scientific reports

OPEN

Transforming CCTV cameras into NO_2 sensors at city scale for adaptive policymaking

Mohamed R. Ibrahim[1,2] & Terry Lyons[1,3]

Air pollution in cities, especially NO_2, is linked to numerous health problems, ranging from mortality to mental health challenges and attention deficits in children. While cities globally have initiated policies to curtail emissions, real-time monitoring remains challenging due to limited environmental sensors and their inconsistent distribution. This gap hinders the creation of adaptive urban policies that respond to the sequence of events and daily activities affecting pollution in cities. Here, we demonstrate how city CCTV cameras can act as a pseudo-NO_2 sensors. Using a predictive graph deep model, we utilised traffic flow from London's cameras in addition to environmental and spatial factors, generating NO_2 predictions from over 133 million frames. Our analysis of London's mobility patterns unveiled critical spatiotemporal connections, showing how specific traffic patterns affect NO_2 levels, sometimes with temporal lags of up to 6 h. For instance, if trucks only drive at night, their effects on NO_2 levels are most likely to be seen in the morning when people commute. These findings cast doubt on the efficacy of some of the urban policies currently being implemented to reduce pollution. By leveraging existing camera infrastructure and our introduced methods, city planners and policymakers could cost-effectively monitor and mitigate the impact of NO_2 and other pollutants.

Keywords Air quality, Computer vision, Deep learning, Environmental analysis, Policy-making, Cities

Cities house more than half of the world's population[1], which influence individuals' behaviour[2] as well as their physical[3,4] and mental health[5]. Every day, hundreds of millions of people spend several hours commuting on the spatial network of cities exposed to several risks, including air pollution. There is no dispute about the need for developing a fundamental understanding of how, collectively, individuals move from one location to another in their daily lives. This could be linked with pollution indicators to aid in emission reduction.

Nitrogen dioxide (NO_2) is a major pollutant that can harm severely one's health[6–12]. NO_2 is formed by the combustion of fuels such as natural gas, diesel, petrol, and coal, and it can be found in the air as a result of traffic or a variety of land uses in cities, including industrial processes. NO_2 levels (measured in $\mu g/m^3$) vary in major cities worldwide[13]. Several studies have mapped NO_2 emissions from space[13–21], whether during pandemics[13,22] or after a policy is implemented[14,16,20,23]. While relying on satellite imagery is beneficial for many cases, including understanding the change in emission over a long period or across several large cities[15,16,20,22,24], the spatial and temporal representations are often limited for understanding the dynamics of emission at a neighbourhood, district, or even many of the cities globally. Consequently, a substantial knowledge gap exists in linking micro-level events occurring frequently to their impact on emissions, thereby hindering the ability of policymakers to take localised actions. The objectives of this study are as follows: (1) to what extent the existence of specific traffic modes influences the surface NO_2 level, (2) what effect congestion and stationary modes have on the level of NO_2, and (3) whether there is a significant temporal lag between what happens in traffic now and its impact on the future level of NO_2 at a given location.

Analysing urban dynamics at the street level through visual data can uncover details that may be missed when when observing from space[25]. Recent progress in deep learning for predicting traffic flow[26] aids in estimating pollutant levels in cities. Multi-modal sensor fusion has advanced by integrating data from various sensors to improve environmental predictions[27]. These techniques could enable air quality estimation by combining CCTV visuals with other sensor data. Effective sensor deployment is crucial for urban-scale monitoring to ensure comprehensive coverage and reliable data collection[28]. In this study, we introduce innovative techniques that leverage statistical analysis and graph neural networks to sense ambient ground-level NO_2 concentrations and their underlying factors using CCTV camera feeds on a citywide scale. This approach proves invaluable, especially in cities lacking an extensive network of environmental sensors. It provides an automated means

[1]The Alan Turing Institute, London, UK. [2]Institute for Spatial Data Science, University of Leeds, Leeds, UK. [3]Mathematical Institute, Oxford University, Oxford, UK. email: geomi@leeds.ac.uk

of detecting the concentration of NO_2 levels and their causes related to the dynamics of traffic, empowering urban planners, and policymakers to actively monitor and respond to emerging issues in real-time, guided by the dynamic flow patterns within cities. Our methodology offers a non-physical (hardware-free) solution for monitoring ground-level NO_2 in urban areas where CCTV cameras are prevalent but NO_2 sensors are scarce, a situation encountered in numerous cities worldwide.

Results

Multi-level spatiotemporal representation of traffic modes

To understand the influence of individual road users and their transportation modes on NO_2 ground-levels within the city, adopting a bottom-up approach that details individual trajectories is crucial. This strategy is invaluable for accurately assessing the real-time NO_2 concentrations at specific locations and times, as well as evaluating the exposure that individuals face during their commutes. Previous research across various domains has explored the use of human trajectories from GPS data for similar assessments[29–32]. However, the limited availability of such data and substantial privacy concerns complicate the widespread replication of these methods. Therefore, it is imperative to discover alternative data sources that can accurately reflect traffic dynamics and roadway user behaviours while preserving anonymity. Successfully identifying such sources is key to advancing this study and enabling its future application across global urban landscapes to enhance our understanding of ground-level NO_2 distributions and their impacts on public health.

We used an open-access video data set provided by Transport For London (TfL), which includes unidentifiable human subjects and road users[33]. We recorded and analysed 133,132,866 sequential frames representing 112 unique hours in 907 London locations. We recorded many features of road users by utilising deep learning in our proposed framework. We refers to 'flows' as the movement patterns of road users captured by CCTV cameras across different locations and times within the city. These flows represent the dynamic interactions and traffic patterns, identified through the analysis of sequential frames in video data. By "flows," we mean the aggregated and continuous movement of vehicles and pedestrians detected and tracked through video footage. This term encompasses both the spatial and temporal dimensions of traffic, enabling us to infer NO_2 levels from the volume and behaviour of traffic over given periods.

Figure 1 illustrates the variables analysed and the structured hierarchy used to represent data for this study's various components. The data aims to depict diverse events and aspects of urban environments across different spatial and temporal scales (Fig. 1A,C). For example, the spatial distribution of data derived from camera streams does not necessarily match the spatial distribution of NO_2 sensors (Fig. 1B). Additionally, the temporal characteristics of data sourced from cameras, static spatial features, and NO_2 measurements differ (Fig. 1C). At a micro-level of a given street, we extracted road users which were given a unique ID across the frame sequence of a given video file of a time increment of a given hour. Afterwards, a unique traffic modal flow (o) for a given hour is defined as where q is the different modal flows and F is the number of different video files representing time increments of a given hour. At a city scale, the data is combined for each unique hour (H) of a given date (d) and hour (t). The overall Spatiotemporal representations of the CCTV data (X) is structured as $X \in \mathbb{R}^{HXNXFXC}$ and the generated NO_2 (Y) as $Y \in \mathbb{R}^{HXM}$, where H is the number of unique hours, N is the number of cameras' locations, C is the number of features, including modal flows and locational urban features, and M is the number of NO_2 sensors' locations where $M \neq N$. The spatiotemporal representations of cameras' data and NO_2 sensors differ in position and temporal resolution, and they are aligned based on the sparse availability at hourly rates of NO_2 sensor data. The static urban features of a specific site are combined with the aligned locations of both sensor data. Time resolution remains as a variable depending on a given scale; moving from 0.04 s at a frame level to 4 min in a trajectory level and finally to 1 h at an aggregate higher level. The construction of a non-linear tree data structure allows for the insertion, search, and relocation of new branches over time. It also supports this research by responding to stated questions that may require different spatial and temporal resolutions.

Traffic composition at micro-scale

To address how we can use high-frequency data (0.04 s) of the number of road users and their behaviour (moving, stationary, etc.) to provide meaningful statements for NO_2 at an hourly city level, we must first collect and understand the collective patterns of road users at a micro-level that derive the overall traffic in London. We demonstrate, in Fig. 2, how to transform the sequential frames of a given video to spatial and temporal representations of road users, and georeferencing their representation in a bird's eye view map blended with Google map. We determined the modal flows based on the monitored unique ids of road users through the length of a given file to avoid re-counting the same users (Fig. 2B). Lastly, to provide a unique summary of the observed sequence of the events of multidimensional streams of road users based on their types and behaviour at a given camera, we computed a signature, based on rough path theory[34–36], Sig^N of depth $N = 3$ for a given stream $X \in \mathbb{R}^{n\times f\times c}$, given that n is the number of cameras ($n = 906$), f is the number of file increments that make an hour of traffic modes ($f = 11$) and c is the number of channels for traffic modal flows and their stationary status ($c = 13$). The collection of computed signatures for all cameras for a given hour is invariant to path reparameterization. This provides (1) a natural characteristic of linear functionals, which only capture the main aspects of the provided path by mapping the sequence of the stream's information rather than mapping the exact position of the path at each occurrence, and (2) the ability to retrieve the original stream of road users and their behaviour from the lower-dimensional signature, minimising computational and memory footprint (Fig. 2C).

