


ORIGINAL PAPER

Open Access



Evaluating the impact of urban traffic patterns on air pollution emissions in Dublin: a regression model using google project air view data and traffic data

Pavlos Tafidis^{1,2*} , Mehdi Gholamnia², Payam Sajadi², Sruthi Krishnan Vijayakrishnan² and Francesco Pilla²

Abstract

Air pollution is a significant and pressing environmental and public health concern in urban areas, primarily driven by road transport. By gaining a deeper understanding of how traffic dynamics influence air pollution, policymakers and experts can design targeted interventions to tackle these critical issues. In order to analyse this relationship, a series of regression algorithms were developed utilizing the Google Project Air View (GPAV) and Dublin City's SCATS data, taking into account various spatiotemporal characteristics such as distance and weather. The analysis showed that Gaussian Process Regression (GPR) mostly outperformed Support Vector Regression (SVR) for air quality prediction, emphasizing its suitability and the importance of considering spatial variability in modelling. The model describes the data best for particulate matter (PM_{2.5}) emissions, with R-squared (R²) values ranging from 0.40 to 0.55 at specific distances from the centre of the study area based on the GPR model. The visualization of pollutant concentrations in the study area also revealed an association with the distance between intersections. While the anticipated direct correlation between vehicular traffic and air pollution was not as pronounced, it underscores the complexity of urban emissions and the multitude of factors influencing air quality. This revelation highlights the need for a multifaceted approach to policymaking, ensuring that interventions address a broader spectrum of emission sources beyond just traffic. This study advances the current knowledge on the dynamic relationship between urban traffic and air pollution, and its findings could provide theoretical support for traffic planning and traffic control applicable to urban centres globally.

Keywords Air pollution, Google project air view, Regression model, Traffic congestion

1 Introduction

Air pollution stands as one of the most pressing environmental challenges of the twenty-first century, posing significant threats to human health and the well-being

of ecosystems around the world. This issue is not merely a local or regional concern; it has far-reaching global implications that demand international cooperation and innovative solutions. In urban areas, where population density, industrial activities, and transportation converge, air pollution tends to be more concentrated and its impact more severe [27]. Cities are often hotspots for emissions of pollutants such as nitrogen dioxide (NO₂), particulate matter (PM_{2.5}), and carbon dioxide (CO₂) [2]. As a result, urban populations are particularly vulnerable to the adverse effects of polluted

*Correspondence:

Pavlos Tafidis
pavlos.tafidis@ed.ac.uk

¹ School of Engineering, The University of Edinburgh, King's Buildings, Edinburgh EH9 3FB, Scotland, UK

² School of Architecture, Planning and Environmental Policy, University College Dublin, Richview Campus, Belfield, Dublin D04 V1W8, Ireland



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

air, including respiratory illnesses, cardiovascular diseases, and reduced quality of life [9].

Dublin, the capital and largest city of Ireland in terms of population, is not exempt from these challenges. Firstly, it is consistently ranked as one of the most congested cities in the world, as reported by TomTom in 2022 [24]. Furthermore, according to a recent report from the European Environment Agency (EEA) in 2022, the city's air quality standards are below acceptable levels [5]. More particularly, air quality monitoring results from 2021 showed that $PM_{2.5}$ mainly from burning solid fuel in our homes, and NO_2 mainly from road transport, remain the main threats to good air quality. EPA monitoring shows that $PM_{2.5}$ and NO_2 levels are within the current European Union (EU) legal limits, however these pollutants exceed the World Health Organisation (WHO) Air Quality guidelines (AQGs) for health.

The correlation between traffic and air pollution is well-established in the literature, and the impact of different variables such as road incline, vehicle type, fuel, etc., has been analysed [1, 6–8, 12–14, 26]. However, detailed insights into this relationship within the specific context of Dublin remain limited. On separate occasions, Tang et al. [21–23] assessed the impact of different traffic management strategies on air quality and public health in Dublin, based on modelled traffic conditions from 2013. Findings revealed the influence of various strategies on air emissions and traffic, highlighting that urban transport systems are dynamic ecosystems. Changes in one location can have effects that extend beyond just the immediate area. Furthermore, Quintyne et al. [18] examined the potential impact of COVID-19 transport restrictions in Ireland, particularly emphasizing their repercussions on air quality and respiratory hospital admissions. During the COVID-19 pandemic, Ireland implemented several transport restrictions, including a significant reduction in public transportation services, restrictions on non-essential travel, and stringent lockdown measures that limited movement to within a certain distance from homes. Results showed that during the period of transport restrictions, there was a reduction in the annual mean NO_2 , and a decrease in hospital admissions for respiratory system diseases. Despite the valuable insights these studies provide, a gap exists in our understanding of the complex dynamics between urban traffic patterns and air pollution emissions in Dublin, particularly through a data-driven lens. A data-driven approach is crucial as it will allow for empirical analysis based on real-world data, ensuring more accurate and context-specific insights [20] into the relationship between traffic patterns and air pollution. Access to high-resolution datasets offers the opportunity to dive deeper

into this relationship, potentially revealing patterns and correlations previously unseen.

This study seeks to address the aforementioned gap by implementing a data-driven approach that combines state-of-the-art machine learning (ML) algorithms with high-quality datasets. Utilizing the Google Project Air View (GPAV) [3] dataset and local traffic data [4], we delve into the dynamics between Dublin's traffic patterns and its air pollution levels. Over 50 million street-level air quality data measurements were released by Dublin City Council (DCC) and Google as part of GPAV. The mapped, street-by-street air quality data is a first for an Irish city and provides unique street level insights that will help inform current and future environmental and climate policies, and planning efforts. The findings are intended to guide policymakers and urban planners in creating effective strategies to improve air quality in Dublin.

The remainder of the paper is structured as follows: Section 2 provides a detailed overview of the datasets used, the methods employed, and the machine learning algorithms implemented. Section 3 presents the results of our analysis, highlighting key findings and the relationships discovered between urban traffic patterns and air pollution levels in Dublin. Section 4 offers a discussion on the implications of these results, their significance in the broader context of urban planning and policymaking, and the potential applications for other urban areas. Finally, Section 5 concludes the study with a summary of the main points, recommendations for future interventions, and suggestions for further research in this domain.

