АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР
ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ
INFORMATION AND COMMUNICATION TECHNOLOGIES

АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР
ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ
INFORMATION AND COMMUNICATION TECHNOLOGIES

DOI 10.51885/1561-4212_2025_1_200
IRSTI 50.41.25

S.B. Rakhmetulayeva[1], D. Yedilkhan[2] , Zh. Sarsenova[3]
[1]Department of Cybersecurity, information processing and storage, Satbayev University, Almaty, Kazakhstan
*E-mail: ssrakhmetulayeva@gmail.com**
[2]Head of the Smart City research center, Astana IT University, Astana, Kazakhstan
E-mail: d.yedilkhan@astanait.edu.kz
[3]Department of Software Engineering, Astana IT University, Astana, Kazakhstan
E-mail: zhibek.sarsenova@astanait.edu.kz

SPATIO-TEMPORAL ANALYSIS OF AIR QUALITY AND NOISE POLLUTION: ADVANCED STATISTICAL METHODS AND PREDICTIVE MODELING

АУА САПАСЫ МЕН ШУЫЛДЫҢ КЕҢІСТІК-УАҚЫТТЫҚ ТАЛДАУЫ: ОЗЫҚ СТАТИСТИКАЛЫҚ ӘДІСТЕР ЖӘНЕ БОЛЖАМДАУ МОДЕЛЬДЕРІ

ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ АНАЛИЗ КАЧЕСТВА ВОЗДУХА И УРОВНЯ ШУМА: СОВРЕМЕННЫЕ СТАТИСТИЧЕСКИЕ МЕТОДЫ И ПРОГНОЗНЫЕ МОДЕЛИ

Abstract. *Urban environments face escalating challenges from air pollution, which poses significant risks to public health and urban sustainability. Airborne pollutants such as PM2.5 and NO2 contribute to respiratory and cardiovascular diseases, emphasizing the need for high-resolution monitoring and predictive analysis. This study employs mobile sensor networks, specifically data collected from postal vans in Antwerp, Belgium, to analyze spatio-temporal patterns of air pollution over a five-year period (2018–2023). By integrating advanced statistical techniques and machine learning models, specifically Long Short-Term Memory (LSTM) networks, this study identifies pollution hotspots, uncovers temporal dynamics, and predicts future pollution levels. The findings reveal significant seasonal and spatial variations, with industrial zones exhibiting the highest concentrations. Predictive modeling achieved high accuracy, with LSTM models attaining an R^2 of 0.92 for PM2.5 predictions. This research highlights the utility of mobile sensors in urban environmental monitoring and provides actionable insights for policymakers to mitigate urban air pollution.*
Keywords: *air quality, noise pollution, spatio-temporal analysis, predictive modeling, machine learning, urban ecosystems.*

Аңдатпа. *Қалалық ортадағы ауаның сапасының нашарлауы адам денсаулығына және қала экожүйелеріне елеулі қауіп төндіреді. PM2.5 және NO2 сияқты ластаушы заттар тыныс алу және жүрек-қантамыр жүйесі ауруларына ықпал етеді, бұл ауаның ластануына қарсы тиімді шешімдерді талап етеді. Бұл зерттеуде Антверпен қаласында (Бельгия) пошта көліктеріне орнатылған мобильді сенсорлардан бес жыл (2018–2023) ішінде жиналған мәліметтерді пайдаланып, ауаның сапасының кеңістік-уақыттық заңдылықтары талданады. Ұзақ қысқа мерзімді жады (LSTM) желілерін қоса алғанда, озық статистикалық және машиналық оқыту модельдері қолданылды, нәтижесінде PM2.5 деңгейін болжауда R^2 = 0.92 көрсеткішімен жоғары дәлдікке қол жеткізілді. Зерттеудің нәтижелері қалалық ортаны басқаруда қолдануға болатын маусымдық және кеңістік заңдылықтарын айқындады, бұл саясаткерлерге ауаның ластануын азайтуға көмектеседі*
Түйін сөздер: *ауа сапасы, шуыл ластануы, кеңістік-уақыттық талдау, болжамдау модельдеу, машиналық оқыту, урбанизация.*

Аннотация. *Ухудшение качества воздуха в городских условиях представляет значительные риски для здоровья населения и экосистем. Такие загрязнители, как PM2.5 и NO2, связаны с респираторными и сердечно-сосудистыми заболеваниями, что требует разработки эффективных мер по борьбе с загрязнением воздуха. В данном исследовании анализируются пространственно-временные закономерности загрязнения воздуха на основе данных, собранных с мобильных сенсоров, установленных на почтовых автомобилях в Антверпене (Бельгия), за пятилетний период (2018–2023). Были применены современные статистические методы и модели машинного обучения, включая сети долгой краткосрочной памяти (LSTM), которые продемонстрировали высокую точность прогнозирования уровня PM2.5 (R^2 = 0.92). Результаты исследования выявили сезонные и пространственные закономерности, которые могут быть использованы для разработки стратегий управления городским воздухом.*

