Origin-destination specific traffic emissions and data-driven NO_2 pollution-optimal routing in urban environments⋆

Samantha Ivings[a,*], James A. King[b], Alexander Roocroft[c], Patricio Ortiz[a], Toby Willis[a], Maria Val Martin[b], Hadi Arbabi[d] and Giuliano Punzo[a]

[a]School of Electrical and Electronic Engineering, University of Sheffield, Sir Frederick Mappin Building, Sheffield, S1 3JD, South Yorkshire, UK
[b]School of Biosciences, University of Sheffield, Alfred Denny Building, Western Bank, Sheffield, S10 2TN, South Yorkshire, UK
[c]Delft University of Technology, Mekelweg 5, Delft, 2628 CD, South Holland, Netherlands
[d]School of Mechanical, Aerospace, and Civil Engineering, University of Sheffield, Sir Frederick Mappin Building, Sheffield, S1 3JD, South Yorkshire, UK

ARTICLE INFO

Keywords:
Traffic Assignment
Transport Theory
Optimisation
Wardrop Equilibrium
Pollution
Air Quality

ABSTRACT

Urban air pollution from traffic poses serious public health risks. Pollution exposure can be minimised through traffic routing systems; these currently rely on detailed local environmental information, which is often difficult to collect or generalise within and across cities. Here, we introduce a new data-driven approach for ready application to different urban road networks by directly relating NO_2 to traffic density in a time-dependent and weather-corrected manner. We demonstrate this application by comparing pollution-optimal routings, using our novel direct NO_2/density approach, to the conventional traffic assignment minimising user travel time, in a case study of Sheffield, UK. There, we find user-optimal traffic flows result in 21% higher total NO_2 concentrations than pollution-optimal routings, while saving only 9% in total travel time: an average of 0.3 minutes per road. Our generalisable framework offers a practical alternative to current emissions-based models for air-quality-aware traffic control and environmental zone planning.

1. Introduction

Urban air pollution remains a defining environmental and public health challenge, with nitrogen dioxide (NO_2) strongly linked to motor traffic. In the UK, 68% of roadside NO_2 concentration measurements in 2023 originated from nitrogen oxides (NO_x) emissions from exhaust pipes (DEFRA, 2025). This has resulted in widespread non-compliance with the annual mean concentration limit for NO_2 set by the Air Quality Standards Regulations (UK Government, 2010), particularly at many urban roadsides. Traditional routing systems optimise for efficiency – minimising time, distance, or fuel – but ignore the spatial distribution and magnitude of pollutant exposure caused by the chosen routings. This disconnect represents a missed opportunity for cities seeking to balance mobility with air quality outcomes.

The EU standard tool for calculating vehicle-associated road emissions is COPERT (European Environment Agency, 2011). This tool requires inputs such as vehicle type, vehicle population, mileage, and speed to model various output emissions, including carbon dioxide (CO_2) and those in the NO_x family. Limited environmental conditions may be optionally input to the model, such as minimum/maximum/ambient temperatures, humidity, and Reid vapor pressure. However, important factors such as wind speed – which plays a large role in pollutant dispersion – are ignored (Rodríguez, Van Dingenen, Putaud, Dell'Acqua, Pey, Querol, Alastuey, Chenery, Ho, Harrison et al., 2007; Grundström, Hak, Chen, Hallquist and Pleijel, 2015). The COPERT model is highly adapted for shaping policies around air quality, but does not provide tools required for scenario analysis or optimisation of traffic routing.