2 Methods

The following section describes the methods and steps adopted for the purpose of this study (Fig. 1). To assess the relationship between urban traffic patterns and air pollution emissions in Dublin, two primary very high-resolution data sources were used: 1) GPAV dataset, which provides detailed spatiotemporal resolution air quality measurements (1 s intervals), including levels of Nitric Oxide (NO), Ozone (O_3), Carbon Monoxide (CO), NO_2 , $PM_{2.5}$, and CO_2 at street level; 2) Traffic counts dataset, collected from the DCC's Sydney Coordinated Adaptive Traffic System (SCATS), providing information on hourly vehicle counts at various locations throughout the city. The GPAV dataset was made available through collaboration with Dublin City Council and Google. This data is publicly accessible at Google Air View Data–Dublin City. The traffic count data from SCATS is also publicly available at Traffic Counts Datasets. The availability of these datasets ensures that the results can be replicated by other researchers. Finding correlations between

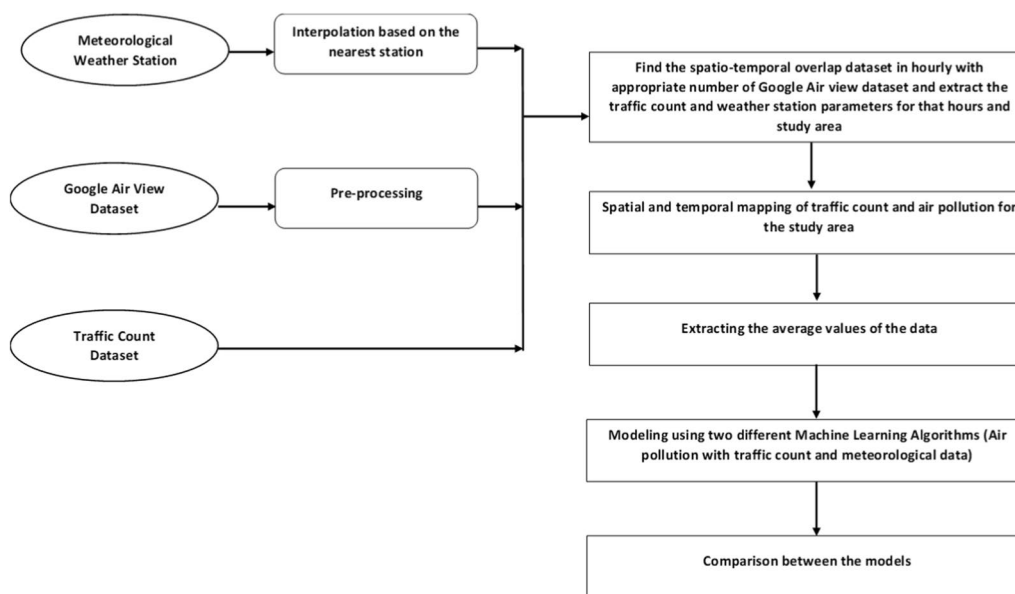


Fig. 1 Methodological flow diagram

traffic counts and the mentioned pollutants is important for several reasons such as public health, environmental impact, policy and planning and public awareness. While O_3 is not directly emitted from transportation sources, traffic emissions are the primary sources of its main precursors: NO_x and volatile organic compounds (VOCs). Therefore, it is imperative to investigate the relationship between traffic conditions and O_3 pollution. Additionally, the Meteorological Weather Station Dataset including temperature, wind speed, wind direction, pressure, and rainfall, publicly available from Met Éireann, the Irish Meteorological Service, was leveraged [16]. The data was interpolated to align with the nearest weather station, ensuring relevant and localized meteorological inputs were incorporated into the analysis. Finally, the datasets were synchronized based on spatiotemporal overlap. This ensured that each hourly air quality measurement from the GPAV dataset corresponded with traffic counts and meteorological data. An initial exploratory data analysis was conducted to visualize and understand the underlying patterns and relationships in the data.

The case study encompasses areas proximal to the city centre of Dublin, including parts of both the inner and outer city sectors (Fig. 2). The study area includes 29 intersections, defined as the points where two or more roads cross or meet, which are depicted with yellow stars in the figure. However, traffic data was available for only 19 of them. The time period concerned 12 months from May 2021 to April 2022. The time period selected for this analysis was determined by the availability of the GPAV data. This timeframe includes critical seasonal variations

and the residual impacts of COVID-19 restrictions, which are essential for understanding the dynamics of urban traffic patterns and air pollution in Dublin. These factors contribute to the observed trends and provide a valuable context for interpreting the results. Analysing data over a one-year period is important for several reasons, especially in the context of studying phenomena with temporal variations, such as traffic and air quality. Many factors such as seasonal variability, anomalies, peak and off-peak patterns are not neglected allowing robust results, a deeper understanding of the monitored phenomena, and improved policy formulation.

Besides estimating the correlation between air pollution and traffic using Pearson's r , we utilized two distinct machine learning algorithms to better examine the relationship between traffic patterns, meteorological factors, and air pollution levels:

- Support Vector Regression (SVR): A deterministic regression algorithm renowned for its computational efficiency [19]. SVR was employed due to its capacity to handle vast datasets efficiently, drawing focus on support vectors to predict air pollution levels based on traffic volume and meteorological insights.
- Gaussian Process Regression (GPR): A probabilistic regression algorithm, GPR was chosen for its ability to provide intrinsic uncertainty estimates, allowing for a more nuanced understanding of the predictions [10]. Though more computationally intensive than SVR, its adaptability in modelling intricate data patterns offers a robust approach for this study.

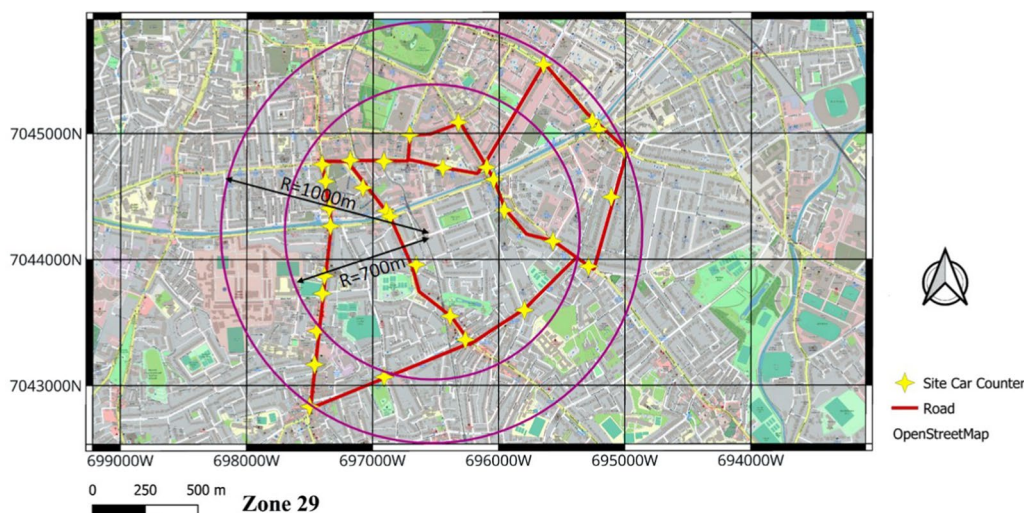


Fig. 2 Study area with buffer zones (UTM zone 29)

Each model was trained using traffic volume data from SCATS, meteorological data from Met Éireann, and air quality data from the GPAV dataset (70% training–30% testing). Performance was assessed using Root Mean Square Error (RMSE), and R-squared (R^2) score. The R^2 value indicates how well the model's predictions match the actual data. It ranges from 0 to 1, where a higher value signifies a better fit. An R^2 value closer to 1 means that the model explains a larger portion of the variance in the data. In environmental modelling, R^2 values around 0.5 or higher are generally considered acceptable, though this can vary based on the complexity and variability of the data. RMSE measures the average magnitude of the prediction errors, providing insight into the model's accuracy. Lower RMSE values indicate better predictive performance, with values closer to zero representing minimal differences between predicted and observed values. The acceptable range for RMSE depends on the specific context and units of the data being analysed. The probabilistic nature of GPR also allowed for an exploration into uncertainty estimates, providing a layer of depth to the analysis.