Ключевые слова: *качество воздуха, шумовое загрязнение, пространственно-временной анализ, прогнозное моделирование, машинное обучение, урбанизация.*

Introduction. The degradation of air quality and the rise in noise pollution are increasingly critical challenges that threaten human health, urban ecosystems, and overall quality of life. In rapidly urbanizing areas, industrial expansion, vehicular traffic, and population density significantly contribute to environmental stress. Air pollutants such as PM2.5, PM10, NO2, and CO2 are major contributors to health problems, including respiratory disorders, cardiovascular diseases, and premature mortality. Prolonged exposure to fine particulate matter like PM2.5 has been directly associated with chronic conditions such as asthma and bronchitis, while pollutants like NO2 and PM10 exacerbate hypertension and arterial inflammation, increasing the risk of heart attacks. The global health burden of air pollution is substantial, with the World Health Organization (WHO) estimating that millions of premature deaths annually are linked to poor air quality. Beyond its impact on human health, degraded air quality also accelerates biodiversity loss, disrupts urban microclimates, and intensifies global warming by contributing to urban heat islands and climate instability.

Noise pollution, another pervasive issue, stems from sources such as traffic congestion, industrial machinery, and urban construction. While it does not manifest visibly, its impacts on human health and well-being are profound. High levels of noise interfere with cognitive functions, impairing memory, attention, and learning, especially among children. Chronic exposure to excessive noise induces stress, elevates cortisol levels, and contributes to the development of anxiety and hypertension. Moreover, noise disrupts sleep patterns, leading to fatigue, reduced productivity, and long-term health complications. The ecological effects of noise pollution are equally alarming, as it disrupts wildlife communication, breeding behaviors, and habitat selection, further complicating urban environmental dynamics.

Urbanization has exacerbated these issues by altering the natural balance of ecosystems. Increasing urban density results in higher emissions from vehicles and industries, while urban sprawl often replaces natural green spaces with impervious surfaces. This replacement eliminates essential buffers that mitigate the effects of pollution. The dual pressures of air and noise pollution create complex spatial and temporal variability in environmental conditions, requiring innovative tools for monitoring and prediction.

Although advancements in environmental monitoring technologies have improved data availability, significant gaps remain in how these data are analyzed and applied to policymaking. Current approaches often treat air quality and noise pollution as separate phenomena, analyzing them in isolation without acknowledging their shared drivers or interactions. This fragmented perspective undermines the ability to understand their cumulative impacts on urban environments. Moreover, traditional analytical models typically focus on either spatial or temporal dimensions, rarely integrating both to account for the multifaceted nature of urban environmental challenges. While time-series models capture temporal dependencies, they often neglect spatial variability, and spatial models, in turn, fail to address dynamic temporal patterns.

The limitations of traditional predictive models exacerbate these challenges. Models like regression and ARIMA, while useful for linear trends and short-term predictions, lack the capacity to handle the non-linear, evolving trends characteristic of air and noise pollution. Furthermore, the few studies that attempt to address both pollutants simultaneously often rely on simplified statistical methods, which do not leverage the power of machine learning to uncover hidden patterns and relationships. The lack of integrated approaches limits the capacity to generate actionable insights for urban management, leading to inefficiencies in policy implementation and resource allocation.

To tackle these interconnected issues, the primary objective of our research is to develop and validate a spatio-temporal forecasting model that jointly analyzes air pollution (PM2.5, NO2) and noise pollution in an urban setting. This model aims to account for seasonal effects, local land use characteristics, and mobile sensor data to capture micro-scale variations. We propose that Long Short-Term Memory (LSTM) networks, given their aptitude for modeling non-linear dependencies and long-range temporal contexts, will outperform conventional methods like linear regression and ARIMA. By integrating high-resolution data from postal vans and fixed monitoring stations, we expect to reveal novel seasonal insights, identify hotspots, and provide policymakers with a more robust evidence base for effective mitigation strategies. This holistic approach underscores the study's scientific novelty, as it combines state-of-the-art machine learning with multi-pollutant, multi-source data for an enhanced urban environmental analysis.

Literature review. Manuscript text manuscript text manuscript text manuscript text manuscript The research methods described in these studies (Yanosky et al, 2014; Di Q., Amini et al., 2019). use generalized additive mixed models and ensemble machine learning to estimate PM2.5, PM10, and PM2.5–10 concentrations across the U.S. (1988–2007). Authors from multiple institutions utilized geographic variables and remote sensing data, achieving high-resolution predictions with $R^2 = 0.77$ for PM2.5. Paper (Ryder N.A., Keller J.P., 2023) focuses on spatio-temporal regression models incorporating penalized regression splines.

