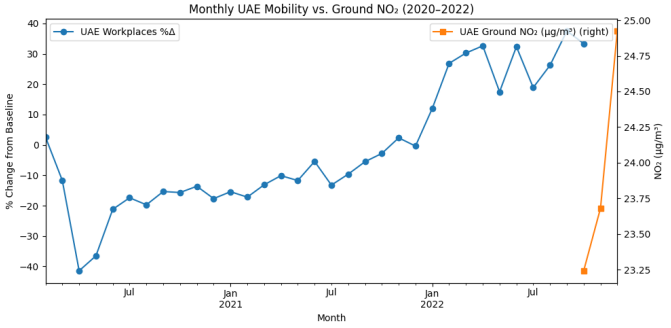

Fig. 16. Monthly averages of UAE workplace mobility vs. ground-level NO_2 concentration (2020–2022). While mobility dropped significantly during early COVID-19 periods, NO_2 levels followed similar seasonal dips, supporting traffic-pollution linkages.

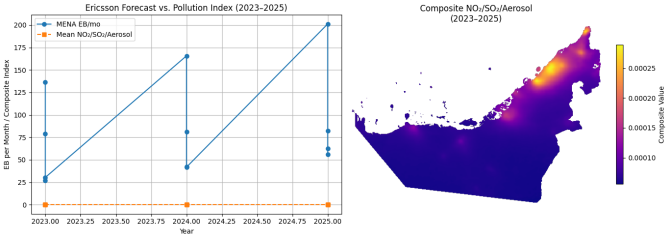

Fig. 17. Ericsson forecast of UAE mobility (EB/month) vs. projected pollution index (2023–2025). The right panel shows the composite pollution raster, where elevated NO_2/SO_2/Aerosol levels remain persistent in northern metro areas.

mobile device activity, indicative of broader growth in urban movement and energy use. In contrast, the spatial raster analysis for 2023–2025 suggests that pollution intensity may remain stagnant or increase in specific high-density corridors unless mitigated.

*4) **Recommendations for Sustainable Urban Air Management:**:* To address the varying pollution drivers across UAE urban zones, targeted interventions must be designed according to each zone's environmental–behavioral typology. Type C zones (characterized by high mobility and high pollution) should be prioritized for sustainable mobility solutions, including the implementation of green public transport systems, low-emission zones, and real-time air quality monitoring. In contrast, Type B zones (low mobility but high pollution) require stricter industrial controls, particularly around SO_2 emissions, and should be included in cleaner energy transition initiatives. Urban planning strategies that promote densification must also integrate air quality safeguards to prevent the concentration of emissions in compact areas. Moreover, the continued use of satellite-integrated raster data is essential to support scalable and automated monitoring of pollutants across diverse geographies. Overall, these insights highlight the importance of adopting multimodal, data-driven environmental policies. Effective air quality management must move beyond traffic regulation alone, addressing both mobile sources such as vehicles and stationary sources like industry and power generation.

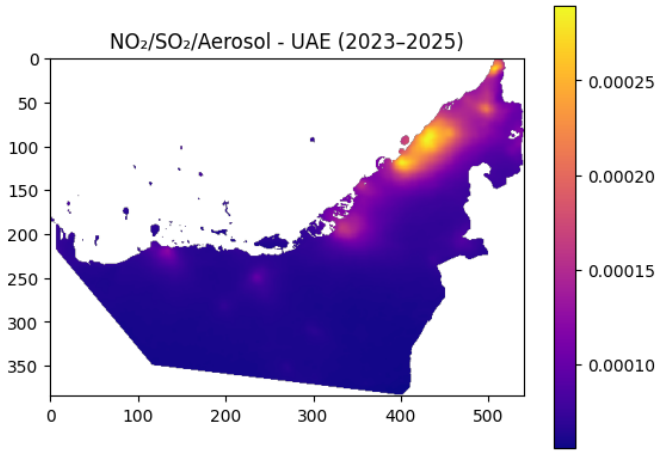

Fig. 18. Projected NO_2, SO_2, and aerosol index distribution in the UAE (2023–2025). Intensified pollution is expected in coastal urban centers, consistent with mobility and industrial growth forecasts.