The effect of location and environment on ground-level NO_2

Geographical factors, such as the proximity to farmland, industrial zones, or various land uses, significantly influence traffic patterns and, as a result, levels of NO_2 (See Fig. 3A). To investigate the spatial relationship

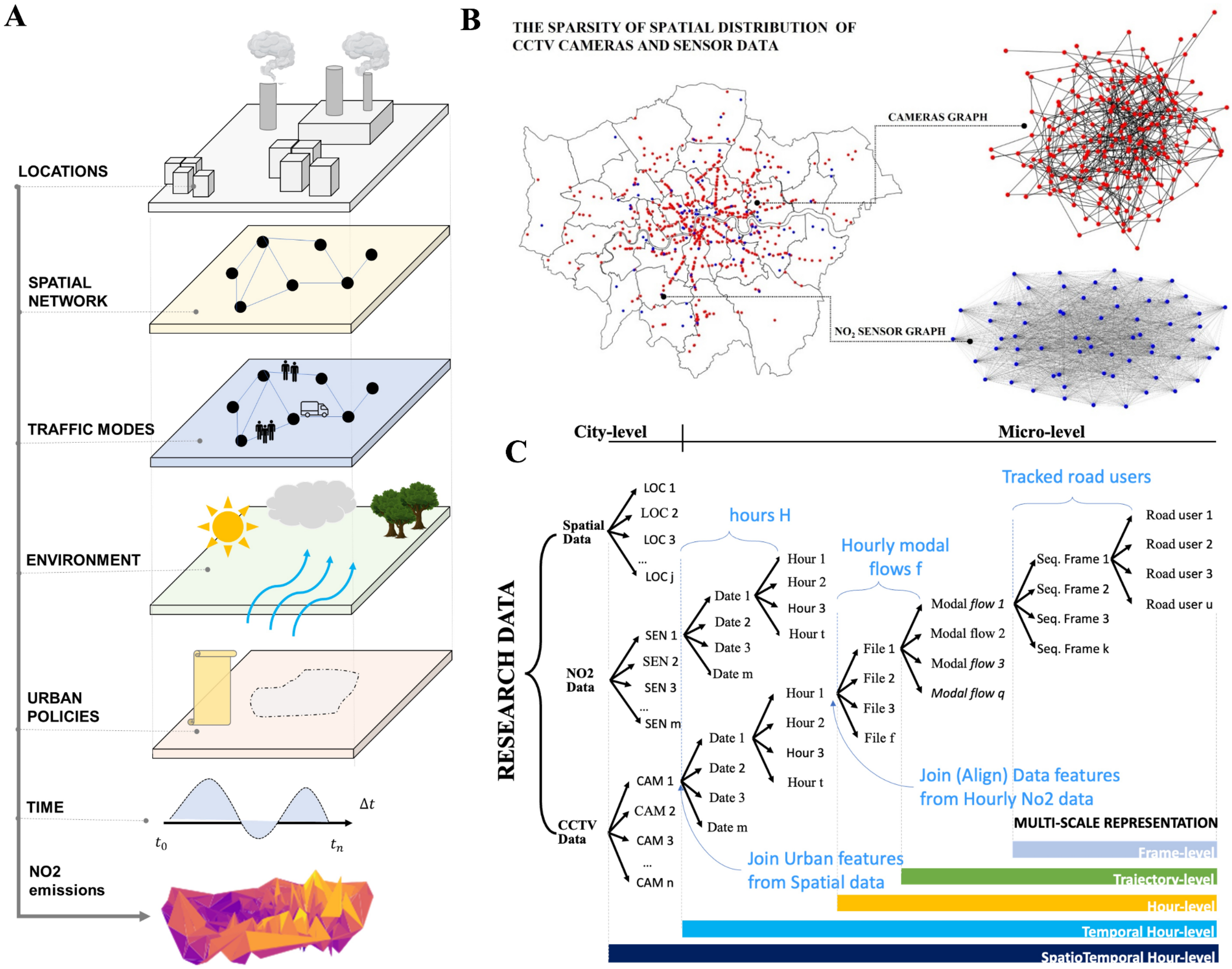


Fig. 1. Multi-level representation of all data sources. (**A**) The six layers of factors presented in this research. (**B**) The spatial representation as graph of knowledge of the camera (red nodes) and NO_2 (blue nodes) inputs. The locations of the cameras and NO_2 sensors do not need to align. The figure was created by the first author using Python programming language (using geopandas and matplotlib). (**C**) The multi-level representation of the studied data modalities shows several spatial and temporal resolutions in which different data modalities are aligned to conduct this research.

between NO_2 and traffic, we developed a hot spot analysis to cluster total traffic and NO_2 levels based on the spatial dependency of neighbouring high or low values, yielding statistically significant clusters ($p < 0.05$) of spatial outliers. Here, we show a spatial lag when examining the locations of hot spots for both variables at a given time (See Fig. 3B,C).

We observe a spatial lag which could be attributed to confounding variables related to environmental factors such as rainfall, wind speed, and direction that either concentrate or disperse emissions from their sources. Moreover, the observed spatial lag may also be linked to the lifetime of NO_2[20,37–42], which introduces a temporal delay between traffic emissions and the resultant ground-level concentration of NO_2 detected in a specific area. We will further investigate this in the following section by relying on Granger Causality analysis, which helps in understanding and measuring the delayed effects of traffic emissions on NO_2 ground-level concentrations. However, as a first step, we used a spatial two-stage least squares model to investigate various variables related to geographical characteristics, environment, and day of the week (See Fig. 3D). We discovered that proximity to industrial zones within one mile ($\beta = 2.156, p = 0.000$), boroughs within Ultra Low Emission Zones (ULEZ)[43] ($\beta = 3.075,\ p = 0.000$), wind speed ($\beta = 2.843,\ p = 0.000$), sun hours ($\beta = 6.438,\ p = 0.000$), rainfall ($\beta = 43.571,\ p = 0.000$), South West winds ($\beta = 6.761,\ p = 0.0001$), congestion ($\beta = 0.060,\ p = 0.000$), and the change in atmospheric pressure ($\beta = 3.243,\ p = 0.000$) are more likely to contribute linearly to the level of NO_2 at a given location. Conversely, the number of wet hours in a given day ($\beta = -33.782,\ p = 0.000$), the change in average temperature ($\beta = -2.486,\ p = 0.000$), North East wind ($\beta = -26.877,\ p = 0.000$), average speed limit of a given road ($\beta = -0.042,\ p = 0.000$) and proximity to farmland ($\beta = -0.805,\ p = 0.0013$) are negatively linear with the emission. We further investigate the temporal dependency of traffic modes within a given hour of the day.

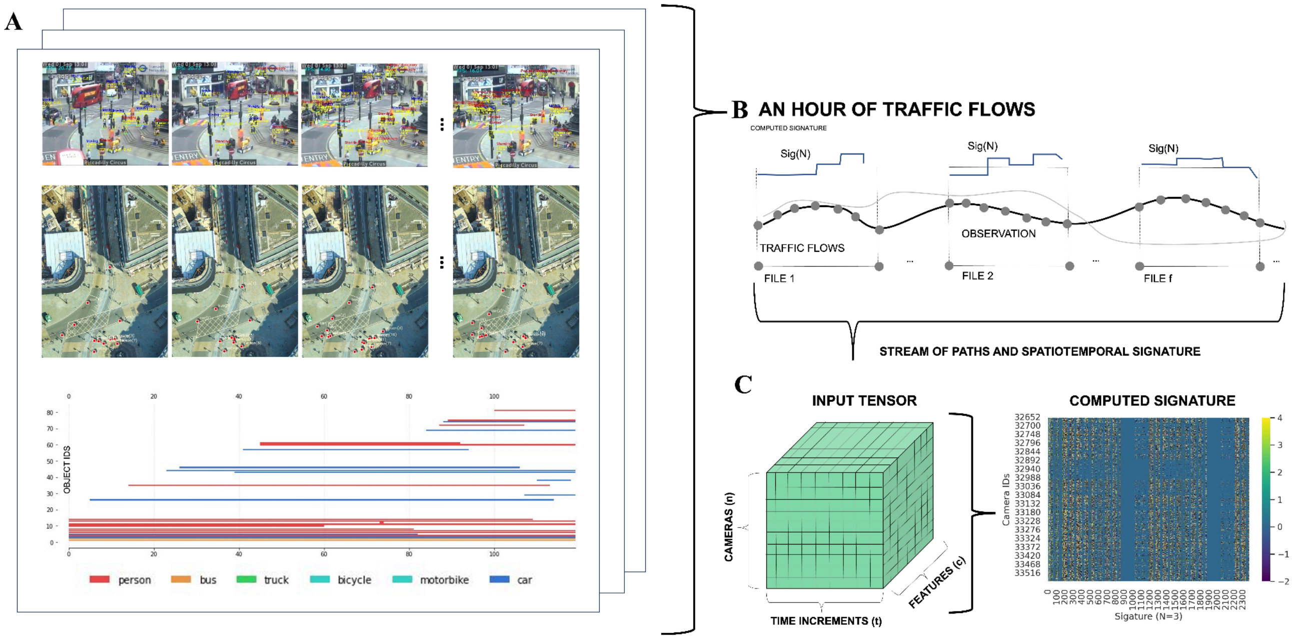


Fig. 2. Capturing the order of events from micro-level to city-scale (**A**) Shows (1) sequential frames of a given video file at Piccadilly circus in London as an input of one given CCTV camera, (2) vector representation of road users in an estimated bird's eye view map with Google Maps to validate the geographical localisations of road users, and (3) temporal representation of road users within a given file based on the tracked system. (**B**) The relationship between hourly observed traffic flow data and the unseen temporal intervals among various file increments representing the stream of paths $X \in \mathbb{R}^{(nXtXc)}$, given that n is the number of cameras ($n = 907$), t is the number of file increments that make an hour of traffic modes ($t = 11$) and c is the number of channels for traffic modal flows and their stationary status ($c = 13$) (**C**) The tensor representation of the generated paths with all its channel and their unique computed signatures (Sig^N, $N = 3$) that summarise the paths of varied traffic modal flows and their actions in a given scene.

The effect of time and the dynamics of traffic modes on ground-level NO_2

Given the relationship between NO_2 levels and total traffic is nonlinear at all times and locations (See Fig. S2-A in supplementary), modelling NO_2 ground-levels requires considering the entire urban landscape as an integrated dynamic system. This approach is especially pertinent because air pollution tends to diffuse and is influenced by numerous factors, such as wind speed, direction, existing green spaces, and proximity to industrial zones or farmlands, in which we have studied. These elements collectively contribute to a nonlinear impact on localised NO_2 levels within the network.

Moreover, NO_2's behaviour in the atmosphere adds another layer of complexity to this topic. NO_2 can have variable lifetimes in the air, ranging from a few hours to a whole day depending on meteorological conditions and the presence of other chemical species[20,37–42]. During daylight hours, UV light from the sun can drive photolytic reactions that convert other nitrogen oxides such as NO into NO_2, further altering the dynamics of air quality. This chemical interplay indicates that emissions and concentrations of NO_2 are fluid, changing not just with traffic flow and industrial activity, but also with the shifting patterns of sunlight and weather.

Despite the complicated dynamics influenced by environmental and chemical processes, there is a discernible linear relationship between NO_2 and types of traffic observed over the course of a day at specific camera locations. This linearity in smaller, more controlled environments suggests that while broader city-wide models must account for complex inter-dependencies and nonlinear behaviours, localised predictions and assessments can successfully utilise simpler linear models. This dichotomy highlights the need for a layered approach in environmental monitoring and management, blending both detailed, location-specific data and broader, systemic perspectives to form a comprehensive understanding of urban air quality.