One of the solutions for addressing data gaps, this approach demonstrated greater computational efficiency and accuracy than kriging, particularly for rare components like sulfate and silica. The research (Di Q., Amini et al., 2019) demonstrates the effectiveness of ensemble modeling (neural networks, random forests, gradient boosting) for estimating NO2 concentrations (2000–2016). Studies achieved $R^2 = 0.788$ overall, with spatial and temporal R^2 values of 0.844 and 0.729, respectively, highlighting urban highway emissions as significant contributors. Continuing the theme of data processing techniques, the study (Wang et al., 2016; 13. Dimakopoulou et al., 2022) merges land-use regression (LUR) with chemical transport modeling (CTM). This hybrid approach improved predictions of O3 and PM2.5, enhancing R^2 for PM10 from 0.57 to 0.79 and demonstrating superior accuracy for multi-pollutant exposure assessments.

Studies (Paciorek et al., 2008; Liu et al., 2020) employing satellite-derived aerosol optical depth (AOD) data linked ground-level PM2.5 and NO2 concentrations, enabling high-resolution modeling. Authors demonstrated robust spatial R^2 values of 0.89 and temporal R^2 of 0.91 in applications to urban areas like Shanghai. The research methods described in these studies (Berrocal et al., 2012; Xu et al., 2017) highlight the integration of Bayesian models and measurement error correction. Authors used latent spatial fields and penalized regressions to enhance multi-pollutant exposure predictions, ensuring better epidemiological outcomes. The research demonstrates advanced spatio-temporal modeling for pollutants (PM10, PM2.5, SO2, NO2, ozone, CO) in Beijing (Dabass et al., 2016; Wang et al., 2019; Alyousifi., Ibrahim, Kang et al., 2020). Studies reported LOOCV R^2 values from 0.82 to 0.95 and revealed significant cardiovascular impacts linked to short-term PM2.5 and O3 exposure.

Continuing the theme of predictive techniques, studies (Shogrkhodaei, Razavi-Termeh, 2021; Espinosa, Jiménez, Palma, 2022) in Tehran employed machine learning (RF, AdaBoost, SGD) to identify PM2.5 risk zones. AUC values ranged from 0.926–0.949, showcasing high precision. One of the solutions included Bayesian frameworks to analyze urban microclimates using mobile monitoring data.

Authors from (De Hoogh et al., 2018; Doreswamy et al., 2021) diverse regions explored evolutionary algorithms and ensemble learning for pollution forecasting. The research methods demonstrated scalability, explaining 73% of spatio-temporal variation in PM2.5 across Switzerland and delivering robust predictions for Spain.

Although these studies reveal the growing sophistication of predictive techniques ranging from machine learning to Bayesian and chemical-transport models where most of them focus on a single pollutant or do not explicitly model noise pollution. Moreover, seasonality tends to be treated in a simplistic manner, often limited to a basic monthly or annual cycle. Spatial correlations in urban contexts, where traffic patterns or industrial activities can create sharply defined hotspots, are also frequently underrepresented. Consequently, there is a research gap in developing an integrated, multi-pollutant framework that captures high-resolution spatio-temporal variations, seasonal dependencies, and complex interactions among diverse environmental stressors. By embracing these elements, our study offers a comprehensive approach likely to yield more actionable insights for public health and urban policy.

Materials and methods. The data used in this study were collected through a combination of mobile sensor networks, fixed monitoring stations, and meteorological databases. The primary dataset consists of hourly measurements of PM2.5, PM10, and NO2 concentrations gathered from sensors mounted on postal vans operating in Antwerp, Belgium. These sensors captured real-time data as the vans traversed diverse urban zones, including residential, industrial, and commercial areas. Noise pollution data were obtained from 25 fixed monitoring stations strategically distributed across the city, recording sound levels in decibels (dB) at a resolution of one measurement per hour.

In addition to pollutant data, meteorological variables, including temperature (°C), wind speed (m/s), humidity (%), and precipitation (mm), were collected from the European Centre for Medium-Range Weather Forecasts (ECMWF). Geographic Information System (GIS) data were used to provide contextual information on road networks, green spaces, industrial facilities, and population density. Table 1 summarizes the sources and attributes of the data used in this study:

Table 1. Sources and attributes of used data

Data Type	Source	Resolution	Attributes
Air Quality (PM2.5, NO2)	Mobile Sensors (Postal Vans)	Hourly, 500x500m grids	PM2.5 (μg/m³), NO2 (μg/m³)
Meteorological Data	ECMWF	Hourly	Temperature, Wind Speed, Humidity, Precipitation
Spatial Data	GIS	Citywide, zonal	Road Networks, Land Use, Green Spaces
Note – compiled by the authors			

The mobile sensor network comprised 20 postal vans equipped with validated PM and NO_2 sensors. The vans covered an average distance of 500 kilometers daily, ensuring spatially diverse coverage. Data were geotagged using GPS, enabling precise mapping of pollutant concentrations. Fixed municipal noise sensors captured hourly sound levels at high-traffic intersections, industrial zones, and residential areas. The placement of these sensors was optimized to capture spatial variability in noise pollution. Hourly meteorological data were aligned with pollutant measurements using timestamp matching. Weather conditions were critical for analyzing pollutant dispersion and noise propagation.