Finally, a head-to-head comparison between the SVR and GPR models was conducted. This comparative analysis aims to underline the strengths and weaknesses of each algorithm, granting researchers, urban planners, and policymakers' clear insights into the complex dynamics of traffic volume, meteorological factors, and air quality in Dublin.

3 Results

In the following section, the findings derived from the analysis are given. Figure 3 displays the distribution of air pollution levels in Dublin for various pollutants during the studied period. The results discussed relate to the entire range of values within each box plot, where the boxes represent the interquartile range (IQR) of the data. The red line within each box plot indicates the median value, providing a measure of the central tendency. As it can be seen, there is an observable peak in NO_2 during specific months, i.e., September to December. The median values during these months also indicate elevated NO_2 levels, suggesting increased vehicular activity or other seasonal factors like heating during these months. Additionally, the spread of values within these months is considerable, pointing to variability in daily NO_2 levels which could be influenced by factors such as traffic volume fluctuations and weather conditions. The months of January and February 2022 also show relatively high NO_2 levels, although not as pronounced as the September to December period. NO levels appear steadier, with only minor month-to-month variations, while the median values are relatively consistent, indicating a more stable emission source. The O_3 concentrations present a distinct pattern, with a noticeable rise during spring months, i.e., March, April, May. The median values during these months are also higher. Examining the CO and CO_2 emissions, it can be observed that both concentrations display spikes in particular months as well, e.g., December. Notably, while the average/median values for CO

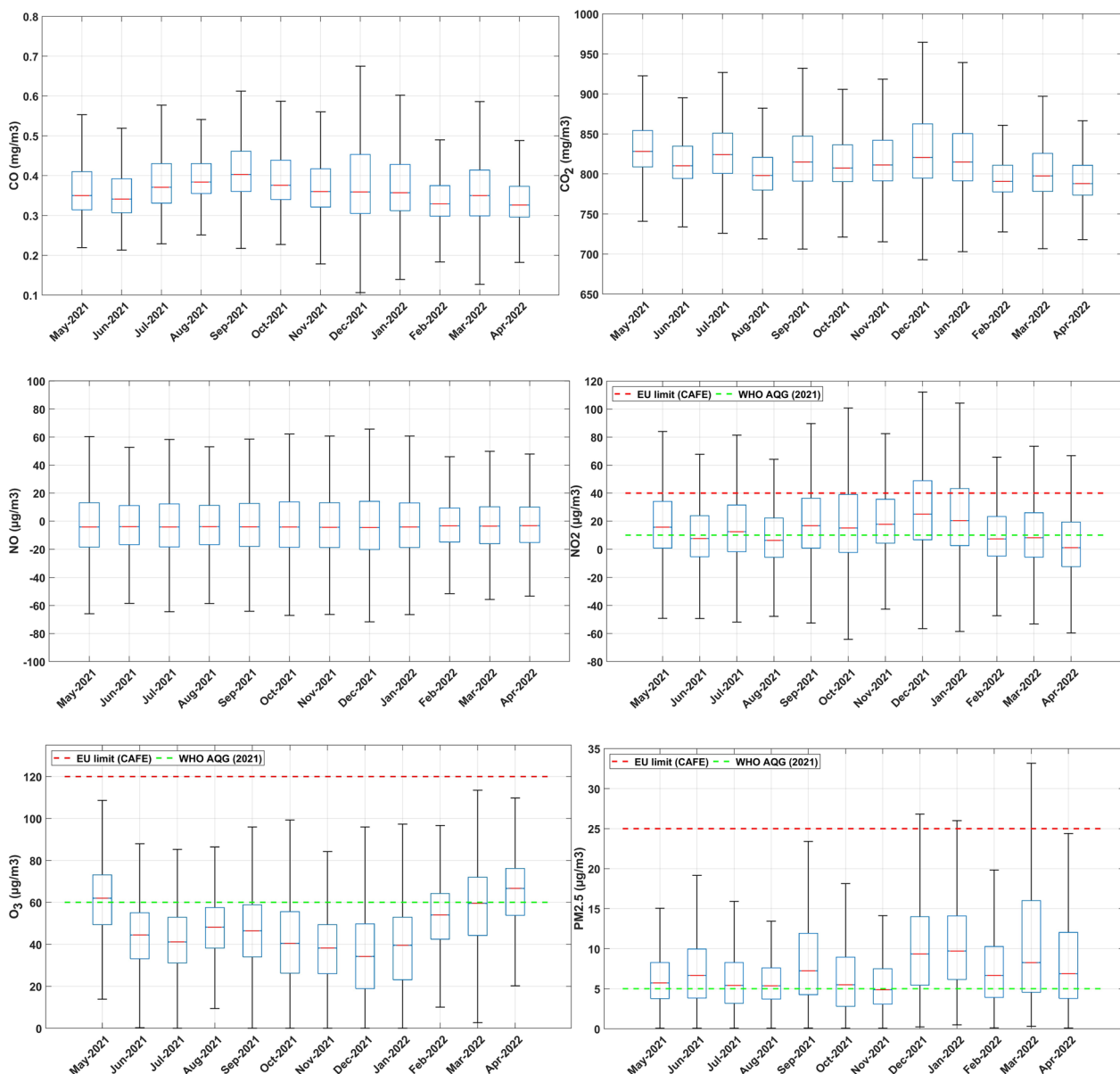


Fig. 3 Monthly distribution of pollutant emissions in Dublin

and CO₂ emissions in September and May, respectively, are higher than those recorded for December, the peak values observed in December highlight significant short-term increases likely influenced by specific events or conditions during that month. Lastly, the PM_{2.5} levels suggest a pronounced concentration during specific intervals, e.g., December to January, and March. The median values during these intervals are also elevated, likely due to increased heating activities and lower atmospheric dispersion rates during the winter months. It is important to note that while this study did not include an analysis of the relationship between these pollutant concentrations

and weather conditions, factors such as wind speed, temperature inversions, and other meteorological conditions can significantly influence air pollution levels. When analysing air quality metrics, we observe distinct differences between regional and global standards. The PM_{2.5} concentration, for instance, is consistently within the EU threshold of 25 µg/m³ during the observed timeframe [5]. Yet, it surpasses the more stringent WHO AQG set at 5 µg/m³ [17]. When examining the air quality pollutants indicative of road transport emissions, particularly NO₂, we find that its concentration repeatedly exceeds the WHO AQG recommendation of 10 µg/m³ over much

of the investigation period. Nonetheless, it adheres to the EU permissible level of $40 \mu\text{g}/\text{m}^3$, mirroring the pattern noticed with $\text{PM}_{2.5}$. Further, with regard to CO, the established thresholds as delineated by the EU and WHO AQG are $10 \mu\text{g}/\text{m}^3$ and $4 \mu\text{g}/\text{m}^3$ respectively. The gathered data suggests that CO levels in the region consistently align with these benchmarks. O_3 , while not a direct emission, is pivotal to monitor due to its standing as a significant anthropogenic greenhouse gas. The data elucidates that its levels are in accordance with the EU's prescribed norms ($120 \mu\text{g}/\text{m}^3$). However, during the summer months of June through August, the concentration surpasses the WHO's stricter threshold of $60 \mu\text{g}/\text{m}^3$. It's important to underscore that the reference points utilized in this study were based on average annual metrics.