D. Forecasting Phase Outlook (2023–2025)

1) Forecasting Urban Mobility and Pollution: To evaluate future pollution trends in light of projected mobility increases and urban growth, a forecasting analysis was conducted for the period 2023–2025 using spatially-resolved pollution raster data. This analysis builds upon the models and environmental baselines established in the 2020–2022 historical phase. Figure 18 presents the geospatial forecast of NO_2, SO_2, and aerosol index (AAI) concentrations across the UAE, based on satellite-derived emissions estimates and urban expansion models. The raster, extracted from UAE_NO2_SO2_Aerosol_Combined_2023_2025.tif, reveals continuity in pollution hotspots previously observed in Abu Dhabi, Dubai, and the Northern Emirates. To quantify future shifts in pollution intensity, a pixel-wise comparison was performed between raster outputs for the historical period (2020–2022) and the projected forecast period (2023–2025). The analysis revealed a mean absolute difference of 3.09×10^{-6} and a maximum pixel-level difference of 2.59×10^{-5}, with a total of 80,422 pixels showing detectable change. While the average change in pollutant concentration appears modest, the geographic spread of these changes is significant, particularly in zones expected to experience increases in human mobility. This spatial pattern supports the hypothesis that, in the absence of intervention, pollution burdens may accumulate incrementally across metropolitan areas, driven by behavioral recovery post-pandemic. The forecasting methodology integrates multiple components: Ericsson mobility projections through 2025, historical NO_2 regression models developed in Section 4.C, and the zone clustering framework from Section 4.B. By combining these elements, the study offers a predictive approach for identifying and monitoring emerging environmental risks at a high spatial resolution. This allows planners and policymakers to proactively respond to pollution shifts before they become embedded in urban air baselines.

2) Strategic Recommendations for Policy and Planning: The forecasts for pollution and mobility from 2023 to 2025 give important guidance for city planners and decision-makers. By combining future mobility trends with detailed pollution maps, the study helps identify specific urban zones that are at high risk and where actions can make the biggest difference. Many of these zones match the earlier-defined Type B (low mobility, high pollution) and Type C (high mobility, high pollution) areas, making them top priorities for future air quality efforts. Several actions are recommended. First, switching public transport to electric vehicles in busy areas can reduce harmful emissions like NO_2 and support climate goals. Second, creating green corridors such as adding trees and plants along roads, can help clean the air, cool down cities, and make walking safer and more pleasant. Third, there should be more air quality monitors on the ground, especially in areas that are growing quickly or where pollution levels are uncertain. These ground sensors should work together with satellite tools to provide detailed and timely pollution data. As shown in Figure 18, some parts of cities are expected to see worse air quality over time. If no action is taken, these places could move from being moderately polluted to having serious long-term problems. Looking ahead, cities could use real-time data and satellite maps to adjust environmental zones as conditions change. Also, better coordination between different government sectors such as transportation, energy, and environment, is needed to avoid policies that conflict with each other. Finally, using digital twin models (virtual simulations of cities) could help test how new buildings or roads might affect air quality before they're even built. These steps would shift cities from reacting to pollution after it happens to planning ahead using smart, data-based solutions, which is especially important for fast-growing countries like the UAE.

E. Forecast Scenario Analysis, 2023–2025)

1) Forecasted Pollution Patterns Across the UAE: The second phase of the study focused on simulating future air quality trends across the UAE by integrating satellite-based pollution forecasts with projected human mobility data. This integration was achieved through the concurrent use of Sentinel-5P forecast rasters for NO_2, SO_2, and aerosol indices, alongside Ericsson's forecasted mobile activity metrics for 2023–2025. The datasets utilized are summarized in Table III, highlighting the multi-source, multi-scale nature of the forecasting pipeline.