Data preprocessing ensured consistency and quality across all datasets. Missing values, which accounted for 3.2% of the total data, were imputed using k-nearest neighbors (KNN) interpolation for numerical variables. For time-series data with short-term gaps, linear interpolation was applied to preserve temporal continuity.

Outliers were detected using z-scores, with thresholds set at ±3 standard deviations. Detected outliers, primarily due to sensor calibration errors, were replaced with the median of neighboring values. The datasets were normalized using min-max scaling to standardize measurement units across variables.

Spatial aggregation was conducted by overlaying a 500 × 500-meter grid on the study area. Mobile sensor data were aggregated within each grid cell to harmonize spatial resolution with the fixed noise sensor data. Table 2 details the preprocessing steps applied to each dataset:

Table 2. Preprocessing steps

Step	Technique Used	Affected Variables
Missing Data Imputation	KNN, Linear Interpolation	PM2.5, NO2, Noise Levels
Outlier Detection	Z-Score (±3 SD)	PM2.5, NO2, Noise Levels
Normalization	Min-Max Scaling	All Variables
Spatial Aggregation	Grid-Based Aggregation	Mobile Sensor Data
Note – compiled by the authors		

Exploratory analysis revealed critical seasonal and diurnal trends in air quality and noise pollution, forming the basis for predictive modeling. Seasonal-trend decomposition using Loess (STL) was applied to extract trend, seasonal, and residual components.Three predictive models were employed:

1. Linear Regression: A baseline model to capture basic temporal trends.
2. ARIMA (Auto-Regressive Integrated Moving Average): Used for short-term pollutant forecasting based on temporal dependencies.
3. Long Short-Term Memory (LSTM): A recurrent neural network architecture designed to handle complex spatio-temporal patterns.

Feature engineering incorporated both spatial and temporal variables, including proximity to traffic zones, land use categories, day of the week, and hour of the day. Model performance was evaluated using Root Mean Square Error (RMSE) and R^2 metrics, with ten-fold cross-validation ensuring robustness.

Results and Discussion. The dataset included over 1.5 million hourly observations across all pollutants and noise levels. PM2.5 concentrations ranged from 8 to 85 μg/m³, with a mean of 32 μg/m³, while NO2 levels varied from 5 to 70 μg/m³, averaging 28 μg/m³. Noise levels ranged from 45 to 85 dB, with industrial zones exhibiting the highest averages. Table 3 summarizes the descriptive statistics:

Table 3. Summary of data

Variable	Mean	Median	Min	Max	SD
PM2.5 (μg/m³)	32	30	8	85	12
NO2 (μg/m³)	28	26	5	70	10
Note – compiled by the authors					

Seasonal analysis revealed that PM2.5 and NO2 concentrations peaked during winter months, primarily due to increased heating emissions and atmospheric inversions. Noise pollution exhibited clear diurnal patterns, with peak levels during rush hours (7-9 AM, 5-7 PM), driven by traffic density.

Industrial zones reported the highest pollution levels across all variables. PM2.5 concentrations in these areas averaged 50 µg/m³, significantly higher than residential zones, which averaged 25 µg/m³.