Figure 4 provides an overview of the monthly distribution of traffic counts in the study area. Throughout this period, a notable consistency emerges in the median traffic counts, underscoring a stable flow of traffic for the majority of the year. This consistency is further mirrored in the middle 50% of the data, especially evident during months like June and July 2021. However, there is discernible variability in the extreme values. For instance, December 2021 witnesses a significant decrease in the minimum traffic count, possibly suggesting seasonal effects or an external influencing event during this

period. During this time, Ireland experienced new public health measures due to COVID-19, including restrictions on hospitality, reduced capacity for indoor events, and advice to limit social contacts, which likely contributed to the observed decrease in traffic counts. Conversely, an upward trend in maximum traffic counts becomes apparent as we transition from the latter part of 2021 to early 2022, with months such as February and March 2022 showcasing heightened traffic on their busiest days. Despite these fluctuations in extremes, the Interquartile Range remains largely unwavering across months, implying that the core traffic activity, or the central majority of observations, maintained its steadiness. The overall spread of the data and the position of the median within each monthly box also hint at a slight skew in traffic counts for certain months.

Upon cross-referencing Figs. 3 and 4, potential patterns and observations that should be highlighted are: a) Concentration Peaks and Traffic: The notable peak in NO_2 from September to December could potentially coincide with factors such as increased vehicular activity or heating during colder months. However, the traffic data doesn't show a corresponding surge in this period, suggesting other factors might be influencing NO_2 concentrations; b) Decreased Traffic, Increased Emissions: An intriguing pattern emerges in December 2021 where the

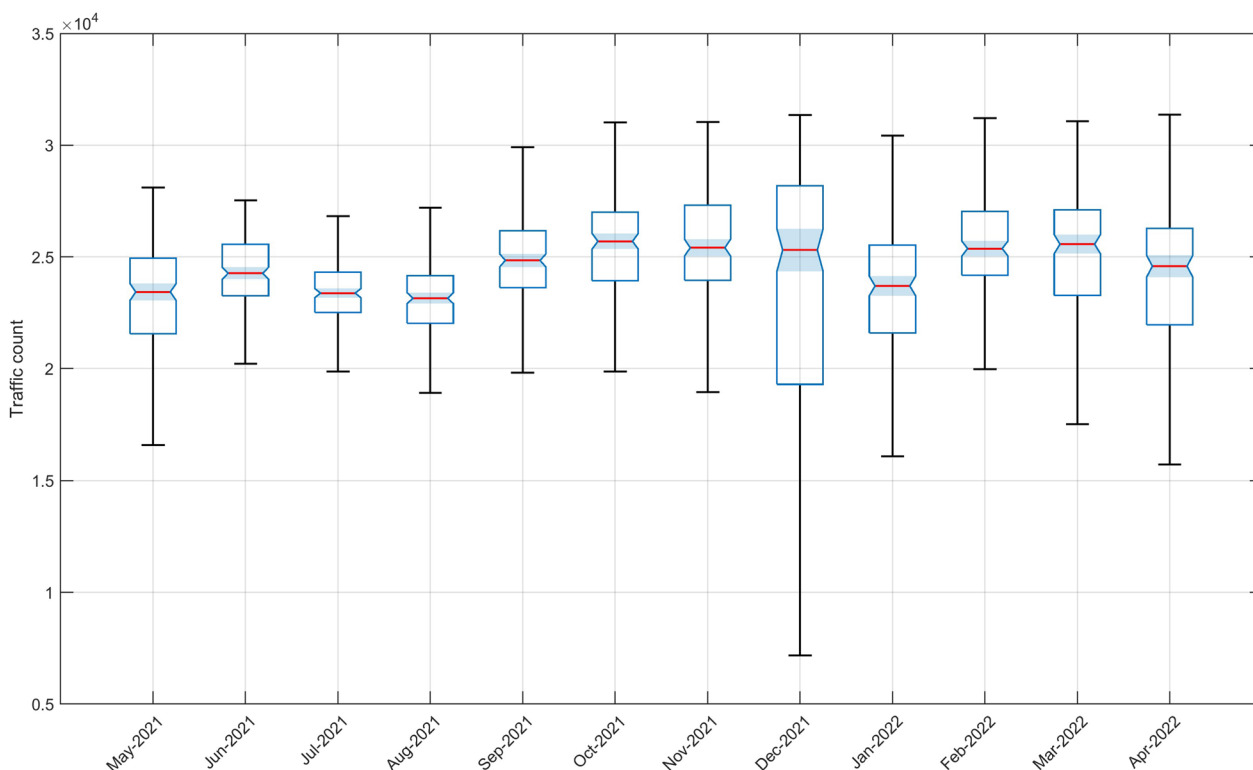


Fig. 4 Monthly distribution of traffic in study area

data shows substantial variance in traffic counts alongside elevated levels of multiple pollutants (NO_2 , CO , CO_2 , and $\text{PM}_{2.5}$). The month records the lowest traffic counts on several days but also witnesses the highest counts of the year on others, resulting in a wide distribution of traffic data for December. The median traffic count for December sits above the monthly average, underscoring that while many days experienced low traffic, there is a significant number of days with elevated traffic counts as well. The lower quartile (25%) for December traffic is markedly low, showcasing the minimum traffic days as the lowest of the year. Conversely, the upper quartile (75%) represents the highest traffic counts observed throughout the entire year. This divergence between low and high traffic days in December might correlate with the observed spike in pollutant concentrations. The days with minimal traffic show a disconnection with the elevated pollutant levels, suggesting that the sources of pollution on those days might not be traffic-related. These might be attributed to alternative emission sources, such as residential heating, industrial activities, or other non-vehicular pollution contributors that are typically higher in the winter months due to colder weather conditions, which increase the need for heating, and energy consumption patterns. On the other hand, the days with exceptionally high traffic in December could be contributing significantly to the pollution levels, potentially due to increased vehicular emissions from a larger volume of cars or possibly less

efficient vehicle operation in colder temperatures. Additionally, the occurrence of thermal inversion, which are stronger in the winter season, may be a cause of the high levels of air pollution in December. There will be less vertical circulation as the warmer air rises above and the cooler air stays near the surface. Hence, pollutants from vehicular traffic also become trapped in the lower level of the atmosphere, leading to higher concentrations of air pollutants [15, 25]; c) O_3 Observations: While O_3 is not a direct emission from traffic, its higher levels during spring (and surpassing WHO guidelines during June–August) might be linked to chemical reactions influenced by vehicular emissions under specific meteorological conditions; d) Stability of Traffic and NO Levels: The NO levels remain relatively steady, aligning with the observed stability in median traffic counts. This might suggest a direct correlation between consistent traffic flow and stable NO levels.