Building on this, the temporal dynamics play a crucial role in analysing the patterns of NO_2. To dissect how each factor influences NO_2 levels at distinct times, we implemented two distinct statistical methodologies. Firstly, we employed a spatial regression model for each hour of the day, resulting in 24 unique models. This method helps identify the direct impact of various factors on NO_2 levels at specific hours. Secondly, to explore how each factor may influence future levels of NO_2, we developed a Granger Causality analysis model for each factor (8 models in total). This technique is particularly useful for pinpointing significant temporal lags and understanding the predictive relationship between the factors and subsequent NO_2 concentrations. These approaches allow us to identify not only the immediate effects of factors on NO_2 levels but also their delayed impacts, thus providing a more comprehensive understanding of the temporal dynamics at play. This layered analysis ensures a more nuanced insight into the cyclic and predictive behaviours of NO_2 in relation to traffic and environmental influences.

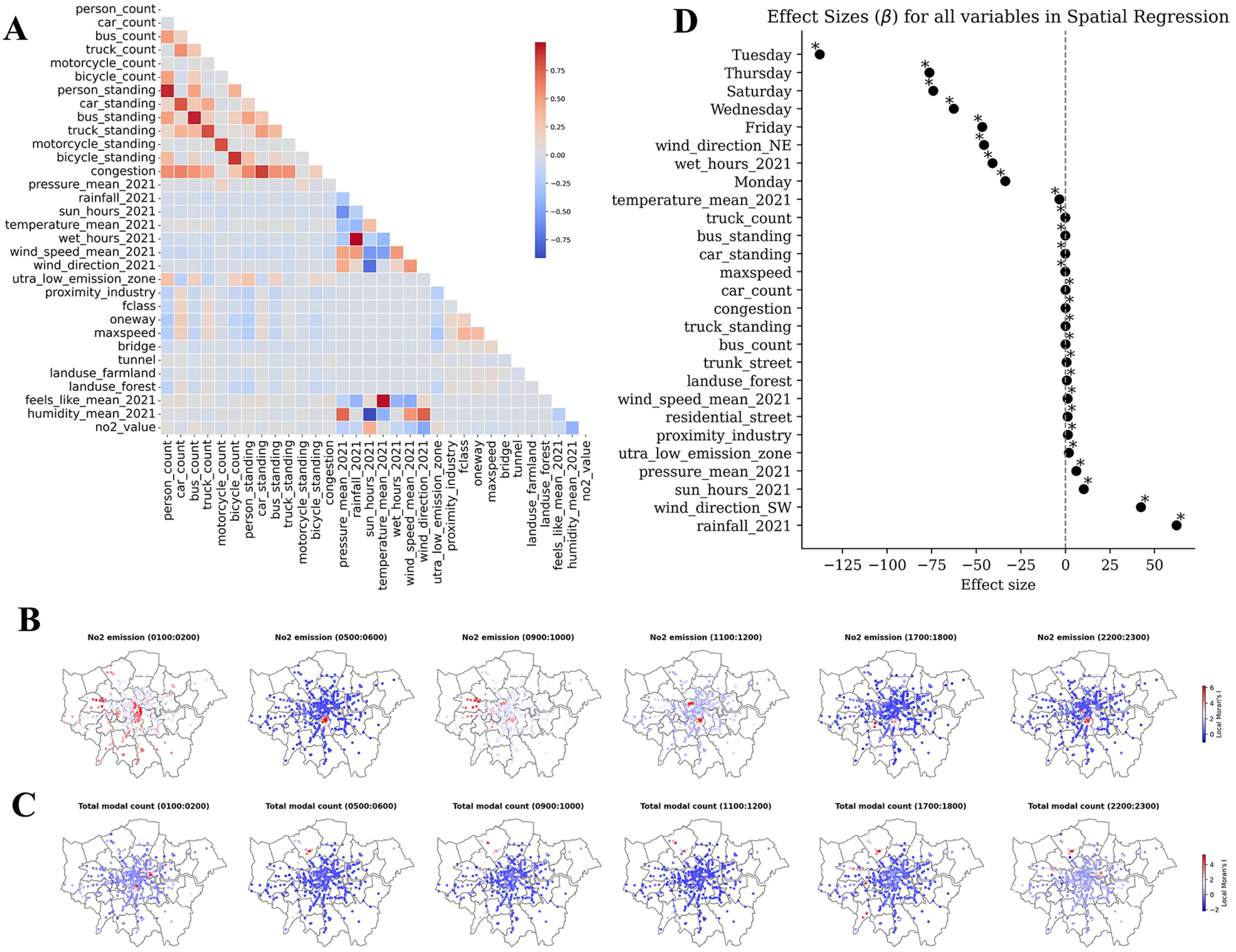


Fig. 3. The spatial patterns of NO_2 and traffic at city-scale. (**A**) The association between the studied variables relies on Pearson's correlation. (**B**) Hot spot analysis using significant Moran's I z-value ($P < 0.05$) to highlight the outliers of NO_2 across different hours of the day (the rest of the 24 h are presented in supplementary). (**C**) Hot spot analysis using significant Moran's I z-value to highlight the outliers of total flow across different hours of the day. (**D**) Statistically significant results ($p < 0.05$, $r^2 = 0.4$, $spatialr^2 = 0.23$,and $df = 88{,}020$) of the spatial two-stage least-square model, variables are shown based on the sign and weight of their β value.

Furthering our understanding of the temporal dynamics, Fig. 4 shows a novel visual representation of the NO_2 clock, showcasing statistically significant linear relations between certain factors and NO_2 levels, characterised for each hour of the day. This graphical display helps to encapsulate NO_2 levels and the main associations observed: for instance, trucks exhibit a consistent linear correlation with NO_2 during midday, night, and the early hours of the morning. In contrast, buses tend to influence NO_2 levels predominantly during the morning and afternoon peak traffic periods. Stationary cars contribute to air pollution during the peak morning hours around 10 am, and their influence extends into midday, primarily while idling in traffic jams. This is different from other periods when stationary vehicles, mainly parked, have little or no impact on pollution. During busy traffic, however, the idling of these cars significantly elevates NO_2 levels. Expanding on these observations, the data also reveals that stationary buses notably contribute to NO_2 during the morning rush hours (8–9 AM). Furthermore, locality factors such as proximity to industrial areas (within a one-mile radius) demonstrate a substantial effect on NO_2 concentrations during specific times-specifically in the evening (7–8 PM) and early morning hours. These insights underscore not only the diverse temporal relationships between different vehicles and NO_2 concentrations but also illuminate the role of geographic and stationary factors in influencing air quality at different times of the day. This level of detail enriches our understanding of urban air pollution dynamics and highlights the critical interplay between temporal, vehicular, and locational determinants in shaping urban NO_2 levels.

Expanding on the analysis of significant temporal lags where specific traffic modes influence and Granger-cause future NO_2 levels, our data demonstrates that the time series of each traffic mode Granger-causes the series of NO_2 with notable statistically significant lagged values. For instance, car flows are likely to Granger-cause NO_2 levels with lag times ranging from 2 to 6 h, varying by location. Meanwhile, stationary cars manifest a more

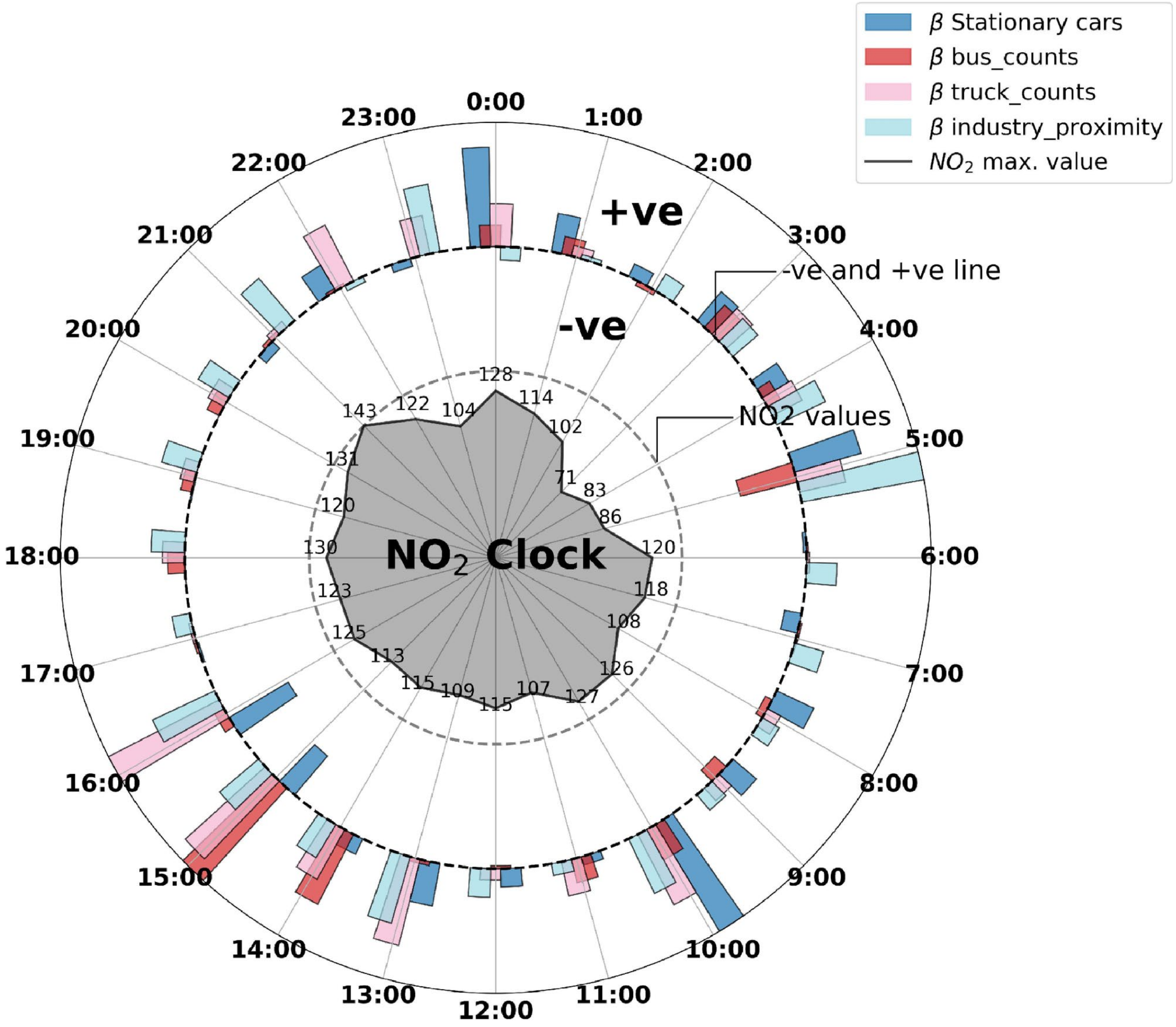


Fig. 4. NO_2 clock. It shows the hourly average levels of NO_2 and the factors influencing these levels, based on a spatial regression model for each hour. It features three concentric circles: the innermost represents the average NO_2 concentration per hour, the middle circle shows factors negatively correlated with NO_2, and the outermost highlights positively correlated factors. Each factor's influence is quantified by a β value, indicating its effect size relative to the hourly covariates, factors, and overall impact on NO_2. All β values are standardised across all hours. For simplicity and clarity, the figure displays only four variables, although the full model considers a more extensive range of variables detailed in Table S1.