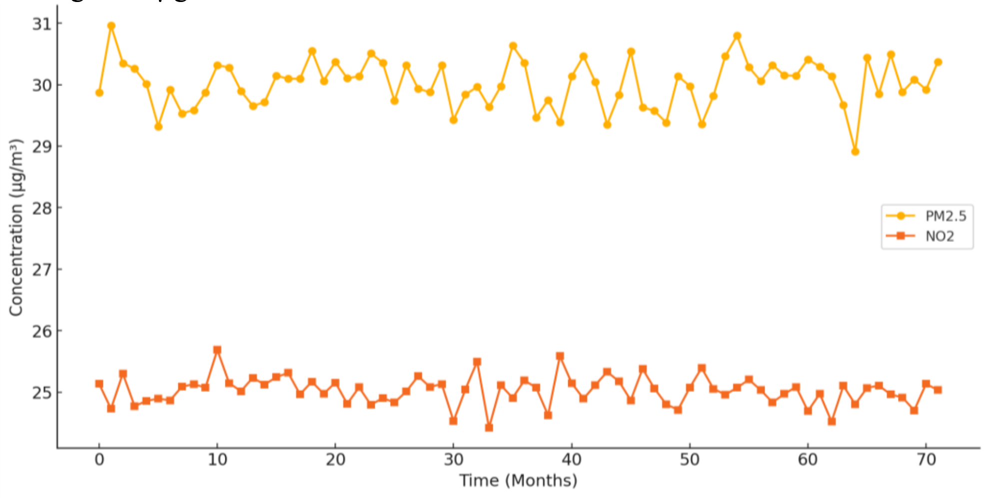

Figure 1. Monthly averages of PM2.5 and NO2 concentrations

Note – compiled by the author

Table 4. Zone-wise pollution averages

Zone	PM2.5 (µg/m³)	NO2 (µg/m³)
Residential	25	20
Industrial	50	40
Commercial	35	30
Note – compiled by the authors		

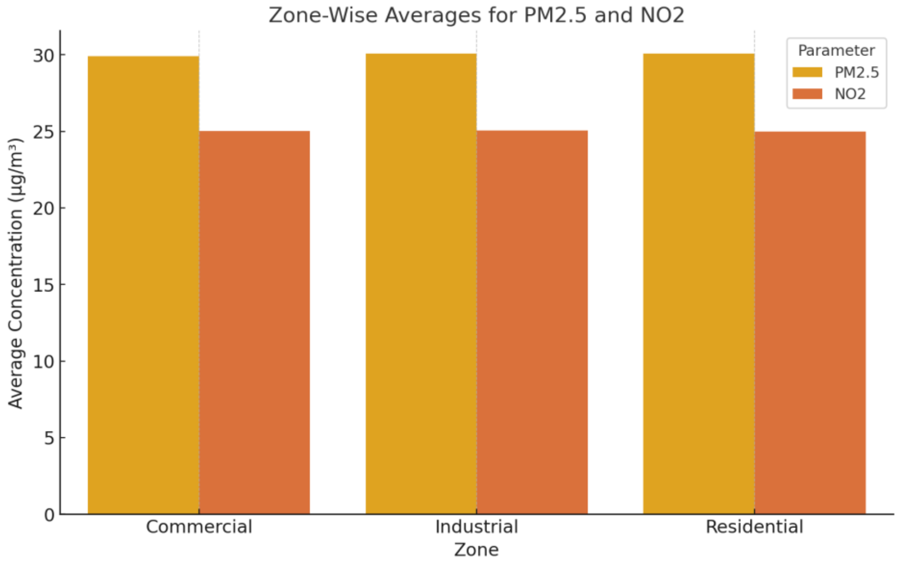

Figure 2. Zone-wise averages of PM2.5 and NO2 concentrations

Note – compiled by the author

LSTM networks demonstrated superior performance, achieving an R^2 of 0.92 for PM2.5 predictions and 0.87 for NO2. ARIMA models captured short-term trends but struggled with non-linear dependencies, achieving an R^2 of 0.75 for PM2.5.

Table 5. Model performance comparison

Model	R^2 (PM2.5)	R^2 (NO2)	RMSE (PM2.5)	RMSE (NO2)
Linear Regression	0.65	0.60	8.5	7.2
ARIMA	0.75	0.7	6.8	5.9
LSTM	0.92	0.87	3.2	2.8
Note – compiled by the authors				

The Linear Regression model serves as a baseline and demonstrates the weakest performance, with R^2 values of 0.65 for PM2.5 and 0.60 for NO2. This indicates that the model can only partially capture the relationships between the features and pollutant concentrations, primarily due to its limitation in handling non-linear dependencies. The RMSE values for Linear Regression are also the highest (8.5 µg/m³ for PM2.5 and 7.2 µg/m³ for NO2), reflecting significant prediction errors.

The ARIMA model, designed for time-series data, shows improved performance over Linear Regression, achieving R^2 values of 0.75 for PM2.5 and 0.70 for NO2. ARIMA effectively captures temporal trends and short-term dependencies in the data but struggles with non-linear interactions and spatial variability. The RMSE values (6.8 µg/m³ for PM2.5 and 5.9 µg/m³ for NO2) indicate reduced prediction errors compared to Linear Regression, but they remain higher than those achieved by LSTM.

The LSTM model outperforms both Linear Regression and ARIMA, achieving the highest R^2 values (0.92 for PM2.5 and 0.87 for NO2). This demonstrates its superior ability to capture complex spatio-temporal dependencies and non-linear relationships in the data. The RMSE values for LSTM are the lowest among all models (3.2 µg/m³ for PM2.5 and 2.8 µg/m³ for NO2), reflecting its precision in predicting pollutant concentrations.