Spatial visualizations generated from these data layers reveal that major coastal urban centers—including Abu Dhabi, Dubai, and Sharjah—are expected to continue exhibiting elevated pollutant concentrations. These areas align with the highest projected increases in population density and digital mobility usage, suggesting a strong correlation between human activity intensity and environmental stressors. As illustrated in Figure 19, these urban cores are anticipated to remain critical risk zones without intervention.

Although the mean absolute difference (MAD) between historical and projected NO_2 values was relatively small

(3.09×10^{-6}), the number of geohash-level pixels exhibiting detectable change exceeded 80,000. This finding underscores the fact that modest average changes can still mask significant spatial variability—a key consideration in urban air quality management. From a policymaking perspective, this reinforces the need for high-resolution, location-specific monitoring to guide interventions at the neighborhood or district scale.

The forecast also captured peripheral and inland regions with emerging pollution risk, likely tied to expanding industrial zones and spillover effects from urban sprawl. These trends emphasize the need for dynamic planning strategies that go beyond reactive regulation and instead adopt proactive environmental zoning informed by predictive analytics.

Overall, the forecasted pollution patterns indicate that while average conditions may appear stable, the geographic spread of risk is growing more uneven. This spatial fragmentation makes the case for integrating forecasting tools into regulatory frameworks—particularly to anticipate future hotspots before thresholds are breached.

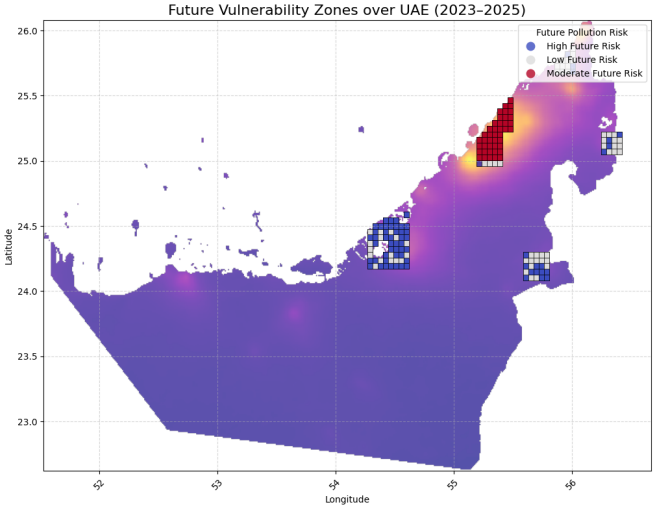

Fig. 19. Future Vulnerability Zones over UAE (2023–2025), highlighting areas of high projected pollution risk.

2) Prediction Model Accuracy and Residual Behavior: We applied a Gradient Boosting Regressor (GBR) to estimate future NO_2 concentrations based on forecasted urban mobility data and projected population trends. This ensemble learning method was selected for its capacity to model nonlinear interactions and its robustness against overfitting when hyperparameters are properly optimized. The model demonstrated outstanding predictive performance, achieving a test $R^2 = 0.959$ and a cross-validation $R^2 = 0.966$, suggesting that over 95% of the variance in NO_2 levels is accurately captured.

The alignment between predicted and sampled pollution values is visually confirmed in Figure 21, where points closely follow the 1:1 diagonal reference line. Furthermore, the model's low mean absolute error (1.0×10^{-5}) and root mean squared error (1.4×10^{-5}) are consistent with the narrow distribution of residuals shown in Figure 20. These results collectively indicate that the model is both precise and generalizable.