immediate impact on NO_2 concentrations, typically with a 2-h lag. In terms of heavier traffic elements, congested traffic flows and stationary buses exert a more prolonged effect on NO_2 levels, showing significant impacts at lags of 5 and 6 h. Stationary trucks, on the other hand, show a swift influence with only a 1-h lag, suggesting their emissions rapidly integrate into the local atmosphere. Conversely, moving trucks have a more extended influence, where the current flows can predict NO_2 levels up to 5 h into the future. These findings are also linked to the chemical behaviour of NO_2 in urban air. The timeline of influence observed ties back to the variable atmospheric lifetime of NO_2[20,37–42], which can differ from several hours to a full day, influenced by ambient conditions such as sunlight and temperature. Solar radiation promotes the photolytic cycle that converts NO to NO_2, fundamentally affecting how quickly emissions from traffic transform into atmospheric NO_2. Therefore, the timing of traffic flows and their characteristic effects on NO_2 can directly correlate with these natural diurnal variations, reinforcing the need to consider both chemical kinetics and traffic dynamics when analysing urban air quality patterns. This multi-faceted approach provides a richer, more accurate depiction of NO_2 ground-level, particularly in dense urban environments where traffic and industrial emissions often overlap.

The impact of policies on the dynamics of ground-level NO_2

Not only do factors connected to place and time have a significant impact on NO_2 levels, but so do the measures and regulations implemented in London driven by specific location and time. According to our Granger analysis, the effect of traffic in a given location on the level of NO_2 can appear after several hours, we found that limiting certain traffic modes, such as trucks, under certain policies (i.e. London Lorry control scheme35) may not be an effective measure for controlling NO_2, especially in residential areas, given that if the traffic of heavy lorries and trucks is concentrated at night times, its effect will still appear in the morning peak hours when the majority of people are travelling.

Finally, there is still less than one percent of electric cars in London compared to petrol cars, implying that their positive effect on reducing NO_2 levels is likely to be negligible when compared to the entire number of existing petrol and diesel automobile flows. Furthermore, there are still a small number of electric trucks and buses, which we believe, along with stronger steps to restrict emissions from industrial zones, are more likely to cut NO_2 levels in London.

Transforming CCTV cameras into NO_2 sensors with a graph-to-graph neural network

Building on our understanding of the complex spatiotemporal dynamics of NO_2 levels, we are faced with the challenge of deducing these levels from the complex and nonlinear interactions among various variables. To address this, we developed a Graph-to-Graph deep model using deep learning[44,45], specifically geometric deep learning[46–50], to learn the presented spatiotemporal links and other latent ones that could contribute to the level of NO_2 at a given location while accounting for the dynamics of the entire network, traffic flows in London, and fluid dynamics derived from wind direction and speed. Figure 5A shows the overall conceptual framework of the developed pipeline to forecast NO_2 in London using hourly traffic modal flows in London. The introduced framework also integrates additional secondary data such as weather conditions and spatial features, among other variables (See Fig. S1 in supplementary). The developed model learns in semi-supervised settings from both the states of a given node represented in terms of traffic flows for each mode and the links between nodes represented in their adjacency and their potential influence elsewhere.

Given that the positions of both cameras and environmental sensors are not constrained to one another (as previously shown in Fig. 1B), the stated problem shifts from identifying regressor values on the same graph to generating a whole graph of a different adjacency matrix than the one given as an input. It is important to note that we used a weighted graph in which fewer links for traffic modes are identified based on the number of nearest neighbours to mimic the actual spatial network, whereas, for the graph of environmental sensors, we used a fully-connected network because air can diffuse freely from one location to another without the spatial constraints of a given network. The model was able to learn to create spatially distributed NO_2 values, resulting in a surface of NO_2 concentration over London at a given hour, using the described method (See Fig. 5). We also trained several models to assess our method (refer to the methodology section and Table S5).

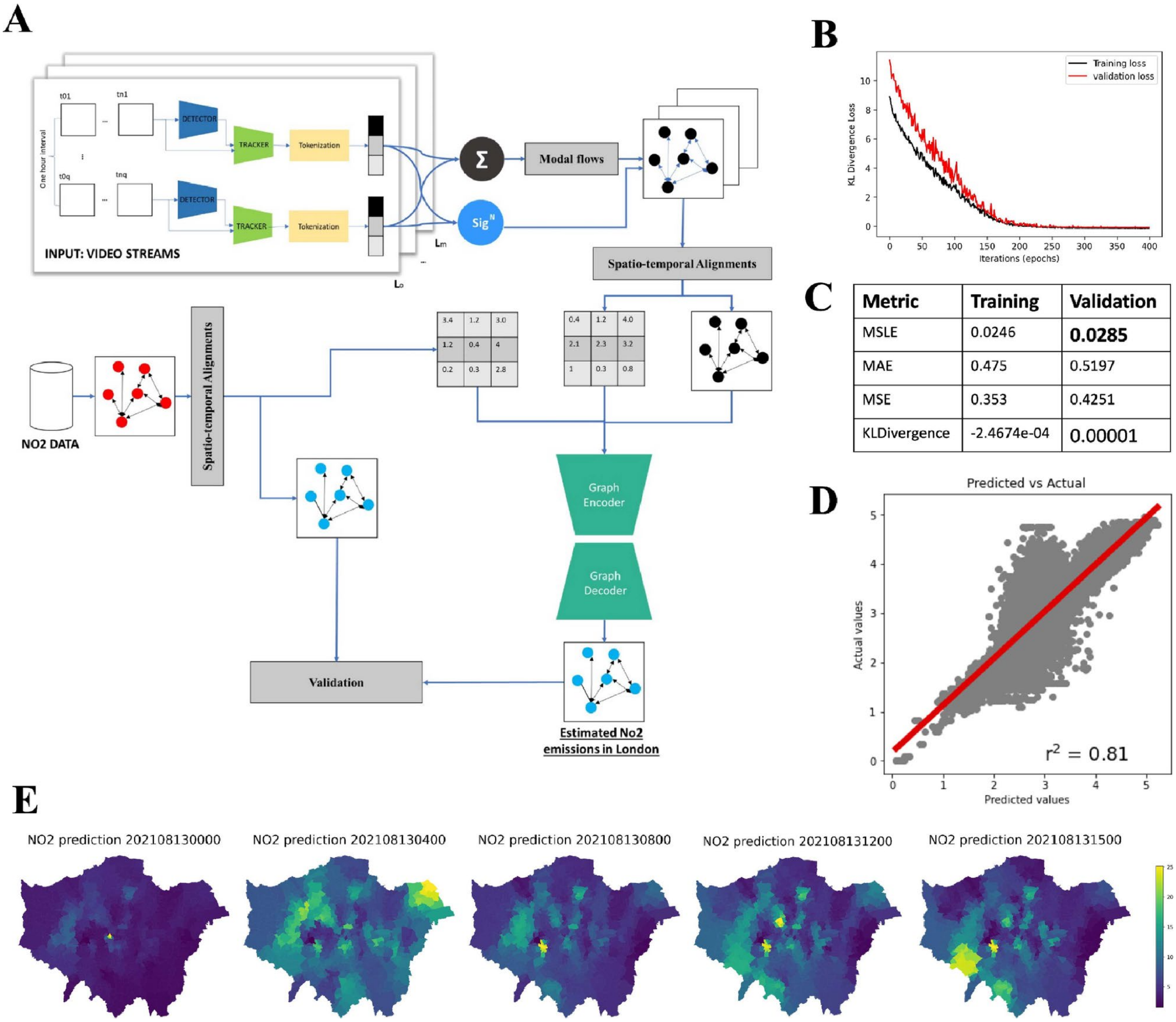
Discussion

Monitoring the dynamics of the environment and tracking the progress of environmental policies remains a difficult but critical issue in achieving urban sustainability. In this study, we demonstrated how CCTV cameras and autonomous vision systems using artificial intelligence can aid in monitoring NO_2 levels and evaluating our daily activities in cities that are substantially linked to different NO_2 levels. We demonstrated how human behaviours related to urban mobility and choice of mobility mode can influence the level of NO_2 differently depending on the dynamics of location and time. We presented novel analyses and insights into the multifaceted nature of the stated issue, such as the impact of time, location, natural and built environments, and urban policies. We demonstrated how CCTV cameras and additional spatial data can be utilised to infer NO_2 levels at the city scale when environmental sensors are unavailable or have sparse coverage when they exit. This technology could benefit numerous cities around the world that lack the infrastructure to monitor pollutants.

Based on this research, various learning lessons and policy implications can be applied to London and other cities across the globe. When it comes to decreasing emissions in cities, the majority of urban policies rely on (1) locational restrictions, (2) temporal constraints, or (3) a combination of temporal and locational constraints. Our findings suggest an alternative approach for developing environmental legislation that considers overall emissions across all locations and times of day. We demonstrated that there are temporal lags between current traffic and their impact on future NO_2 emissions. This implies the need for new policy reform that considers a minimal overall emission during different hours of the day rather than temporal constraints and concentrating unwanted traffic at a given time of the day. Given that our findings suggest that if trucks, for example, only drive at night within the inner parts of the city, their impact on emissions will be more likely to appear in the morning (with a lag of up to 6 h), where more people may be affected.