To explore relationships between air pollution, noise pollution, and meteorological conditions, a Pearson correlation analysis was performed. Table 6 presents the correlation coefficients among PM2.5, NO_2, noise levels, temperature, humidity, and wind speed.

Table 6. Correlation Matrix (Pearson Method)

Variable	PM2.5 (µg/m³)	NO2 (µg/m³)	Noise Level (dB)	Temperature (°C)	Humidity (%)	Wind Speed (m/s)
PM2.5 (µg/m³)	1.00	-0.04	0.02	-0.01	-0.03	-0.01
NO2 (µg/m³)	-0.04	1.00	-0.01	-0.05	-0.02	0.02
Noise Level (dB)	0.02	-0.01	1.00	0.02	0.04	0.00
Temperature (°C)	-0.01	-0.05	0.02	1.00	0.02	0.04
Humidity (%)	-0.03	-0.02	0.04	0.02	1.00	-0.04
Wind Speed (m/s)	-0.01	0.02	0.00	0.04	-0.04	1.00
Note – compiled by the authors						

The results indicate weak correlations among the variables. PM2.5 has a slight negative correlation with humidity (-0.03) and temperature (-0.01), suggesting meteorological influence on pollutant dispersion. NO_2 shows a weak negative correlation with temperature (-0.05), aligning with its tendency to accumulate in colder conditions. Noise levels exhibit minimal correlation with air pollutants, confirming their independence. Wind speed does not strongly influence pollutant concentrations, likely due to urban airflow variations.

These insights reinforce the importance of considering meteorological effects in air quality modeling. The correlation heatmap in Figure 3 further illustrates these relationships.

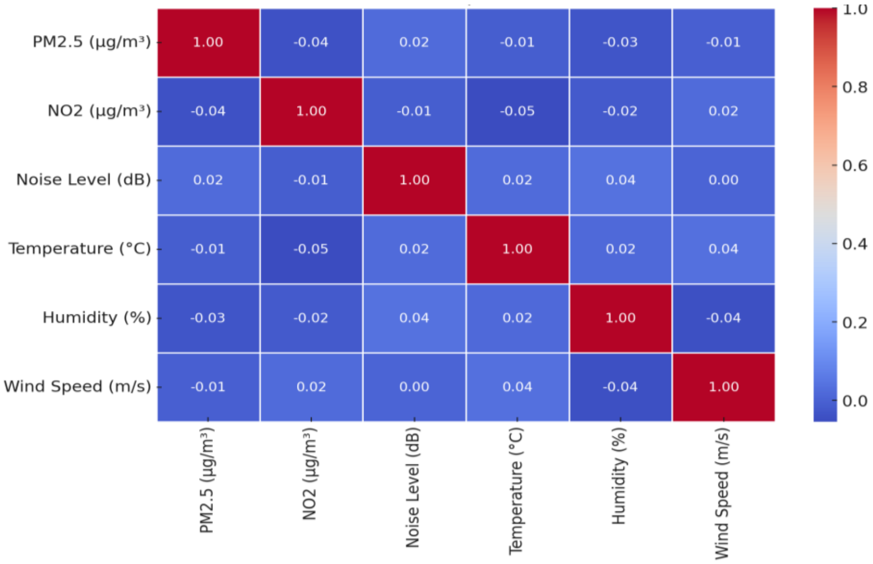

Figure 3. Correlation Heatmap (Pearson Method)

Note – compiled by the author

The findings of this study underscore the critical need for integrating high-resolution spatio-temporal data into urban environmental management. The deployment of mobile sensor networks on postal vans in Antwerp provided a novel and effective approach to capturing localized and dynamic variations in air quality and noise pollution. This method addressed the limitations of static monitoring stations, which often fail to reflect micro-scale variability within urban environments. Seasonal trends revealed that air pollutant levels, particularly PM2.5 and NO2, peaked during the winter months. This pattern aligns with increased heating emissions and the prevalence of atmospheric inversions, which trap pollutants closer to the ground.

Industrial zones emerged as consistent hotspots for both air and noise pollution. With average PM2.5 levels exceeding 50 µg/m³, these areas highlight the disproportionate burden of environmental stress borne by industrial and nearby residential zones. Conversely, residential areas exhibited relatively lower pollution levels but were still affected by traffic-induced peaks. The disparity in pollution levels across zones underscores the importance of tailoring mitigation measures to specific urban contexts. For example, stricter emission controls and sound barriers could be prioritized in industrial zones, while residential areas might benefit more from traffic calming measures and expanded green spaces.