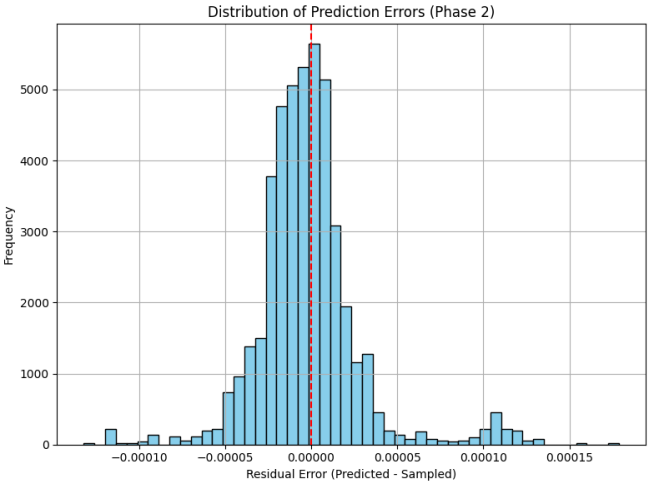

Fig. 20. Distribution of Prediction Errors (Phase 2). Mean error is close to zero, indicating low bias.

To further assess residual behavior, we plotted standardized residuals against sampled pollution levels in Figure 22. The distribution reveals no major outliers or heteroscedastic patterns, reinforcing the model's reliability across different pollution intensities. These diagnostics validate the GBR's suitability for urban air quality forecasting and suggest its potential utility in predictive planning systems. Table VI summarizes the key evaluation metrics supporting this assessment.

TABLE VI
PREDICTION ERROR SUMMARY (PHASE TWO MODEL)

Metric	Value
Test R^2	0.959
Cross-Validation R^2	0.966
Mean Absolute Error (MAE)	1.0×10^{-5}
Root Mean Squared Error (RMSE)	1.4×10^{-5}
Pixels with Detectable Change	80,422

3) Zones of Concern and Policy Implications: Beyond aggregate performance metrics and spatial trends, this phase also identified specific urban zones that are projected to exceed the World Health Organization's NO_2 guideline of 40 $\mu g/m^3$. These threshold-exceeding areas are especially important because they represent locations where health impacts and regulatory non-compliance are most imminent. Table VII lists the geohash-coded zones flagged for concern, based on forecasted NO_2 values derived from the model.

TABLE VII
URBAN ZONES EXCEEDING WHO NO_2 LIMIT (40 μG/M^3)

Geohash	NO_2 ($\mu g/m^3$)	Mobility Score (%)
thqem	40.40	-18.93
thqf4	76.80	+12.10
thqf8	45.15	-27.53

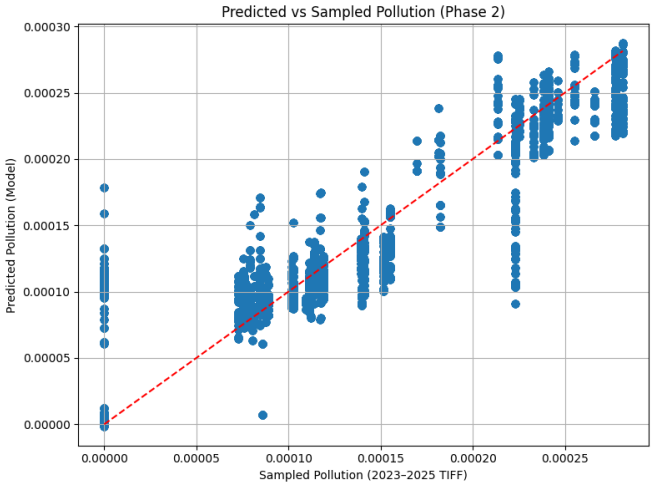

Fig. 21. Predicted vs. Sampled Pollution (Phase 2). Dashed red line indicates perfect alignment.

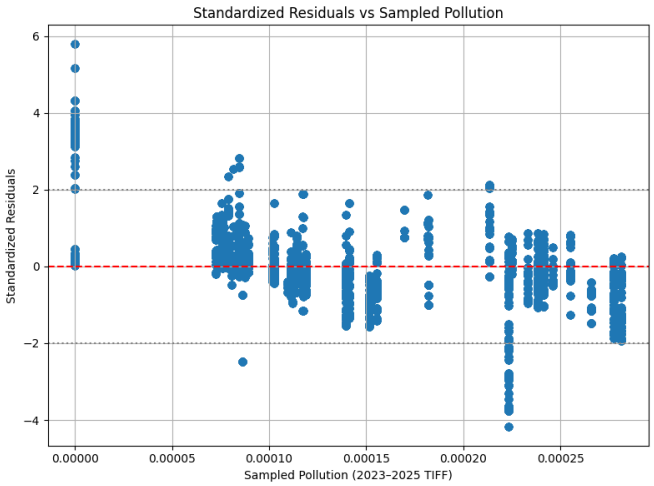

Fig. 22. Standardized Residuals vs. Sampled Pollution. Most values fall within ±2 standard deviations.