Limitations

There are still data uncertainties in big data, particularly video streams, making the presented traffic counts an approximation of day-to-day operations in Greater London. These uncertainties stem from factors such as camera field of view, obstruction, or biases due to the chosen locations for sensors[51]. Effective sensor deployment is essential for urban-scale monitoring to ensure comprehensive coverage and reliable data collection[28]; however, this study assumes both cameras and NO_2 sensors are provided and does not cover sensor placement. The placement of CCTV cameras can introduce biases into our NO_2 predictions. Cameras are typically located in high-traffic areas, which may not fully represent overall urban air quality. We have discussed this limitation and the measures taken to mitigate its impact. As a result, we considered numerous strategies such as recognising outliers and data stationary wherever it is acceptable for a certain method. Furthermore, many features derived from data tend to follow rational thinking of patterns that are predicted to be shown, according to descriptive



3.4	1.2	3.0
1.2	0.4	4
0.2	0.3	2.8

0.4	1.2	4.0
2.1	2.3	3.2
1	0.3	0.8

Metric	Training	Validation
MSLE	0.0246	**0.0285**
MAE	0.475	0.5197
MSE	0.353	0.4251
KLDivergence	-2.4674e-04	0.00001

Fig. 5. Graph-to-Graph model to predict NO_2 surface at a given time from camera feeds. (**A**) The overall method for the developed a signature-based graph neural network to generate a surface of NO_2 ground-level from camera inputs. The arrows represent the flow of information from the CCTV footage to the prediction of the NO_2 ground-level. (**B**) A scatter plot for the actual and predicted data for all sensor locations and all dates. (**C**) The results of training and validation loss and evaluation metrics for training and validation sets. (**D**) A scatter plot for the actual and predicted data for all sensor locations and all dates. (**E**) NO_2 prediction for different hours of the day, aggregated at a borough level. This figure is created by the first author using python programming.

analysis. For example, cars contribute to traffic congestion but not bicycles, the two traffic peaks of a given day when the total flow is distributed throughout all hours of a given day, and the negative relationship between cycling and the level of NO_2, among other things.

While the presented models require minimal inference time (< 0.1 s) to generate NO_2 at a given hour, it is critical to understand the centralised computational requirements for computing and extracting traffic flow data from CCTV video feeds at scale. The supplied data across all cameras and all days were retrieved using 84 days of computing on a single GPU. Accordingly, finding alternative solutions to minimise the time for deployment at a scale of a given city is necessary. Two approaches can be used to do this: (1) learning the complete traffic flow at a city level for a given time from only fewer camera inputs, and (2) decentralised computations at the edge by relying on AI-enabled cameras that deploy lightweight models on minimal hardware sensors. This method might enable real-time NO_2 data processing and inference, as well as proactive sensing of its determinants at any given time and place.

Methods

Our study enhances CCTV-based analysis and NO_2 monitoring by demonstrating the use of existing infrastructure for environmental sensing, which is especially beneficial for cities with limited access to specialised air quality sensors. Our method can be implemented in other cities with a sufficient number of CCTV cameras. For city-wide NO_2 prediction, our model utilises traffic data extracted from cameras, along with environmental and locational factors, and the computed signature of this data to predict city-wide NO_2 levels. The camera and NO_2 sensor locations do not have to coincide, providing flexibility in applying and transferring this method to any location. The input data comprises traffic data extracted from CCTV camera footage, including various road users' modal flows and their stationary statuses. Additionally, we included environmental factors such as average wind speed, wind direction, wet hours, sun hours, rainfall, average pressure, average humidity, average temperature, and proximity to industrial zones. The ground truth data for training and validating our models were sourced from hourly NO_2 sensor measurements across multiple locations within London. The target features for our models were the NO_2 levels, either at specific sensor locations or across a generated surface for city-wide prediction. By integrating the computed signature of the traffic data with locational and environmental features, our models provided accurate predictions of NO_2 levels, demonstrating the feasibility of using existing CCTV infrastructure for environmental monitoring and policy-making. Here, we describe the materials and methods utilised to develop this research.

Here we describe our materials and the different methods utilised to develop this research.

Materials

All raw data sources can be accessed online.

1. *London CCTV data* We collected video streams that represent 892 unique camera locations across London for 56 different hours of scattered days in the year 2021. This data includes 65,493,858 sequential frames, in which the total data or a subset of it has been used for different analyses represented in the paper. We also collected additional video data for a given camera (ID) for a given hour (12 am–1 pm) across all the days of the year to show the seasonal dynamics of traffic patterns. The data can be accessed via API permissions from Transport for London (TfL).
2. *Hourly* NO_2 *data* We extracted hourly N02 data of 144 unique sensors that link to the extracted video hours. The raw data can be accessed through an API from London Air: https://www.londonair.org.uk/london/asp/annualmaps.asp.
3. *Weather data* We linked the camera and NO_2 data to the weather day based on a day resolution. We included nine variables as a representation of the environmental conditions of a given day. This data, includes (1) average wind speed, (2) wind direction, (3) wet hours, (4) sun hours, (5) rainfalls, (6) average pressure, (7) average humidity, (8) average temperature, and (9) average feels like temperature. The raw data can be accessed from: http://nw3weather.co.uk/wxdataday.php?vartype=wmean&year=2021.
4. *Spatial data* We used GIS shapefile data for the spatial representations of London's boroughs, spatial network, and the boundary of the city. The spatial network data included (1) whether a given street is two-directional, (2) average speed and (3) the type of the street. The raw data can be accessed from Greater London Authority: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london.
5. *Car flows based on engine types* To evaluate the percentage of electric cars to petrol and diesel ones in each borough, we used the traffic flow data provided by London Council. This data is used for statistical analysis to account for the ratio of cars based on the engine types that we observe in CCTV cameras at a given location. The data is entitled: "laei-2019-major-roads-vkm-flows-speeds" and can be accessed from: https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory-laei-2019.
6. *Proximity to industrial zones* We used Strategic Industrial Location Points to calculate a buffer zone of 1 mile and account for the camera's locations that are within this zone. The raw dataset can be accessed online from: https://data.london.gov.uk/dataset/strategic-industrial-location-points-london-plan-consultation-2009.

Extracting road users from video streams

To extract the six types of road users from video streams and their relevant information, we used a deep learning framework that comprises multiple deep models including, You Look Only Once (YOLO) architecture[52,53]. Particularly, we relied on YoloV5m[54] coupled with DeepSort architecture[55] to detect and track road users throughout a given video file. DeepSort architecture is built on a deep learning model with Sort algorithms[56] to account for object occlusion. We used a pre-trained weight of YOLOV5m model trained on COCO dataset[57]. It's worth mentioning that computing this data and transforming it from raw video streams to vector data took almost 18 h for analysing one hour across all cameras for a given day (84 days in total) on a single GPU.

Projecting road users in a bird's eye view map

Transforming moving objects from CCTV footage to a top-view perspective is crucial for accurately analysing and verifying various traffic factors. This perspective allows for the consistent identification and tracking of road users, regardless of obstructions within the camera's field of view. By projecting the traffic data onto a bird's eye view, we can effectively distinguish between stationary and non-stationary road users, offering a clearer and more precise understanding of traffic dynamics. Additionally, this transformation ensures geographic consistency when integrating data with mapping services, enhancing the overall spatial accuracy of our traffic flow analyses. This step is integral to mitigating common issues associated with perspective distortion in street-level imagery, ensuring reliable data for predicting NO_2 levels.

We relied on the TopView framework to transform objects from the camera view to the bird's eye view without knowing the camera models that include both intrinsic and extrinsic parameters[58]. The framework relies on a deep learning model to detect the vanishing point (VP) in a given scene, whereas four points in the camera view can be automated and correspond to four points in world coordinates and accordingly to a bird's eye view map based on geometric transformation and homography[58–61]. We used the VP model and paired points in the two views to determine the homography matrix H as follows:

$$\begin{bmatrix} z_i x_i' \\ z_i y_i' \\ z_i \end{bmatrix} = H \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \tag{1}$$

where $\mathrm{dst}(i)$ = (x_i', y_i'), $\mathrm{src}(i)$ = (x_i, y_i), $i = 0, 1, 2, 3$.

Given that src and dst are the coordinates of the quadrangle vertices in the camera view and world coordinates respectively, (x_i, y_i) and (x_i', y_i') are the paired coordinate points in the camera and the bird's eye view planes respectively and H is the transformation of the homography matrix that is computed as:

$$H = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \tag{2}$$

Given that H is calibrated based on the four paired points that are produced by the camera and top-view planes, respectively. And therefore, the detected object in the camera plane may be changed into the top-view plane by resolving H. For further explanation, see the full explanation of the TopView method[62].

Tokenizing road users and counting flows

To detect modal flows, we first tracked road users in a given file, where each road user has a unique ID, and then the number of road users is counted throughout the file. The road users are vectorized based on their tracked ID data and visualised based on when they appear and disappear in the video files while keeping in mind that multi-dimensional data, such as stationary status, road user categories, and trajectory line in the bird's eye view, has been retrieved.

Ranks of traffic composition

We estimated the ranks of traffic composition by separating the total counts into unique values that indicate nodes (n = 1, n = 2, etc.) to grasp the collective behaviour of road users from the local site of all cameras to the city scale. Following that, we computed the unique patterns across each node value (i.e., in the case of n = 2, the possible scenarios are vehicle and person, car and car, etc.) and assigned a unique id to each unique pattern. Instead of summing the counts for each mode, we sum the structure at the city level, for example (1–1 + 2–2 + 3–1) up to the number of files.

Granger causality

Granger causality[63–66] is tested in the context of linear regression, and it is significant when the previous values of a given variable X_1 contribute to the forecasting of the current value of variable X_2 or vice versa. By considering a bivariate autoregressive model for these two variables:

$$X_1(t) = \sum_{j=1}^{p} A_{11,j} X_1(t-j) + \sum_{j=1}^{p} A_{12,j} X_2(t-j) + \varepsilon_1(t) \tag{3}$$

$$X_2(t) = \sum_{j=1}^{p} A_{21,j} X_1(t-j) + \sum_{j=1}^{p} A_{22,j} X_2(t-j) + \varepsilon_2(t) \tag{4}$$

given that p represents the number of lagged observations in the model order. The matrix A comprises the coefficients of the model such as the contributions of each lagged observation to the predicted values of $X_1(t)$ and $X_2(t)$, and ε_1 and ε_2 are the model residuals for each time series.