The following figure (Fig. 5) represents the average distribution of CO_2 emissions in the study area and the average traffic volumes in each intersection. As it can be noticed, not a clear pattern can be identified from the data presented in the figure. It becomes evident that the relationship between CO_2 emissions and traffic volumes is not straightforward. Unlike some other environmental and traffic studies that reveal a strong correlation between increased traffic and higher emissions [11, 14], our data does not exhibit such a straightforward trend.

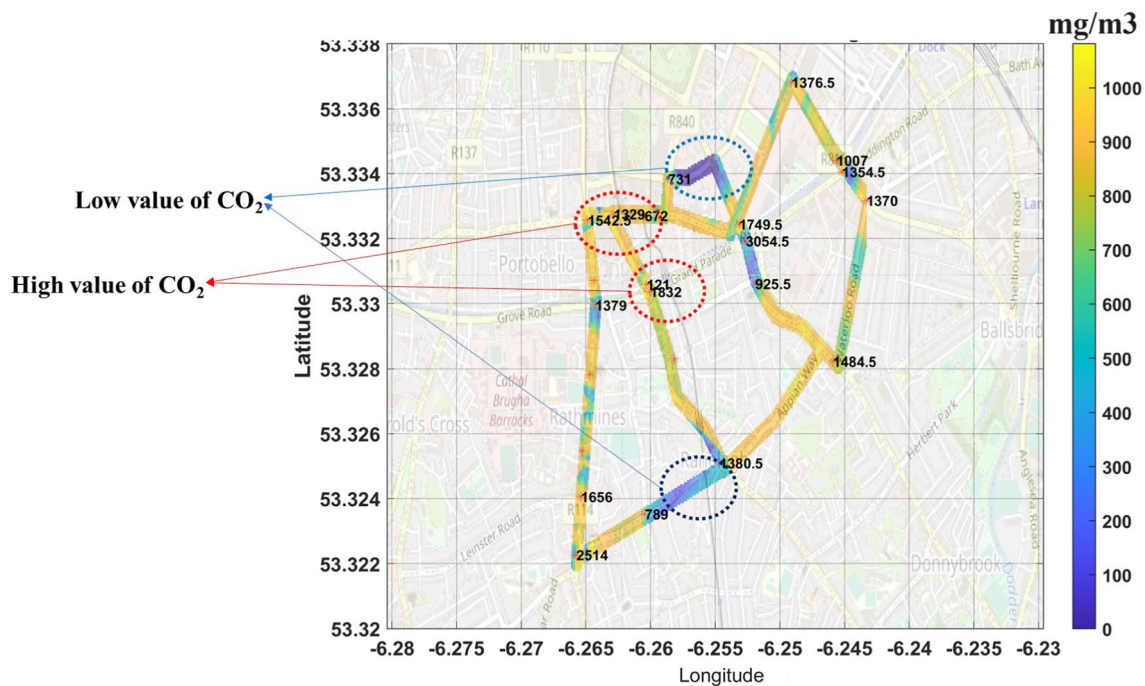


Fig. 5 Average distribution of CO_2 emissions in study area

Several factors could contribute to this lack of a clear pattern. It's important to note that the CO₂ emissions measurements obtained from GPAV data encompass various emission sources within the area, not limited solely to traffic-related emissions. This broader scope of emissions data accounts for multiple factors, including industrial processes, residential activities, etc. On the other hand, some intersections in close proximity experienced higher levels of air pollution, potentially due to increased congestion and vehicle idling. However, this isn't the case for the central intersections of the studied area. Even though the average traffic values at these intersections are significant, high values of CO₂ were observed in only one instance. Overall, the highest concentrations of CO₂ emissions were recorded on major streets.

No distinct trends were identified for the other pollutants as well. Figure 6 presents the average distributions of NO₂ (Fig. 6a) and PM2.5 (Fig. 6b) for the examined period. It is worth noting that, when comparing the distributions of CO₂, NO₂, and PM2.5, specific locations or links exhibit notably high concentrations of the respective pollutants.

To better comprehend the relationship between traffic volumes and concentrations of key pollutants, Pearson's correlation coefficients were estimated. Overall, no significant correlations were found. Among the examined pollutants, CO and CO₂ revealed the most notable correlations ($r=0.2401$ and $r=0.2377$), with their p -value of 0.0833 and 0.0866 indicating no significance.

Figure 7 presents the comparison between the GPR and SVR model results for all the pollutants in relation to the distance of the centre of study area.

In the analysis of air pollution parameters, it's important to note that these parameters can be affected by

factors beyond the immediate study area. Moreover, within the study region, we observe varying levels of air pollution at different points. To address this complexity, we have employed a spatial averaging approach. This approach considers the influence of traffic parameters and meteorological data at various distances or radii from the centre of the study area, allowing us to model changes in air pollution parameters comprehensively. In this context, we have measured average pollution parameter levels at varying distances from the centre of our study area. These average values serve as the basis for our pollution modelling. To be specific, we conducted averaging within radii ranging from 700 to 1700 m from the study area's centre, and subsequently assessed accuracy and modelling errors.

More analytically: For PM2.5, it can be observed that for the SVR model, the highest accuracy and the lowest error are at a distance of 1200 m from the study area centre, with an R² value of 0.13 and a RMSE of 3.7. In contrast, the GPR model achieves its best performance at a distance of 900 m, where the R² value is 0.54 and the RMSE is 3.1. The substantial difference in R² values between the two models indicates that GPR may be better equipped to handle the intricacies and spatial variability associated with PM2.5 emissions. For CO, the GPR model achieved a R² value of 0.33 at a distance of 1300 m with a RMSE value of 0.06. The model performed overall better than the SVR model where very low R² values were recorded. Concerning the CO₂ levels, the SVR model performed slightly better with an R² value of 0.35 at 700 m and RMSE value of 0.01, highlighting better effectiveness in capturing CO₂ concentration trends. For NO the GPR model returned negative R² values, although for NO₂, delivered better results at most distances compared