Among these, the geohash `thqf4` is of particular interest. With a projected NO_2 concentration of 76.80 $\mu g/m^3$ and a positive mobility change of +12.10%, this zone reflects a compounding effect of increasing human activity and pollutant concentration. The presence of both elevated NO_2 and growth in digital mobility usage suggests the convergence of traffic emissions, urban densification, and possibly fixed industrial sources.

By contrast, zones like `thqem` and `thqf8` show moderate to high pollution despite decreasing mobility, implying potential contributions from non-traffic sources such as power generation, construction, or long-range pollution transport. These cases illustrate that mobility data alone are not always sufficient to explain air quality changes and that multi-source analysis is critical for accurate attribution.

From a policy perspective, these spatial insights provide actionable intelligence. High-mobility, high-pollution zones (Type C) could benefit from sustainable transport infrastructure—such as low-emission zones, electrified public transit, or dynamic congestion pricing. Meanwhile, zones with persistent pollution but stagnant mobility (Type B) likely require fixed-source interventions, including stricter industrial permitting, improved stack emissions monitoring, and real-time pollutant surveillance.

V. Conclusion and Future Work

This study created a strong data-based system to understand how traffic and air pollution are connected in cities across the UAE. It used detailed satellite images such as (Sentinel-5P), local air quality sensors like (OpenAQ), and population-based movement data from Google. The work was implemented in two parts where one looked at the past (2020–2022) and the other predicted future trends (2023–2025). A special mapping method called 'geohashing' was used to group data by location. This was useful in helping compare patterns across different areas in a clear and detailed way. Also, this helped uncover both expected and unusual links between pollution and how people move around. In the historical analysis, the study confirmed that when people move around less, the NO_2 levels also drop which is supporting the idea that car emissions are a major cause of air pollution in cities. However, some exceptions stood out such as zone thqf4, where pollution stayed high even though there wasn't much change in how people moved. This shows that factories and other fixed sources can also cause a lot of pollution. Therefore, the environmental policies should not only focus on traffic but also consider these non-mobile sources. The prediction models showed that although the number of people move around affects pollution, population density has the strongest and most consistent impact over time and across different places. Many machine learning (ML) models like Random Forest and Gradient Boosting performed better than simple linear models because they could detect complex patterns in the data. The future predictions suggest that if things stay the same, pollution levels might remain high or even get worse, particularly in busy city areas or industrial zones. In addition, our framework used simulations based on made-up changes in movement, location shifts, and population factors in order to test how future environmental risks might play out. These tests showed that machine learning models are useful not only for studying past data, but also for helping cities plan better for the future.

Future Work

To enhance the precision, applicability, and impact of the proposed framework, several future directions are envisioned:

- **Integration of Meteorological and Topographical Variables:**
  Including factors like wind direction, temperature layers in the atmosphere, air pressure, and land elevation can make pollution models more accurate. These variables help explain why pollution levels such as PM and SO_2,

differ from one area to another, and how pollutants spread across different zones.