Despite these limitations, COPERT has facilitated the development of traffic-control strategies based on air quality in cases where in-region emissions sensors are not available (Ingole, Mariotte and Leclercq, 2020; Kovács, Leelőssy, Tettamanti, Esztergár-Kiss, Mészáros and Lagzi, 2021). Using the COPERT IV model (Ntziachristos, Gkatzoflias, Kouridis and Samaras, 2009), a gating control strategy to influence road users' route choice in an urban environment found a network-wide reduction of 9.3% in NO_x emissions compared to the uncontrolled scenario (Ingole et al.,

⋆

*Samantha Ivings
s.e.ivings@sheffield.ac.uk (S. Ivings)
ORCID(s): 0000-0002-1442-239X (S. Ivings)

2020). This demonstrates how, given a fixed origin-destination (OD) travel demand, re-routing traffic can lead to an improvement in air quality. However, placing gates around the city perimeter requires infrastructural investment; policy interventions aimed at changing driving behaviour offer an alternative. In cities such as Nanjing, the rise of NO_2 concentrations has been significantly decreased by alternating groups of private vehicles able to circulate daily, with groups split by license plate characteristics (Tu, Xu, Wang, Wang and Jin, 2021). This supports the assertion that local transport policy alone can positively impact air quality, although reliance upon an arbitrary decision metric – such as splitting the vehicle fleet based on license plates – lacks an optimised strategy.

Traffic assignment (TA) models are useful tools for the evaluation of changes to transport networks, as they predict traffic flow patterns based on a fixed OD demand and network topology. When paired with emission models, TA models allow the assessment of road-specific environmental impacts, providing critical information into the relationship between traffic distribution and air pollution.

In the literature, a small number of examples are presented where TA has been used to generate traffic flows that minimise functions relating traffic-associated emissions (Patil, 2016; Macedo, Tomás, Fernandes, Coelho and Bandeira, 2020a). Given a particular traffic regime, an estimation of the corresponding pollutant levels may be output. To minimise travel time (as is traditional for static TA models), the bureau of public roads (BPR) congestion function is commonly used (Carlier, Jimenez and Santambrogio, 1964), which relates travel time to flows along each road segment. By re-formulating emissions-speed functions as emissions-flow, and combining this with the BPR equation, a functional relationship between emissions and traffic flow may be incorporated into TA models. This approach has been applied to minimise CO_2 across hypothetical transport networks, using popular BPR-specific parameter values in the CO_2 cost function (Patil, 2016). Counter-intuitively, though the regime minimised CO_2 emissions for each user, higher levels of CO_2 were output than in the regime to minimise user travel times.

It is appropriate to use models tuned with actual data, as the differing characteristics of roads can result in divergent traffic and emission outcomes, even within the same city. Data-informed model selection facilitates more accurate impact evaluations of in-region policy changes. Moreover, as different traffic profiles show varied emissions with respect to either traffic flow or density, it may be more appropriate to consider emissions models that depend on traffic density (Tsanakas, Ekström and Olstam, 2017; Tsanakas, 2021). For example, when a road becomes heavily congested, vehicle flow virtually drops to zero, but exhaust-pipe emissions remain high (Shi, Di, Zhang, Feng and Svirchev, 2018). This highlights that density and vehicles' spatial distribution, rather than flow, may better capture emission dynamics under certain conditions (Tsanakas, Ekström and Olstam, 2020). Finally, in addition to the spatial distribution of vehicles, the fleet composition – such as vehicle type, age, and fuel technology – has been highlighted as a key determinant for output emissions, sometimes even exceeding the influence of traffic distribution itself (Tsanakas et al., 2020).

Environmental and meteorological factors, such as wind and temperature, play a significant role in shaping local air pollution profiles. Wind affects NO_2 concentrations by dispersing or accumulating NO_2 depending on its strength and direction (Shen, Jiang, Feng, Zheng, Cai and Lyu, 2021), while temperature affects also atmospheric mixing and pollutant dispersion (Zhang, Wang, Hu, Ying and Hu, 2015).

Lower temperatures are associated with elevated NO_2 concentrations, as reduced atmospheric mixing limits pollutant dispersion, and a shallower planetary boundary layer (PBL) that traps emissions near the ground. Relative humidity also influences NO_2 concentrations by enhancing the conversion of NO_2 into nitric acid through reactions with hydroxyl radicals and water vapour. While this reduces NO_2 concentrations, it contributes to overall air quality degradation by forming particulate pollution (Zhang et al., 2015). Although NO_2 has a relatively short atmospheric lifetime, local traffic emissions are typically the dominant source in urban areas, although long-range transport from industrial regions can occasionally impact air quality under favourable weather conditions (Shen et al., 2021; Zhang et al., 2015; Pope, Arnold, Chipperfield, Latter, Siddans and Kerridge, 2018; Pope, Savage, Chipperfield, Arnold and Osborn, 2014).