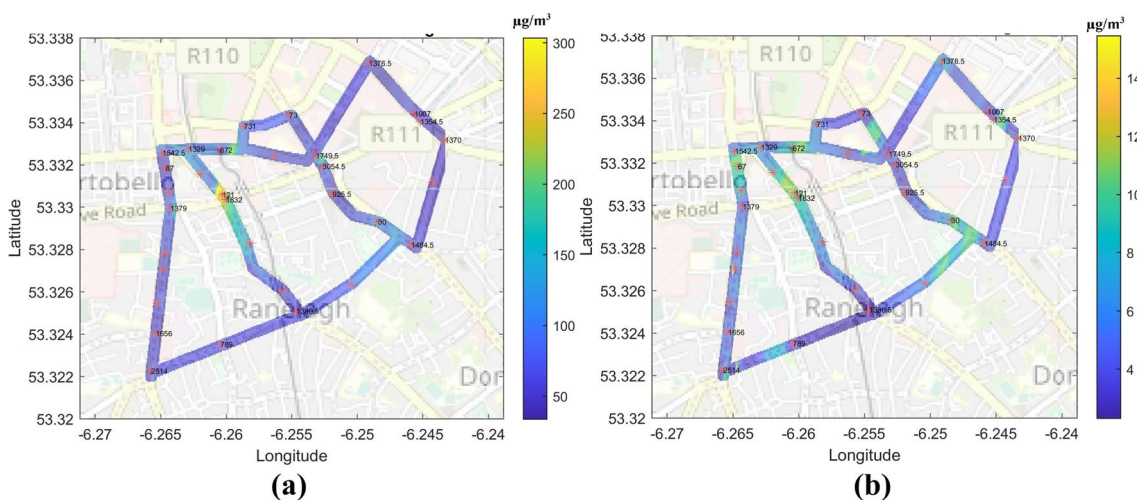


Fig. 6 Average distribution of NO₂ (a) and PM2.5 (b) emissions in study area

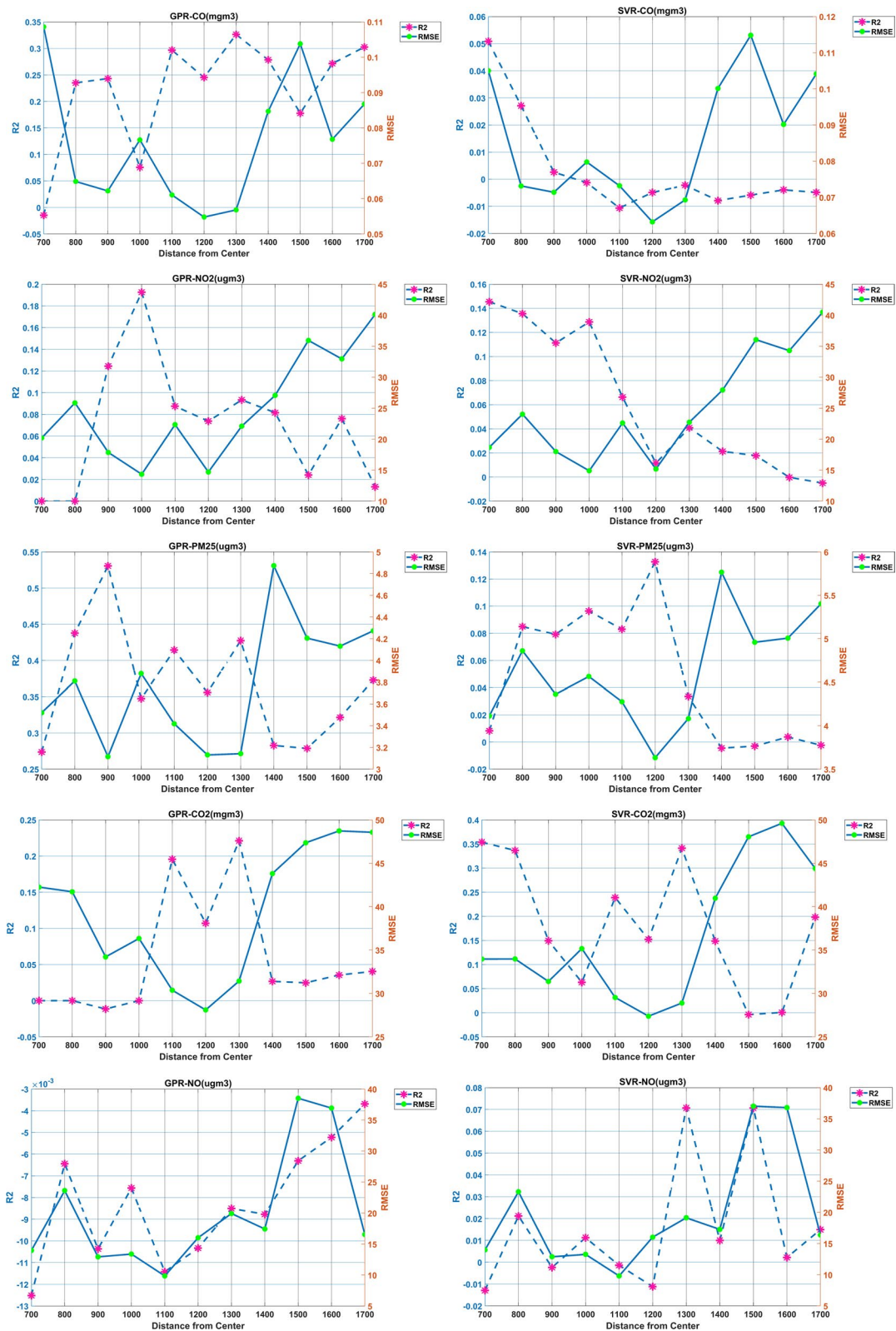


Fig. 7 Comparison between GPR and SVR model results

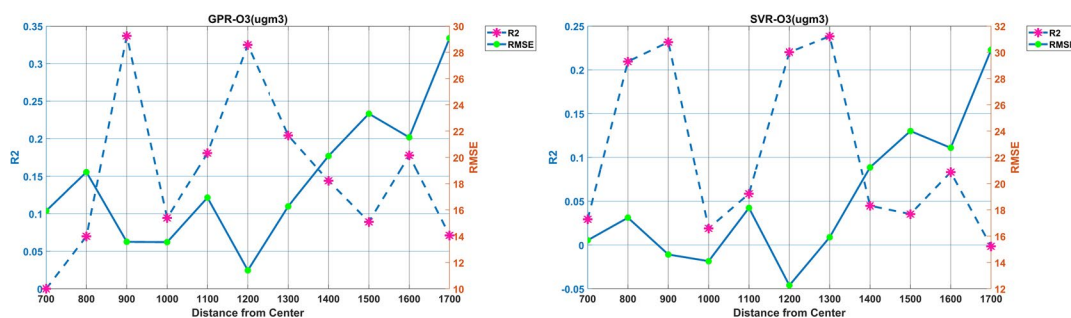


Fig. 7 continued

to SVR. Finally, in terms of O_3 levels, it was observed that the GPR model recorded better performance with an R^2 value of 0.34 at 800 m from the study area centre, and RMSE of 14, while the SVR model achieved an R^2 value of 0.21 and RMSE of 18 respectively. Given the R^2 values of each parameter, we can suggest that although the models provide some insights with moderate fits on some occasions, overall, they leave a lot of variance unexplained demonstrating that there are other variables not captured by the models.