If the coefficients in A_{12} are all considerably different from zero, then $X_2(t)$ Granger causes $X_1(t)$. The model significance is tested by computing an F-test of the null hypothesis that $A_{12} = 0$, assuming that the stationarity of the covariance on $X_1(t)$ and $X_2(t)$. The logarithm of the associated F-statistic can be used to determine the size of a Granger causality interaction[67,68].

According to the Granger test, it is worth mentioning that causality is evaluated on the grounds that (1) the cause precedes the effect and (2) the cause has specific knowledge about the potential outcomes of its impact. To demonstrate the significant findings of Granger testing, we show the results of four parameters, including the parameters for the F-test and ssr-F-test which are based on the F-distribution and the parameters for the ssr-based chi-squared test and the likelihood ratio test, which are based on the chi-square distribution.

Spatial weight

Using the K-Nearest Neighbour weights technique[69], we estimated the spatial weight matrix (ω_{ij_t}) between the various camera sites at a particular time (t). It is a set of neighbours defined by distance-based weights based on (K) observations. We investigated several (K) values and found that 10 was the best approximation of the number of neighbours where the different camera locations closely matched the actual spatial network. We computed a dynamic spatial weight that differs based on the point representation of a given time. We utilised the estimated spatial weight in many analyses, including spatial clustering, the spatial regression model, and the Graph model.

Spatial clustering and outliers detection

We computed statistically significant spatial clusters and hot-spot analysis based on Local Moran's[70,71]. If the value of I is positive, it means that a feature is part of a cluster and that it is surrounded by other features that have similar attributes that are either high or low. A negative value for I implies that an outlier feature has nearby features with values that differ from its own. For the cluster or outlier to be regarded as statistically significant, the p value for the feature must be low enough in both cases.

$$I_i = \frac{\left(x_i - \bar{X}\right)}{S_i^2} \sum_{j=1, j\neq i}^{n} \omega_{ij} \left(x_j - \bar{X}\right) \tag{5}$$

Given that x_i is the attribute for feature i, $\bar{X}$ is the mean for the corresponding attribute, ω_{ij} is the spatial weight between feature i and j.

$$S_i^2 = \frac{\sum_{j=1, j\neq i}^{n} \left(x_j - \bar{X}\right)^2}{n-1} \tag{6}$$

Given that n is the total number of features.

The Z-score for the statistics is defined as:

$$Z_{I_i} = \frac{I_i - E\left[I_i\right]}{\sqrt{V\left[I_i\right]}} \tag{7}$$

$$E\left[I_i\right] = -\frac{\sum_{j=1, j\neq i}^{n} \omega_{ij}}{n-1} \tag{8}$$

$$V\left[I_i\right] = E\left[I_i^2\right] - E\left[I_i\right]^2 \tag{9}$$

Spatial regression model

Given the geographical dependency of the observed variables, we employed a spatial regression model[72,73] rather than a simple regression model to assess the statistically significant links between NO_2 levels and the various values of road users and the built environment. We explored three different approaches in which spatial weight can be applied including, the spatial dependency model, spatial error model, and spatial lag model. First, in the spatial dependency model, the previously computed spatial weight ω_{ij} is accounted in the model as an additional independent variable as follows:

$$\log\left(P_i\right) = \alpha + X\beta + WX\gamma + \varepsilon \tag{10}$$

$$\log\left(P_i\right) = \alpha + \sum_{k=1}^{p} X_{ij}\beta_j + \sum_{k=1}^{p} \left(\sum_{j=1}^{N} \omega_{ij} x_{jk}\right) \gamma_k + \varepsilon_i \tag{11}$$

Second, in the spatial error model, we account for the spatial dependence in the model residual as follows:

$$\log\left(P_i\right) = \alpha + \sum_{k=1}^{p} X_{ki}\beta_k + \mu_i \tag{12}$$

$$\mu_i = \lambda_{ulag-i} + \varepsilon_i \tag{13}$$

$$\lambda_{ulag-i} = \sum_{j} \omega_{ij} u_j \tag{14}$$

Last, the Spatial lag model can be computed as:

$$\log\left(P_i\right) = \alpha + \rho \log\left(P_{lag-i}\right) + \sum_{k=1}^{p} X_{ki}\beta_k + \varepsilon_i \tag{15}$$

NO_2 surface construction from points

We also relied on the triangulation method to generate a 3D surface from the sensors' unique locations by creating triangles by specifying their corners based on three given points.

Signature of paths

This research is concerned with multi-level temporal scales that go from the temporal representation of a certain sequence of a video file at a given location to the hourly temporal representation of video files that can correspond to the temporal scale of NO_2 Data. As a result, in addition to depending on a straightforward strategy of summing the data increments of a given hour at a specific site, we relied on rough path theory and path signature[35,36,74–76] to summarise the multidimensional temporal representation of the presented data. As a result, we developed a method for summarising the key patterns within the video increments of an hour without losing the raw data relying on signature due to its invariance to reparameterisations. The truncated signature of a path γ_t at a given depth N at a given hour is defined as:

$$S_{a,b}\left(\gamma_t\right) = \bigoplus_{n=0}^{N} S^{n}_{a,b}(\gamma), \quad \text{given that} \quad S^{n}_{a,b}\left(\gamma_t\right) = \frac{1}{n!}\left(\gamma_b - \gamma_a\right)^{\otimes n} \tag{16}$$

The signature transform given that $\mathrm{Sig}^N = S(\mathbb{R}^d) \rightarrow \prod_{n=1}^{N} (\mathbb{R}^d)^{\otimes n}$ is computed as:

$$\mathrm{Sig}^N(X) = \left(\int_{0<t_1<\cdots<t_n<1} \frac{df}{dt}\left(t_1\right) \otimes \cdots \otimes \frac{df}{dt}\left(t_n\right)\, dt_1 \ldots dt_n \right)_{1 \leq n \leq N} \tag{17}$$

$$\text{for } 1 \leq n \leq N \tag{18}$$

The log signature of γ_t is defined as:

$$\log S_{a,b}\left(\gamma_t\right) = \bigoplus_{n=0}^{N} \frac{(-1)^{n-1}}{n} \left(\hat{S}^{n}_{a,b}(\gamma)\right)^{\otimes n}, \tag{19}$$

$$\text{given that } S^{0}_{a,b}(\gamma_t) = 1 \text{ and } \hat{S}_{a,b}\left(\gamma_t\right) = \bigoplus_{n=1}^{N} S^{n}_{a,b}(\gamma_t) \tag{20}$$

Graph model architectures

We developed an undirected weighted Graph $G(V, E, A, \omega)$, where V is the set of nodes with $|V| = N$ is the number of nodes, E represents the set of the edges of the graph, A is the adjacency matrix and is an $N \times N$ sparse matrix, and ω_{ij} represents the adjacency matrix between node v_i and v_j. A graph signal $f : V \rightarrow \mathbb{R}$ represents a function defined on the vertices of a graph G which maps every vertex $v_{i\,i=1,\ldots,N}$ to a real number f_i. The graph signal f can be projected to the eigenvectors of the Laplacian matrix L and by assuming that λ_l and μ_l are the l_{th} eigenvalue and eigenvector of the Laplacian matrix L, the graph Fourier transform $\hat{f}$ of the graph signal can be defined as:

$$GF[f]\left(\lambda_l\right) = \hat{f}\left(\lambda_l\right) = \langle f, \mu_l \rangle = \sum_{i=1}^{N} f(i)\mu_l^{*}(i), \quad \text{given that } \mu_l^{*} = \mu_l^{T} \tag{21}$$

In the context of graph[45,47,49], the convolution operation between two functions f and g can be applied by relying on graph Laplacian eigenvectors and can be defined as:

$$(f * g) = IGF[GF[f] \cdot GF[g]], \quad (f * g)(i) = \sum_{l=0}^{N-1} \hat{f}\left(\lambda_l\right) \hat{g}\left(\lambda_l\right) \mu_l(i) \tag{22}$$

The Graph model comprises Lth graph convolution layers, in which each layer constructs an embedding for each node by fusing the embeddings of the neighbours of a given node from the previous layer as follows:

$$Z^{(l+1)} = A' X^{(l)} W^{(l)}, \quad X^{(l+1)} = \sigma\left(Z^{(l+1)}\right) \tag{23}$$

given that $X^{(l)} \in \mathbb{R}^{N \times F_l}$ represents the embedding of the l-th layer for all N nodes, $X^{(0)} = X$, A' is the weighted and normalized adjacency matrix, $W^{(l)} \in \mathbb{R}^{F_l \times F_{l+1}}$ is the feature transformation matrix that will be learned, and $\sigma(\cdot)$ is the activation function for which we implemented an element-wise ReLU.

We also used a Graph Attention layer[46], given an input of a set of node features $\mathbf{h} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N\}$, $\mathbf{h}_i \in \mathbb{R}^F$ where F is the number of features in each node. The layer outputs a new set of node features $\mathbf{F}'$, $\{\mathbf{h}'_1, \mathbf{h}'_2, \ldots, \mathbf{h}'_N\}$, $\mathbf{h}'_i \in \mathbb{R}^{F'}$. The linear transformation of the layer is applied to each node, parameterised by a weight matrix, $W \in \mathbb{R}^{F' \times F}$, in which a shared attentional mechanism is performed on the nodes to indicate the importance of features in a given node j to node i. Their attention coefficients are defined as:

$$e_{ij} = a\left(\mathbf{W}\mathbf{h}_i, \mathbf{W}\mathbf{h}'_j\right) \tag{24}$$

The attention mechanism a can be defined as a single feedforward layer, parametrized by a weight vector $\mathbf{a} \in \mathbb{R}^{2F}$, activated by a LeakyReLU nonlinearity, its coefficients can be defined as:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}'_j\right]\right)\right)}{\sum_{k\in N_i}\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}'_k\right]\right)\right)} \tag{25}$$

Given that $\|$ represents the concatenation operation and $.T$ represents transposition.

We have experimented with both graph layers, in which we have trained multiple models for two different tasks that take different inputs and generate different outputs, as follows:

Task 1: estimating the NO_2 surface for a given hour from traffic flows data

This model takes an input X where $X \in \mathbb{R}^{H\times N\times C}$ and generates NO_2 levels across London for a given hour (Y) as $Y \in \mathbb{R}^{H\times M}$, where H is the number of unique hours, N is the number of cameras' locations, C is the number of features, including modal flows and locational urban features, and M is the number of NO_2 sensors' locations where $M \neq N$.