In addition to meteorological effects, chemical interactions between NO_2, nitrogen oxide (NO), and ozone (O_3) introduce important non-linearities into the relationship between traffic emissions and NO_2 concentrations. At low traffic densities, NO_2 concentrations increase approximately linearly with emissions. However, at high traffic densities, i.e., near roadways, O_3 titration – where O_3 is depleted by excess NO – limits the formation of NO_2, meaning that increases in emissions do not proportionally increase NO_2 concentrations (Richmond-Bryant, Snyder, Owen and Kimbrough, 2018). These complex dynamics, influenced by chemical equilibrium and atmospheric conditions, emphasize the importance of tailoring emissions models to account for both traffic characteristics and local

environmental variability (Tsanakas et al., 2020; Kuik, Kerschbaumer, Lauer, Lupascu and Schneidemesser, 2018; Kendrick, Koonce and George, 2015).

In this study, we introduce a novel OD-specific, pollution-sensitive routing framework that directly links disaggregated traffic density to empirical NO_2 concentrations. Central to our approach is a data-derived pollution cost function that is robust to short-term meteorological variation (de-weatherised) and seasonality, enabling generalisability across urban contexts. Unlike static emissions factors, our cost function captures real-world variation in NO_2 concentrations from traffic patterns and local meteorology. This provides a transferable foundation for route optimisation without requiring infrastructure changes to the transport network.

We demonstrate our framework using traffic and air quality data from the city of Sheffield (UK). However, the method itself is not bound to this urban environment: the cost function is generalisable and requires only edge-wise traffic densities, which can be obtained from standard TA models.

To achieve this, we combine traffic data gathered by the Sheffield City Council (SCC) for TA modelling, with environmental sensor data from the Sheffield Urban Flows Observatory to model the associated NO_2 concentrations caused by transport emissions.

In particular, this paper delivers the following:

a) A formulation of the TA problem with a novel congestion function – fitted to empirical sensor data – that reflects the air quality cost of mobility.
b) A characterisation of the data-driven TA for minimising NO_2 pollution in comparison to the time-based user-equilibrium for an urban case study (Sheffield, UK), including comparisons between optimised traffic flows in the two scenarios, and the associated spatial mapping of NO_2 concentrations.
c) Analysis of the role of vehicle fleet composition on roadside NO_2 concentrations, using vehicle-type data from an automated number plate recognition (ANPR) camera. We show that passenger car flow emerges as the dominant predictor of NO_2, likely reflecting congestion-related emissions rather than tailpipe intensity alone.

This analysis allows us to assess the potential air quality improvements achievable through pollution-optimal TA.

2. Methods and data

2.1. Transport network construction

To construct the road network model, topological data was imported with geographic coordinates from OpenStreetMap (OpenStreetMap contributors, 2017), which was subsequently processed into a graph-based representation (Boeing, 2024). The imported network edges correspond to all roads within Sheffield accessible by car (Figure 1a). Chains were resolved by merging nodes of degree two, resulting in nodes representing major road junctions. The network is directed, as traffic flow along an edge in one direction may not equal that in the opposite direction.

Network reduction was achieved by computationally merging nodes within a distance threshold of 58.6m: the median radius of Sheffield lower layer super output areas (LSOAs). These are statistical boundaries for population, and is the level at which the travel demand data is available. Reducing the network led to an 81% and 74% decrease in nodes and edges, respectively, compared to the full network. The reduced network was used throughout the TA modelling process (Figure 1b). Sensitivity testing was performed on this threshold to ensure nodes were not over-merged (Figure 20 in Appendix).