- **Tracking Multiple Pollutants:**
  Adding more pollutants like PM2.5, carbon monoxide (CO), and ozone (O_3) would give a fuller picture of environmental risks. These pollutants come from different sources and affect health in different ways, so studying them over time and across locations can help us better understand overall air quality.
- **Using Time-Based Deep Learning Models:**
  Models like LSTM or Temporal Convolutional Networks (TCNs) can better understand how pollution and movement change over time. These models are especially useful for spotting patterns related to events such as public holidays, unusual weather, or new regulations, helping predict how such events impact air quality.
- **Mapping Industrial Emissions with Zoning Data:**
  By collecting data on emissions from factories and combining them with zoning maps like residential, commercial, and industrial areas. We can better identify where pollution is coming from, and this would also help assess how well current land-use policies are working in managing environmental impacts.
- **Live Monitoring and Smart Decision Tools:**
  Creating a visual platform that combines real-time satellite images, sensor readings, and prediction results would support smarter city planning. This tool could help issue timely air quality warnings and test the impact of different policy decisions, making it a valuable asset for managing urban environments.
- **Modeling the Impact of Green Policies:**
  Future studies can test how different environmental policies such as traffic charges, switching to electric vehicles, or setting pollution limits might affect air quality and traffic over time. Running these scenarios would help decision-makers create smarter, data-backed policies.
- **Expanding to Neighbouring GCC Countries:**
  Applying this framework to nearby countries like Saudi Arabia, Qatar, and Bahrain could support regional planning by revealing pollution that crosses borders and highlighting shared environmental and infrastructure challenges.

Building on these directions, this study supports the shift toward smarter and more adaptive environmental management in urban areas. By merging technologies like satellite sensing, AI, and dynamic policy modelling, cities can move beyond reactive measures and adopt proactive, evidence-based planning. The UAE's unique combination of urban growth, environmental challenges, and investment in innovation makes it a key location for testing these forward-looking approaches.

REFERENCES

[1] H.-Y. Liu, E. Skjetne, and M. Kobernus, "Mobile phone tracking: in support of modelling traffic-related air pollution contribution to individual exposure and its implications for public health impact assessment," *Environmental Health*, vol. 12, no. 1, p. 93, Nov. 2013.

[2] M. Bauwens, S. Compernolle, T. Stavrakou, J.-F. Muller, J. van Gent, H. Eskes, P. F. Levelt, R. van der A, J. P. Veefkind, J. Vlietinck, H. Yu, and C. Zehner, "Impact of coronavirus outbreak on no_2 pollution assessed using tropomi and omi observations," *Geophysical Research Letters*, vol. 47, no. 11, Jun. 2020.

[3] F. Liu, I. M. Vrekoussis, X. Zeiher, M. Beirle, B. de Foy, J. Kuhlmann, J. D. Shan, M. Li, Y. Liu, and T. Wagner, "Abrupt decline in tropospheric nitrogen dioxide over china after the outbreak of covid-19," *Science Advances*, vol. 6, no. 28, p. eabc2992, Jul. 2020.

[4] X. Dong, W. Chen, J. Liu, and J. Zhao, "Population based air pollution exposure and its influence factors by integrating air dispersion modeling with gis spatial analysis," *Scientific Reports*, vol. 10, pp. 1–11, 2020.

[5] P. K. Chauhan, S. Tiwari, D. K. Gupta, A. Kumar, V. Pratap, and A. K. Singh, "Assessment of equivalent black carbon variations and its source apportionment over varanasi, indo-gangetic basin," *Atmospheric Pollution Research*, vol. 15, p. 102061, 2024.

[6] A. S, V. K. Patel, J. Kuttippurath, and H. Varikoden, "The covid-19 lockdown induced changes of so2 pollution in its human-made global hotspots," *Global Transitions*, vol. 6, pp. 152–163, 2024.

[7] J. Obiefuna, A. A. Oduwaye, A. C. Udofia, and K. C. Uzochukwu, "Geospatial assessment of ambient air quality footprints in relation to urban landuses in nigeria," *Environment and Ecology Research*, vol. 9, no. 6, pp. 312–327, 2021.

[8] T. Zhou, B. Liu, D. Zhang, and Q. Li, "Urban air pollutant mapping and tracing using multi-points in situ measurements combined with clustering and trajectory analysis," *Environmental Monitoring and Assessment*, vol. 197, no. 3, 2025.