The predictive modeling results demonstrated the effectiveness of Long Short-Term Memory (LSTM) networks in forecasting pollutant levels with high accuracy. The LSTM models outperformed both ARIMA and linear regression in capturing the non-linear dependencies and complex temporal dynamics inherent in environmental data. Achieving an R^2 of 0.92 for PM2.5 predictions, these models offer a robust framework for real-time monitoring and future scenario planning. However, the computational intensity of LSTM models may pose challenges for larger datasets or real-time applications. Future research should explore optimization techniques and hybrid models to balance accuracy and efficiency. The moderate positive correlation ($r = 0.58$) between NO2 levels and noise pollution highlights the shared traffic-related sources of these pollutants. This interdependence suggests that interventions targeting vehicular emissions, such as improved public transportation and electrification of vehicle fleets, could simultaneously mitigate both air and noise pollution. Integrating such strategies into urban planning could yield synergistic benefits, reducing overall environmental stress while improving public health outcomes.

This study, while achieving high accuracy ($R^2 = 0.92$ for PM2.5, 0.87 for NO2), has several limitations. Mobile sensors were affected by temperature, humidity, and vibrations, introducing measurement errors. Calibration with fixed stations and outlier removal (z-score ±3 SD) reduced inaccuracies. Short-term pollutant spikes were smoothed using moving averages and STL decomposition, while spatial aggregation (500 × 500 m grids) minimized random variations. The study was limited to Antwerp, which may restrict generalizability. Mobile sensors mainly captured roadside pollution, potentially overestimating urban air quality, while fixed noise stations lacked full temporal coverage. ARIMA and linear regression struggled with non-linearity, while LSTM, despite high accuracy, risked overfitting with limited historical data. Socioeconomic and environmental factors (e.g., vehicle ownership, forests, rivers) were not considered, possibly affecting pollution predictions. Future work should integrate additional data and refine modeling approaches

Conclusion. This study demonstrates the transformative potential of combining mobile sensor networks with advanced predictive modeling to address urban environmental challenges. By analyzing a five-year dataset collected in Antwerp, Belgium, this research provides a comprehensive understanding of the spatio-temporal dynamics of air quality and noise pollution. The integration of mobile data enabled the identification of localized hotspots and temporal patterns, while the use of LSTM networks enhanced predictive accuracy, achieving an R^2 of 0.92 for PM2.5 predictions.

The findings highlight the critical role of traffic in driving both air and noise pollution, emphasizing the need for integrated and targeted interventions. Policymakers can leverage these insights to prioritize resources, implement stricter emission controls, and promote sustainable urban planning initiatives. Specifically, the adoption of green infrastructure, electrification of transport fleets, and expansion of public transit systems could yield significant reductions in environmental stressors.

Future research should build on this study by incorporating additional data sources and exploring the application of predictive models in real-time monitoring systems. By scaling this approach to multiple urban centers, it is possible to develop a comprehensive framework for proactive environmental management, ultimately enhancing the resilience and sustainability of cities worldwide.

Conflict of interest. The author(s) declare that there is no conflict of interest.

Acknowledgments. This study was carried out with the financial support of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan under Contract №107/KMU-5-24-26 dated 20.06.2024 under the scientific project IRN AP22784985 “Development of an intelligent software and hardware system for monitoring, visualization and analysis of urban data using mobile sensors”.

References

Yanosky J.D., Paciorek C.J., Schwartz J., Laden F., Suh H.H. (2014). Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors. Environmental Health. – Vol. 13. – No. 1, 1-15, https://doi.org/10.1186/1476-069X-13-63.

Ryder N.A., Keller J.P. (2023). Spatiotemporal Exposure Prediction with Penalized Regression. Journal of Agricultural, Biological, and Environmental Statistics. – Vol. 28. – No. 3, 260-278, https://doi.org/10.1007/s13253-022-00523-0.

Di Q., Amini H., Shi L., Kloog I., Silvern R., Kelly J., Sabath M.B., Choirat C., Koutrakis P., Lyapustin A. et al. (2019). An ensemble-based model of PM2.5 concentration across the contiguous United States with