Task 2: estimating NO_2 at a given location from the graph knowledge of traffic flows

This model takes an input X where $X \in \mathbb{R}^{H\times N\times C}$ and generates NO_2 concentration at a given location for a given hour (Y) as $Y \in \mathbb{R}^{H}$, where H is the number of unique hours, N is the number of cameras' locations, C is the number of features, including modal flows and locational urban features.

Further results for all models and their hyperparameters are provided in supplementary, in Table S5.

Training objective loss

We trained our models based on mean squared logarithmic error (MSLE), defined as:

$$\text{Loss} = (\log(x+1) - \log(y+1))^2 \tag{26}$$

Given that x and y are the true and predicted values of NO_2 levels of a given location at a given hour.

NO_2 model validation metrics

Furthermore, we computed different metrics to compare the results of the trained models and to validate their performances. We calculated Kullback–Leibler divergence, or known as relative entropy denoted as $D_{KL}(P\|Q)$ and is defined as:

$$D_{KL}(P\|Q) = \sum_{x\in X} P(x)\log\left(\frac{P(x)}{Q(x)}\right) \tag{27}$$

given that P and Q are two discrete probabilities distributions on the same sample space X representing the distributions of true values and predicted ones. Second, we computed mean absolute error (MAE), known as L1 loss, and is defined as:

$$L_1(x,y) = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{28}$$

given that y_i, x_i are the predicted and true values of NO_2 levels respectively, and n is the batch size.

Training setup and implementation details

We report on 20 models with different hyperparameters and architecture (See Table S5 in supplementary). All models are trained based on the input of the normalized numerical values of traffic flows and categorical values of all factors explained previously after being factorised and transformed into dummies. However, they vary, in terms of input, based on whether (1) the computed signature is included as an input, (2) the adjacency matrix of the NO_2 sensor data is included, besides the adjacency matrix of the CCTV cameras and (3) the number of nearest neighbours when computing the edge or the adjacency matrix. To account for the current state-of-the-art baselines, we trained different architectures as follows:

Graph attention model

We trained several models based on the architecture of three graph attention layers, in which each layer comprises 6 attention heads and each computing 907 features, followed by an ELU nonlinearity layer. The final layer is used to output NO_2 values, containing 1 feature (in case of inferring a NO_2 value for a single location) or N features based on the number of NO_2 sensors (In case of inferring spatially distributed NO_2 values or inferring traffic flows in N cameras), followed by activation of a logistic sigmoid function. We applied dropout[75,77] within the three-layer blocks to avoid over-fitness. We trained the models based on a batch size of 8 graphs for 100 training cycles (epochs). All models are initialized by Glorot initialization[78] and trained to minimise the introduced loss function based on Adam stochastic gradient descent optimiser[79], with an initial learning rate of 0.01 and an early stopping strategy based on the validation loss, with patience of 20 epochs.

Graph convolution model

Similar to the graph attention model we trained a graph model based on three graph convolution layers instead of the graph attention model. We followed a close implementation of the originally introduced method and best practice guidelines to provide a baseline[45]. All models based on graph convolution are trained based on hidden units of 50 features and a dropout of 0.5. The models are trained based on a batch size of 64 using Adam optimiser, with an initial learning rate of 0.01 and an early stopping strategy based on the validation loss, with a patience of 20 epochs.

Multi-branch graph model

This model architecture takes six inputs, including camera nodes, camera edge, categorical feature, numerical features, and environmental sensor adjacency matrix (or five inputs without signature). Each input is encoded through an isolated branch of three 1D Convolutional layers of 32 filters, kernel size of 1 and activated with a ReLU function followed by a Dropout layer of size 0.4. Finally, a Flatten layer and a fully connected layer of 50 features are used. After each encoder, all outputs are concatenated and passed to a Fully connected layer and a final output of N features that is equivalent to the number of nodes in The NO_2 surface for a given hour, activated based on the Softplus function. The model is trained with a batch size of 2 graphs, and for 300 epochs, following similar procedures of the previous architectures.

Transformer model

We also trained several models based on transformer architecture without an explicit graph structure like the case in the first graph architectures. We replaced the convolutional layer in the introduced architecture of the multi-branch graph model, with three transformer layers. Each transformer layer comprises 6 attention heads and projection dimensions of 907 features, followed by a skip connection, a normalization layer, a Multi-layer Perceptron (MLP) and a second skip connection layer. Afterwards, we used a layer normalization and calculated attention weights, in which the product of both attention weights and the previous layer outputs are passed to a single fully connected layer. The final layer is used to output NO_2 values, containing 1 feature (in case of inferring a NO_2 value for a single location) or N features based on the number of NO_2 sensors (In case of inferring spatially-distributed NO_2 values or inferring traffic flows in N cameras), followed by activation of a Softplus function. We also applied dropout to avoid over-fitness. We trained the models based on a batch size of 2 for 300 epochs. We used AdamW stochastic gradient descent optimiser to minimise the introduced loss function, with an initial learning rate of 0.001 and an early stopping strategy based on the validation loss, with a patience of 20 epochs.

Evaluating models under different environmental conditions

We trained various model architectures with different hyperparameters to create a baseline and validate our method using different evaluation metrics (see Table S5). We conducted an error analysis to assess model performance under various weather conditions. This involved analysing the impact of factors like rain, wind speed, and temperature on NO_2 levels, providing insights into the robustness of our models. Additionally, we evaluated model performance over different time periods, such as hourly, daily, weekly, and monthly intervals, to ensure consistency. We also assessed the models at different locations within the study area to account for spatial variability in NO_2 levels. Through these thorough evaluations, we aim to demonstrate the reliability and accuracy of our models in predicting NO_2 levels under various real-world conditions. In our study, several models showed promising results. For example, the Graph Convolutional Model with Signature (Model ID 1) exhibited good performance with a mean squared logarithmic error (MSLE) of 0.0375 and a mean absolute error (MAE) of 0.6558. This model integrates graph convolution operations, which are effective in capturing spatial dependencies in the data. The Attention-based Graph Model without Signature (Model ID 3) introduces attention mechanisms within the graph neural network framework. Although this model has significantly more parameters (120,342,324) and longer training time, it presented robust results with an MSLE of 0.0454 and an MAE of 0.6842. The attention mechanism helps in focusing on the most relevant parts of the graph, providing better feature representation. At City wide prediction, the Conv1D-based multiple branch model with Signature (Model ID 19) demonstrated strong performance, providing accurate predictions and showing a high correlation with actual NO_2 levels. By incorporating signature information (N = 3), the model enhances its predictive accuracy. The multi-branch design allows the model to process various data aspects in parallel, boosting its learning capacity.

Data availability

All raw data sources are listed in the Materials and Methods section.

Received: 19 September 2024; Accepted: 13 January 2025
Published online: 29 January 2025

References

1. Sun, L., Chen, J., Li, Q. & Huang, D. Dramatic uneven urbanization of large cities throughout the world in recent decades. *Nat. Commun.* **11**, 5366 (2020).
2. Coutrot, A. Entropy of city street networks linked to future spatial navigation ability. *Nature* **604**, 104–110 (2022).
3. Anza-Ramirez, C. The urban built environment and adult BMI, obesity, and diabetes in Latin American cities. *Nat. Commun.* **13**, 7977 (2022).
4. Badland, H. M. Association of neighbourhood residence and preferences with the built environment, work-related travel behaviours, and health implications for employed adults: Findings from the urban study. *Soc. Sci. Med.* **75**, 1469–1476 (2012).
5. Lee, K. O., Mai, K. M. & Park, S. Green space accessibility helps buffer declined mental health during the Covid-19 pandemic: Evidence from big data in the United Kingdom. *Nat. Ment. Health* **1**, 124–134 (2023).
6. Beelen, R. Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study). *Environ. Health Perspect.* **116**, 196–202 (2008).