4 Discussion

This study aimed to investigate the complex relationship between urban traffic patterns and air pollution emissions in Dublin using high-resolution datasets. By leveraging advanced machine learning techniques, we sought to uncover detailed insights that could inform urban planning and policy-making. The findings reveal several key patterns and implications that are discussed below.

Air Quality Patterns: The concentration patterns of various air pollutants unveiled in this study offer critical insights into Dublin's air quality dynamics. Notably, the peak in NO_2 levels during specific months, particularly from September to December, suggests the influence of seasonal factors and potential sources that warrant further investigation. In contrast, NO levels exhibited steadier values with minor month-to-month variations, indicating a more consistent emission source or less sensitivity to seasonal changes. O_3 concentrations displayed a seasonal rise during spring months. This seasonality has implications for public health and necessitates measures to monitor and manage ground-level O_3 concentrations during these periods. The observed spikes in CO and CO_2 emissions during specific months underscore the need for a more in-depth investigation into the contributing factors. Additionally, the pronounced concentration intervals of $PM_{2.5}$ levels highlight the complexity of fine particulate matter pollution in the Dublin area.

Traffic Patterns: The stability of traffic flow throughout most of the year indicates a consistent traffic activity

in the study area. However, the observed fluctuations in extreme values, such as the decrease in the minimum traffic count in December 2021 and the increase in maximum traffic counts in early 2022, warrant further investigation.

Emissions and Traffic: Preliminary exploration of the data showed that simple linear models might be insufficient to capture the depth and breadth of the relationships inherent in the datasets. The not clear relationship between emissions and traffic volumes suggests the presence of multiple emission sources within the study area. This observation led to the employment of advanced machine learning models like SVR and GPR.

Model Performance: The better performance of the GPR model compared to SVR for various pollutants at specific distances emphasizes the importance of considering spatial variability in air quality modelling. Overall, the R^2 values suggest a narrative of models offering some insights but with considerable unexplained variance, hinting at external variables or deeper complexities not captured by the current modelling framework.

Transferability and Broader Implications: While our study focuses on Dublin, the methodologies and insights presented here offer a robust framework for other urban areas. Cities with different industrial activities, public transport usage, and climatic conditions can adapt our modelling techniques to their specific needs. The principles underlying our approach to understanding the interplay between traffic and air pollution are widely applicable, providing valuable theoretical support for urban traffic planning and control globally.

The main limitations of this study concern the following areas: While the SVR and GPR are advanced modelling techniques, they rely heavily on the quality and comprehensiveness of the data they are trained on. Their inability to capture all the variance in our study may indicate that there are vital variables missing or that there are intricate non-linear relationships yet to be identified. We centred our analysis on traffic and meteorological conditions as the primary influences on air pollution.

However, a plethora of external factors, such as industrial activities, changes in the usage of public transport, or significant events in the city, were not integrated into our assessment but could substantially impact pollutant levels. Geographically, our research is confined to Dublin City Centre, providing valuable insights specific to this urban context. While direct extrapolation to other cities or regions should be approached with caution due to variations in traffic, weather conditions, and urban infrastructure, our findings contribute to a broader understanding of urban air pollution dynamics. They highlight the importance of considering local context in environmental management and policy formulation. Cities with different industrial activities, public transportation usage, or climate conditions may require tailored approaches to effectively address their unique air quality challenges.

The findings of this study also underscore the intricate relationship between traffic patterns and air pollution levels in Dublin, providing crucial insights for policy formulation, and environmental management. The observation that traffic volume doesn't always align with pollution peaks suggests the presence of other significant contributors to the city's air quality issues, urging policymakers and urban planners to adopt a broader perspective beyond just traffic management. As another novel aspect of our research, we tried to determine the optimal averaging distance for modelling air pollution parameters alongside various environmental factors in an urban setting. Our analysis showed that different distances worked better for different air pollution parameters, with 900 m being the most suitable for some, while others required different distances to match the data effectively. These results also illuminate the path for future research, emphasizing the need for more encompassing studies that integrate additional variables, explore potential non-linear relationships, and extend the temporal scope to capture long-term trends.

5 Conclusions

In this study, we tried to understand the impact of urban traffic on air pollution in the Dublin area by utilizing high-quality datasets. Leveraging cutting-edge ML regression techniques, including SVR and GPR, we developed prediction models for various air pollution parameters. Our study integrated urban traffic count data and meteorological information, exploring multiple buffer zones from the study area's centre. Our findings suggest that while there exists a connection between traffic and air pollution, the relationship is dynamic, influenced by a range of factors beyond mere vehicle counts.

The applied machine learning models, SVR and GPR, provided valuable insights into pollutant trends. Yet, the variance left unexplained suggests room for

refining these models, incorporating additional parameters, or exploring potential non-linear relationships. The effectiveness of our models was assessed based on higher R^2 values and lower RMSE values. Overall, GPR outperformed SVR for most air pollution parameter, showcasing its potential for more accurate air quality predictions. This study represents a preliminary step toward unravelling the complex relationship between urban traffic and air pollution. The application of advanced ML methodologies to novel data sources underscores the importance of leveraging modern technology for environmental research.

Dublin being the study's focal point, the insights gathered offer valuable lessons for urban centres globally. Our findings emphasize the complexity of urban air pollution dynamics and highlight the need for holistic, multi-pronged solutions that consider a variety of local factors. While direct extrapolation to other cities should be approached with caution due to unique local conditions, the methodologies and approaches used in this study provide a robust framework that can be adapted and applied to different urban contexts.

This study has revealed the intricate interplay between traffic patterns and air pollution levels, underscoring that traffic volume alone does not account for pollution peaks. This suggests that urban air quality management requires a comprehensive understanding of multiple contributing factors, including industrial processes, residential activities, public transportation usage, and meteorological conditions. The novel aspect of determining optimal averaging distances for modelling air pollution parameters also offers valuable insights. Different distances worked better for different pollutants, suggesting that urban planners and researchers should tailor their approaches based on specific environmental and infrastructural conditions.

Future studies should explore the nuances of air quality dynamics, delve into the identification of specific pollution sources, and consider the influence of various factors beyond traffic and meteorology.

Acknowledgements

This work is funded by the Dublin City Council through the Google Air Research Bursary.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Pavlos Tafidis, Mehdi Gholamnia, Payam Sajadi, and Sruthi Krishnan V. The first draft of the manuscript was written by Pavlos Tafidis, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This article has been published open access with support of the TRA2024 project funded by the European Union.

Availability of data and materials

Data is available online at the following links: <https://data.gov.ie/dataset/google-airview-data-dublin-city> and <https://data.gov.ie/dataset?tags=traffic-counts>. Data can also be made available upon request.

Declarations**Competing interests**

The authors declare no competing interests.