[9] O. Ghaffarpasand, D. F. Archer, S. Choolun, H. Ngoga, and J. Y. Lee, "The impact of urban mobility on air pollution in kampala, an exemplar sub-saharan african city," *Atmospheric Pollution Research*, 2024.

[10] A. Brunelli, L. Calgaro, P. Scanferla, and E. Badetti, "Effects of pandemic on air pollution and sustainable mobility solutions," *Energies*, vol. 14, no. 17, p. 5729, 2021.

[11] S.-J. Mei, J.-T. Hu, D. Liu, F.-Y. Zhao, Y. Li, and H.-Q. Wang, "Airborne pollutant dilution inside the deep street canyons subjecting to thermal buoyancy driven flows: Effects of representative urban skylines," *Building and Environment*, vol. 149, pp. 592–606, 2019.

[12] A. Brunelli, S. Breda, P. Scanferla, L. Calgaro, A. Marcomini, and E. Badetti, "A methodology to assess a mobile urban street cleaning activity," *Atmospheric Pollution Research*, vol. 14, no. 2, p. 101680, 2023.

[13] S. Kephalopoulos, T. Paviotti, and M. Mylona, "Methods to improve traffic flow and noise exposure estimation in urban environments," *Environmental Pollution*, vol. 213, pp. 731–740, 2016.

[14] K. Singh, D. T. Grinstein, and M. M. Dey, "Monitoring air pollution in philadelphia using low-cost sensors and machine learning," *Frontiers in Built Environment*, vol. 7, p. 648620, 2021.

[15] S. Sadeghi, R. R. Rajkumar, and A. Rahimi, "Citizen-sensing and smart city frameworks for environmental data integration," *Future Internet*, vol. 15, no. 9, p. 263, 2023.

[16] J. Tang, H. Li, Y. Liu, and S. Wang, "Enhancing ozone nowcasting over east asia using a data-to-driven hybrid model," *Atmosphere*, vol. 15, no. 5, p. 699, 2024.

[17] A. A. Jawarneh *et al.*, "Ai-driven urban analytics: Data-driven models for mobility and pollution," *Future Internet*, 2023.

[18] O. Ghaffarpasand *et al.*, "The impact of urban mobility on air pollution in kampala," *Atmospheric Pollution Research*, 2024.

[19] G. Gianquintieri *et al.*, "Atmospheric satellite validation strategies," *Atmosphere*, 2024.

[20] Google, "Covid-19 community mobility reports," 2022, [Online]. Available: https://www.google.com/covid19/mobility/ [Accessed: 26-Apr-2025].

[21] OpenAQ, "Open air quality data platform," 2022, [Online]. Available: https://openaq.org/ [Accessed: 26-Apr-2025].

[22] Copernicus Open Access Hub, "Sentinel-5p satellite data," 2022, [Online]. Available: https://scihub.copernicus.eu/ [Accessed: 26-Apr-2025].

[23] M. Morley *et al.*, "Improving spatial exposure models via geocoding," *Environmental Pollution*, 2016.

[24] S. Oh *et al.*, "Enhancing air pollution forecasts using data-driven deweathering," *Atmospheric Pollution Research*, 2024.

[25] R. Mulyana *et al.*, "K-means clustering for source attribution," *Environmental Monitoring and Assessment*, 2025.

[26] A. Cummings *et al.*, "Urban mobility and mobile source emissions," *Frontiers in Built Environment*, 2021.

[27] R. Dong *et al.*, "Air pollution risk simulations using demographic weighting," *Scientific Reports*, 2020.
[28] Ericsson, "Ericsson mobility report – forecasts for mobile data traffic," 2023, [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report [Accessed: 26-Apr-2025].
[29] O. Obiefuna *et al.*, "Gis-based pollution mapping for urban management," *Environment and Ecology Research*, 2021.
[30] World Health Organization, "Air quality guidelines: Global update 2005," WHO Regional Office for Europe, 2006, available: https://www.who.int/publications/i/item/WHO-SDE-PHE-OEH-06.02 (accessed Apr. 30, 2025).