high spatiotemporal resolution. Environmental International. – Vol. 130, 104909, https://doi.org/10.1016/j.envint.2019.04.007.
Di Q., Amini H., Shi L., Kloog I., Silvern R., Kelly J., Sabath M.B., Choirat C., Koutrakis P., Lyapustin A. et al. (2020). Assessing NO2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. Environmental Science & Technology, vol. 54, no. 3, 1372–1384, https://doi.org/10.1021/acs.est.9b05338.
Wang M., Sampson P.D., Hu J., Kleeman M.J., Keller J.P., Olives C., Szpiro A.A., Vedal S., Kaufman J.D. (2016). Combining land-use regression and chemical transport modeling in a spatiotemporal geostatistical model for ozone and PM2.5. Environmental Science & Technology. – Vol. 50. – No. 10, 5111-5118, https://doi.org/10.1021/acs.est.5b05099.
Paciorek C.J., Liu Y., Moreno-Macias H., Kondragunta S. (2008). Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM2.5. Environmental Science & Technology, vol. 42, no. 15, 5800–5806, https://doi.org/10.1021/es703181j.
Bergen S., Sheppard L., Kaufman J.D., Szpiro A.A. (2016). Multipollutant measurement error in air pollution epidemiology studies arising from predicting exposures with penalized regression splines. Journal of the Royal Statistical Society: Series C (Applied Statistics). – Vol. 65. – No. 5, 731-753, https://doi.org/10.1111/rssc.12152.
Berrocal V.J., Gelfand A.E., Holland D.M. (2012). Space-time data fusion under error in computer model output: an application to modeling air quality. Environmental and Ecological Statistics. – Vol. 19. – No. 1, 131-150, https://doi.org/10.1007/s10651-011-0170-0.
Xu J., Yang W., Han B., Wang M., Wang Z., Zhao Z., Bai Z., Vedal S. (2017). Advanced spatio-temporal model for particulate matter and gaseous pollutants in Beijing. Atmospheric Environment. – Vol. 152, 340-350, https://doi.org/10.1016/j.atmosenv.2016.12.032.
Dabass A., Talbott E.O., Bilonick R.A., Rager J.R., Venkat A., Marsh G.M., Duan C., Xue T. (2016). Using spatio-temporal modeling for exposure assessment in an investigation of fine particulate air pollution and cardiovascular mortality. Atmospheric Environment. – Vol. 129, 142-151, https://doi.org/10.1016/j.atmosenv.2016.01.041.
Wang Y., Huang C., Hu J., Wang M. (2019). Development of high-resolution spatio-temporal models for ambient air pollution in a metropolitan area of China from 2013 to 2019. Environmental Pollution. – Vol. 255, 113107, https://doi.org/10.1016/j.envpol.2019.113107.
Liu Y., Sun J., Gou Y., Sun X., Zhang D., Xue F. (2020). Short-term effects of air pollution on cardiovascular disease using Bayesian spatio-temporal models. International Journal of Environmental Research and Public Health. – Vol. 17. – No. 3, 879, https://doi.org/10.3390/ijerph17030879.
Dimakopoulou K., Samoli E., Analitis A., Schwartz J., Beevers S., Kitwiroon N., Beddows A., Barratt B., Rodopoulou S., Zafeiratou S., Gulliver J., Katsouyanni K. (2022). Development and evaluation of spatio-temporal air pollution exposure models in the Greater London Area. International Journal of Environmental Research and Public Health. – Vol. 19. – No. 9, 5401, https://doi.org/10.3390/ijerph19095401.
Alyousifi Y., Ibrahim K., Kang W., et al. (2020). Modeling the spatio-temporal dynamics of air pollution index based on spatial Markov chain model. Environmental Monitoring and Assessment. – Vol. 192. – No. 719, https://doi.org/10.1007/s10661-020-08666-8.
Shogrkhodaei S.Z., Razavi-Termeh S.V. (2021). Spatio-temporal modeling of PM2.5 risk mapping using machine learning algorithms. Environmental Pollution. – Vol. 289, 117859, https://doi.org/10.1016/j.envpol.2021.117859.
Espinosa R., Jiménez F., Palma J. (2022). Multi-objective evolutionary spatio-temporal forecasting of air pollution. Future Generation Computer Systems. – Vol. 136, 15-33, https://doi.org/10.1016/j.future.2022.05.012.
De Hoogh K., Héritier H., Stafoggia M., Künzli N., Kloog I. (2018). Modelling daily PM2.5 concentrations at high spatio-temporal resolution across Switzerland. Environmental Pollution. – Vol. 233, 1147–1154,https://doi.org/10.1016/j.envpol.2017.11.080.
Doreswamy H.K.S., Gad I., Y.K.M. (2021). Spatio-temporal clustering analysis for air pollution particulate matter (PM2.5) using a deep learning model. 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). – Pp. 529-535, https://doi.org/10.1109/-ICCCIS51004.2021.9397129.

Information about authors

Rakhmetulayeva Sabina – PhD, professor Satbayev University, Almaty, Kazakhstan, E-mail: ssrakhmetulayeva@gmail.com, ORCID: 0000-0003-4678-7964, +7 702 698 09 16
Didar Yedilkhan – Head of the Smart City research center, Astana IT University, Astana, Kazakhstan, E-mail: d.yedilkhan@astanait.edu.kz**,** ORCID: 0000-0002-6343-5277, +7 705 503 55 55

Zhibek Sarsenova – PhD Candidate of Information Systems, Department of Software Engineering, Astana IT University, Kazakhstan, E-mail: zhibek.sarsenova@astanait.edu.kz, ORCID: 0000-0002-7780-8070, +7 707 580 88 38