Received: 11 October 2023 Accepted: 22 August 2024

Published online: 02 September 2024

References

- Abdull, N., Yoneda, M., & Shimada, Y. (2020). Traffic characteristics and pollutant emission from road transport in urban area. *Air Quality, Atmosphere & Health*, 13(6), 731–738. <https://doi.org/10.1007/s11869-020-00830-w>
- Coelho, S., Ferreira, J., Lopes, D., Carvalho, D., & Lopes, M. (2023). Facing the challenges of air quality and health in a future climate: The Aveiro region case study. *Science of the Total Environment*, 876, 162767. <https://doi.org/10.1016/j.scitotenv.2023.162767>
- Data.gov.ie. (2023a). Google project air view data-Dublin City (May 2021–August 2022). Last accessed on 9 October 2023. <https://data.gov.ie/datas/google-airview-data-dublin-city>
- Data.gov.ie. (2023b). Traffic counts datasets. Last accessed on 9 October 2023. <https://data.gov.ie/dataset?tags=traffic-counts>
- EPA. (2022). European city air quality viewer. Retrieved from <https://www.eea.europa.eu/themes/air/urban-air-quality/european-city-air-quality-viewer>
- Fotheringham, A. S., Yue, H., & Li, Z. (2019). Examining the influences of air quality in China's cities using multi-scale geographically weighted regression. *Transactions in GIS*, 23(6), 1444–1464. <https://doi.org/10.1111/tgis.12580>
- Ibarra-Espinosa, S., Ynoue, R. Y., Ropkins, K., Zhang, X., & de Freitas, E. D. (2020). High spatial and temporal resolution vehicular emissions in south-east Brazil with traffic data from real-time GPS and travel demand models. *Atmospheric Environment*, 222, 117136. <https://doi.org/10.1016/j.atmosenv.2019.117136>
- Jeong, C.-H., Wang, J. M., Hilker, N., Deboz, J., Sofowote, U., Su, Y., & Evans, G. J. (2019). Temporal and spatial variability of traffic-related PM_{2.5} sources: Comparison of exhaust and non-exhaust emissions. *Atmospheric Environment*, 198, 55–69. <https://doi.org/10.1016/j.atmosenv.2018.10.038>
- Kaur, J., & Jhamaria, C. (2021). Urban air pollution and human health: A review. *Current World Environment*, 16(2), 362–377. <https://doi.org/10.12944/cwe.16.2.04>
- Kumar, A., Patil, S., Kovacevic, A., & Ponnusami, S. A. (2024). Performance prediction and Bayesian optimization of screw compressors using Gaussian Process Regression. *Engineering Applications of Artificial Intelligence*, 133, 108270.
- Lei, H., Zeng, S., Namaiti, A., & Zeng, J. (2023). The impacts of road traffic on urban carbon emissions and the corresponding planning strategies. *Land*, 12(4), 800. <https://doi.org/10.3390/land12040800>
- Liu, H., Rodgers, M. O., & Guensler, R. (2019). The impact of road grade on vehicle accelerations behavior, PM_{2.5} emissions, and dispersion modeling. *Transportation Research Part D: Transport and Environment*, 75, 297–319. <https://doi.org/10.1016/j.trd.2019.09.006>
- Liu, T., Sun, J., Liu, B., Li, M., Deng, Y., Jing, W., & Yang, J. (2022). Factors influencing O₃ concentration in traffic and urban environments: A case study of Guangzhou city. *International Journal of Environmental Research and Public Health*, 19(19), 12961. <https://doi.org/10.3390/ijerph191912961>
- Ma, J., Xu, M., & Jiang, J. (2023). Mapping high-resolution urban road carbon and pollutant emissions using travel demand data. *Energy*, 263, 126059. <https://doi.org/10.1016/j.energy.2022.126059>
- Martilli, A., Sánchez, B., Santiago, J. L., Rasilla, D., Pappaccogli, G., Allende, F., & Fernández, F. (2022). Simulating the pollutant dispersion during persistent Wintertime thermal Inversions over urban areas. *The case of Madrid. Atmospheric Research*, 270, 106058. <https://doi.org/10.1016/j.atmosres.2022.106058>
- met.ie. (2023). Weather data. Last accessed on 9 Oct 2023. <https://www.met.ie/climate/weather-observing-stations>
- Pai, S. J., Carter, T. S., Heald, C. L., & Kroll, J. H. (2022). Updated world health organization air quality guidelines highlight the importance of non-anthropogenic PM_{2.5}. *Environmental Science & Technology Letters*, 9(6), 501–506. <https://doi.org/10.1021/acs.estlett.2c00203>
- Quintyne, K. I., Kelly, C., Sheridan, A., Kenny, P., & O'Dwyer, M. (2021). COVID-19 transport restrictions in Ireland: Impact on air quality and respiratory hospital admissions. *Public Health*, 198, 156–160. <https://doi.org/10.1016/j.puhe.2021.07.008>
- Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *Journal of Computer-Aided Molecular Design*, 36(5), 355–362. <https://doi.org/10.1007/s10822-022-00442-9>
- Sarker, I. H. (2021). Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, 2(5), 377. <https://doi.org/10.1007/s42979-021-00765-8>
- Tang, J., McNabola, A., & Misstear, B. (2020). The potential impacts of different traffic management strategies on air pollution and public health for a more sustainable city: A modelling case study from Dublin, Ireland. *Sustainable Cities and Society*, 60, 102229. <https://doi.org/10.1016/j.scs.2020.102229>
- Tang, J., McNabola, A., Misstear, B., & Caulfield, B. (2017). An evaluation of the impact of the Dublin port tunnel and HGV management strategy on air pollution emissions. *Transportation Research Part D: Transport and Environment*, 52, 1–14. <https://doi.org/10.1016/j.trd.2017.02.009>
- Tang, J., McNabola, A., Misstear, B., Pilla, F., & Alam, M. S. (2019). Assessing the impact of vehicle speed limits and fleet composition on air quality near a school. *International Journal of Environmental Research and Public Health*, 16(1), 149. <https://doi.org/10.3390/ijerph16010149>
- TomTom. (2022). Dublin in traffic index 2021. Retrieved from <https://www.tomtom.com/traffic-index/dublin-traffic/>
- Trinh, T. T., Trinh, T. T., Le, T. T., Nguyen, T. D. H., & Tu, B. M. (2018). Temperature inversion and air pollution relationship, and its effects on human health in Hanoi City, Vietnam. *Environmental Geochemistry and Health*, 41(2), 929–937. <https://doi.org/10.1007/s10653-018-0190-0>
- Wang, Q., Feng, H., Feng, H., Yu, Y., Li, J., & Ning, E. (2021). The impacts of road traffic on urban air quality in Jinan based GWR and remote sensing. *Scientific Reports*, 11(1), 15512. <https://doi.org/10.1038/s41598-021-94159-8>
- Yuan, C., Ng, E., & Norford, L. K. (2014). Improving air quality in high-density cities by understanding the relationship between air pollutant dispersion and urban morphologies. *Building and Environment*, 71, 245–258. <https://doi.org/10.1016/j.buildenv.2013.10.008>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.