7. Vert, C. Effect of long-term exposure to air pollution on anxiety and depression in adults: A cross-sectional study. *Int. J. Hyg. Environ. Health* **220**, 1074–1080 (2017).
8. Morales-Suárez-Varela, M., Peraita-Costa, I. & Llopis-González, A. Systematic review of the association between particulate matter exposure and autism spectrum disorders. *Environ. Res.* **153**, 150–160 (2017).
9. Roberts, S. Exploration of NO$_2$ and PM$_{2.5}$ air pollution and mental health problems using high-resolution data in London-based children from a UK longitudinal cohort study. *Psychiatry Res.* **272**, 8–17 (2019).
10. Antonsen, S. Exposure to air pollution during childhood and risk of developing schizophrenia: A national cohort study. *Lancet Planet. Health* **4**, 64–73 (2020).
11. Ji, J. S. Air pollution and cardiovascular disease onset: Hours, days, or years?. *Lancet Public Health* **7**, 890–891 (2022).
12. Hu, Y., Ji, J. S. & Zhao, B. Restrictions on indoor and outdoor NO$_2$ emissions to reduce disease burden for pediatric asthma in China: A modeling study. *Lancet Reg. Health West. Pac.* **24**, 100463 (2022).
13. Cooper, M. J. Global fine-scale changes in ambient NO$_2$ during Covid-19 lockdowns. *Nature* **601**, 380–387 (2022).
14. Reuter, M. Decreasing emissions of NO$_x$ relative to CO$_2$ in east Asia inferred from satellite observations. *Nat. Geosci.* **7**, 792–795 (2014).
15. Foy, B., Lu, Z. & Streets, D. G. Satellite NO$_2$ retrievals suggest China has exceeded its NO$_x$ reduction goals from the twelfth five-year plan. *Sci. Rep.* **6**, 35912 (2016).
16. Cuevas, C. A. Evolution of NO$_2$ levels in Spain from 1996 to 2012. *Sci. Rep.* **4**, 5887 (2015).
17. Wells, K. C. Satellite isoprene retrievals constrain emissions and atmospheric oxidation. *Nature* **585**, 225–233 (2020).
18. Stavrakou, T., Müller, J.-F., Bauwens, M., Boersma, K. F. & Geffen, J. Satellite evidence for changes in the NO$_2$ weekly cycle over large cities. *Sci. Rep.* **10**, 10066 (2020).
19. Shams, S. R., Jahani, A., Kalantary, S., Moeinaddini, M. & Khorasani, N. Artificial intelligence accuracy assessment in NO$_2$ concentration forecasting of metropolises air. *Sci. Rep.* **11**, 1805 (2021).
20. Laughner, J. L. & Cohen, R. C. Direct observation of changing NO$_x$ lifetime in North American cities. *Science* **366**, 723–727 (2019).
21. Beirle, S. Pinpointing nitrogen oxide emissions from space. *Sci. Adv.* **5**, 9800 (2019).
22. Badia, A. A take-home message from Covid-19 on urban air pollution reduction through mobility limitations and teleworking. *NPJ Urban Sustain.* **1**, 35 (2021).
23. Song, W. Important contributions of non-fossil fuel nitrogen oxides emissions. *Nat. Commun.* **12**, 243 (2021).
24. Grange, S. K., Lewis, A. C., Moller, S. J. & Carslaw, D. C. Lower vehicular primary emissions of NO$_2$ in Europe than assumed in policy projections. *Nat. Geosci.* **10**, 914–918 (2017).
25. Ibrahim, M. R., Haworth, J. & Cheng, T. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities* **96**, 102481 (2020).
26. Ma, C. et al. Vehicle-based machine vision approaches in intelligent connected system. *IEEE Trans. Intell. Transp. Syst.* **25**(3), 2827–2836. https://doi.org/10.1109/TITS.2023.3276325 (2024).
27. Ma, C., Song, J. & Xu, Y. E. A. Reducing environment exposure to Covid-19 by IoT sensing and computing with deep learning. *Neural Comput. Appl.* **35**, 25097–25106. https://doi.org/10.1007/s00521-023-08712-9 (2023).
28. Song, J. et al. Toward high-performance map-recovery of air pollution using machine learning. *ACS ES &T Eng.* **3**(1), 73–85. https://doi.org/10.1021/acsestengg.2c00248 (2022).
29. Siła-Nowicka, K. et al. Analysis of human mobility patterns from GPS trajectories and contextual information. *Int. J. Geograph. Inf. Sci.* **30**(5), 881–906 (2016).
30. Alessandretti, L., Aslak, U. & Lehmann, S. The scales of human mobility. *Nature* **587**(7834), 402–407 (2020).
31. Kraemer, M. U. et al. Mapping global variation in human mobility. *Nat. Hum. Behav.* **4**(8), 800–810 (2020).
32. Gately, C. K., Hutyra, L. R., Peterson, S. & Wing, I. S. Urban emissions hotspots: Quantifying vehicle congestion and air pollution using mobile phone GPS data. *Environ. Pollut.* **229**, 496–504 (2017).
33. TfL London cameras (2021)
34. Lyons, T. J., Caruana, M. & Lévy, T. *Differential Equations Driven by Rough Paths: École d'Été de Probabilités de Saint-Flour XXXIV-2004*, 1st edn. Lecture Notes in Mathematics, vol. 1908 (Springer, 2007). https://doi.org/10.1007/978-3-540-45886-2 . Part of the book sub series: École d'Été de Probabilités de Saint-Flour. https://doi.org/10.1007/978-3-540-45886-2
35. Lyons, T. *Rough Paths, Signatures and the Modelling of Functions on Streams*. Accessed: 2023-02-12 (2014).
36. Lyons, T. & Qian, Z. *System Control and Rough Paths* (Oxford University Press, 2002). https://doi.org/10.1093/acprof:oso/9780198506485.001.0001.
37. Shah, V. et al. Effect of changing NO$_x$ lifetime on the seasonality and long-term trends of satellite-observed tropospheric NO$_2$ columns over China. *Atmos. Chem. Phys.* **20**(3), 1483–1495 (2020).
38. Matsumi, Y. et al. High-sensitivity instrument for measuring atmospheric NO$_2$. *Anal. Chem.* **73**(22), 5485–5493 (2001).
39. Crutzen, P. J. The role of no and NO$_2$ in the chemistry of the troposphere and stratosphere. *Annu. Rev. Earth Planet. Sci.* **7**(1), 443–472 (1979).
40. Richter, A. et al. Satellite measurements of NO$_2$ from international shipping emissions. *Geophys. Res. Lett.* **31**(23), 1–4 (2004).
41. Mentel, T. F., Bleilebens, D. & Wahner, A. A study of nighttime nitrogen oxide oxidation in a large reaction chamber-the fate of NO$_2$, N$_2$O$_5$, HNO$_3$, and O$_3$ at different humidities. *Atmos. Environ.* **30**(23), 4007–4020 (1996).
42. Liu, F. et al. NO$_x$ lifetimes and emissions of cities and power plants in polluted background estimated by satellite observations. *Atmos. Chem. Phys.* **16**(8), 5283–5298 (2016).
43. Kelly, F. et al. *The London Low Emission Zone Baseline Study* (Health Effects Institute, 2011).
44. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (The MIT Press, 2017).
45. Zhang, S., Tong, H., Xu, J. & Maciejewski, R. Graph convolutional networks: A comprehensive review. *Comput. Soc. Netw.* **6**, 11 (2019).
46. Veliçković, P., et al. *Graph Attention Networks*. Accessed: 2023-02-12 (2018)
47. Hamilton, W., Ying, Z. & Leskovec, J. *Inductive Representation Learning on Large Graphs* (Published date unknown)
48. Gori, M., Monfardini, G. & Scarselli, F. A new model for learning in graph domains. In *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks*, 729–734 (IEEE, 2005).
49. Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A. & Vandergheynst, P. Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Process. Mag.* **34**, 18–42 (2017).
50. Robinson, C., Franklin, R. S. & Roberts, J. Optimizing for equity: Sensor coverage, networks, and the responsive city. *Ann. Am. Assoc. Geograph.* **112**, 2152–2173 (2022).
51. Redmon, J. & Farhadi, A. YOLO9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517–6525 (IEEE, 2017).
52. Redmon, J. & Farhadi, A. *YOLOv3: An Incremental Improvement* (2018).
53. Ultralytics: YOLOv5 (2021).
54. nwojke: DeepSort (2019).
55. Bewley, A., Ge, Z., Ott, L., Ramos, F. & Upcroft, B. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, 3464–3468 (IEEE,2016).
56. Lin, T.-Y. Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014*, 740–755 (Springer, 2014).
57. Zhang, Z. Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, 666–673 (IEEE, 1999).

58. Ibrahim, M. R. TopView: Vectorising road users in a bird's eye view from uncalibrated street-level imagery with deep learning. arXiv:2412.16229 [cs.CV]. Submitted on 18 Dec 2024. 28 pages (2024). https://doi.org/10.48550/arXiv.2412.16229
59. Venkatesh, M. & Vijayakumar, P. A simple bird's eye view transformation technique. *Int. J. Sci. Eng. Res.* **3**, 4 (2012).
60. Mardiati, R., Mulyana, E., Maryono, I., Usman, K. & Priatna, T. The derivation of matrix transformation from pixel coordinates to real-world coordinates for vehicle trajectory tracking. In *2019 IEEE 5th International Conference on Wireless and Telematics (ICWT)*, 1–5 (IEEE, 2019).
61. Escalera, A. & Armingol, J. M. Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration. *Sensors* **10**, 2027–2044 (2010).
62. White, H. & Lu, X. Granger causality and dynamic structural systems. *J. Financ. Econom.* **8**, 193–243 (2010).
63. Bahadori, M. T. & Liu, Y. Granger causality analysis in irregular time series. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, 660–671 (Society for Industrial and Applied Mathematics, 2012).
64. Eichler, M. In *Causal Inference in Time Series Analysis* (eds Berzuini, C. et al.) 327–354 (Wiley, 2012).
65. Sugihara, G. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
66. Geweke, J. Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* **77**, 304–313 (1982).
67. Geweke, J., Meese, R. & Dent, W. Comparing alternative tests of causality in temporal systems. *J. Econom.* **21**, 161–194 (1983).
68. Dudani, S. A. The distance-weighted k-nearest-neighbor rule. *IEEE Trans. Syst. Man Cybern. SMC* **6**, 325–327 (1976).
69. Anselin, L. Local indicators of spatial association-LISA. *Geograph. Anal.* **27**, 93–115 (1995).
70. Getis, A. & Ord, J. K. The analysis of spatial association by use of distance statistics. *Geograph. Anal.* **24**, 189–206 (2010).
71. Florax, R. J. G. M. & Nijkamp, P. In *Misspecification in Linear Spatial Regression Models* (ed. Kempf-Leonard, K.) 695–707 (Elsevier, 2005).
72. Rey, S. J. *Mathematical Models in Geography* 9393–9399 (Pergamon, 2001).
73. Chevyrev, I. & Kormilitzin, A. *A Primer on the Signature Method in Machine Learning.* Accessed: 2023-02-12 (2016).
74. The London Lorry Control Scheme (LLCS) (1985).
75. Dahl, G. E., Sainath, T. N. & Hinton, G. E. Improving deep neural networks for LVCSR using rectified linear units and dropout. In *2013 IEEE International Conference On Acoustics, Speech and Signal Processing (ICASSP)*, 8609–8613 (IEEE, 2013).
76. Ibrahim, M. R. & Lyons, T. Imagesig: A signature transform for ultra-lightweight image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 3649–3659 (2022).
77. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
78. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks.
79. Kingma, D. P. & Ba, J. *Adam: A Method for Stochastic Optimization.* https://arxiv.org/abs/1412.6980. Accessed: 2019-04-23 (2015).

Acknowledgements

Mohamed Ibrahim was funded in part by The Alan Turing Institute, the Data Centric Engineering Programme (under the Lloyd's Register Foundation grant G0095). Terry Lyons was funded in part by the EPSRC [grant number EP/S026347/1], in part by The Alan Turing Institute under the EPSRC grant EP/N510129/1, the Data Centric Engineering Programme (under the Lloyd's Register Foundation grant G0095), the Defence and Security Programme (funded by the UK Government) and the Office for National Statistics and The Alan Turing Institute (strategic partnership) and in part by the Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA).

Author contributions

M.I developed the method and wrote the main manuscript text and M.I. prepared all figures. T.L. applied for project fund. T.L. and M.I. reviewed the manuscript. T.L. supervised the work.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-86532-8.

Correspondence and requests for materials should be addressed to M.R.I.

Reprints and permissions information is available at www.nature.com/reprints.

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