The following assumptions were considered to consolidate the data available with the city-scale modelling:

a) Drivers join the transport network at the closest junction to their origin location.
b) Drivers choose to drive the route which is the shortest time for them.
    (a) To this respect, the literature suggests that drivers often choose a consistent route, not necessarily the fastest (Ciscal-Terry, Dell'Amico, Hadjidimitriou and Iori, 2016). However, in the static TA, the user-equilibrium state is generally modelled such that traffic flows correspond to drivers all choosing the quickest route (Sheffi, 1985).
c) Through traffic - that is, traffic that does not both begin and end on the network - does not have a significant effect on local travel times or air quality. For the size of the area we consider, this assumption is justified considering that the majority of journeys by private vehicle are much shorter than the region's diameter (DfT, 2024).

The major routes (motorways, A-roads, and B-roads) host many of the sensors throughout the city, which have been prioritised to roads with higher traffic and/or locations near schools.

(a) Full network (3943 nodes, 8405 edges), consisting of motorways, A, B, tertiary, and unclassified roads.

(b) Reduced network used for modelling (754 nodes, 2152 edges), with distance threshold 58.6m for node merging.

Figure 1: Car-accessible road networks with administrative boundary for the city of Sheffield. Edge thickness is proportional to road class seniority.

2.1.1. Network origin-destination pairs

Demand between OD pairs was averaged at an hourly resolution across Lower Super Output Areas (LSOAs). This analysis focused on weekday AM and PM peak periods – defined as 07:00–09:00 and 16:00–19:00, respectively – to capture commuting-related travel. To align with the available sensor data, the OD dataset was restricted to the period from February to October. For each LSOA within Sheffield, OD demand matrices were derived from anonymised vehicle movement data collected by The Floow™ through black box devices installed in private vehicles in 2019 (The Floow, 2023). Comparing the average weekday and weekend travel demand within the AM and PM peak periods, we find that weekend demand (11,110 cars/hour) is 34.7% less than weekday demand (17,015 cars/hour). Moreover, an additional 8.3% of all possible OD LSOA pairs have zero average demand on weekends.

A distance-based method was employed to assign representative origin and destination nodes within each LSOA. Each transport network node was first spatially joined to its corresponding LSOA based on boundary containment. The likelihood of a node being selected as an origin or destination was weighted inversely with its distance from the LSOA's population-weighted centroid (PWC). Where LSOAs with valid OD demand did not contain any nodes from the computational network, the node nearest to the PWC was assigned as a proxy.

It is important to note that the TA model does not address intra-LSOA traffic, as such trips are typically not subject to strategic re-routing. Both inter- and intra-LSOA OD demands for private vehicles are shown in Figure 9 in Appendix.

2.2. Data inputs for the traffic assignment models

2.2.1. Road characteristics

Through the OpenStreetMap (OSM) road network import (OpenStreetMap contributors, 2017; Boeing, 2024), three main road features were considered: (1) speed limit, (2) number of lanes, and (3) road class. The road network analysis incorporated a multi-tiered imputation strategy to address missing and anomalous attribute data from OSM. First, physically implausible width values (i.e., roads with widths less than 70% of the minimum expected width based on lane count) were identified, and subsequently treated as missing data. Missing speed limits were imputed using a hierarchical approach based on UK road classification standards (DfT, 2018), while lane counts were estimated using within-class median values supplemented by classification-based defaults aligned with Design Manual for Roads and Bridges guidelines (Highways England, 2020). For width imputation, we used a rule-based estimation model, incorporating lane width parameters calibrated to observed road characteristics (3.4m for high-speed roads, i.e. with speed limit greater than or equal to 50mph, to 2.0m for low-speed roads), with correction factors applied to account for systematic patterns in OSM data quality (Boeing, 2017; Barrington-Leigh and Millard-Ball, 2017). Additional width allowances were included to represent features such as central reservations, hard shoulders, and edge buffers, with the specific values adjusted according to road type and function (Highways England, 2020). This approach allowed for reliable estimation of road parameters, while accounting for known limitations in the OpenStreetMap data. Given that the input road network forms the basis for TA modelling, city-wide sensor data was extensively used to validate the associated road attributes.

Based on the above characteristics, road-specific maximum flow capacities were joined to the network using standard methodology (TfL, 2013), assuming a 50/50 distribution of traffic per direction. Free-flow travel times for

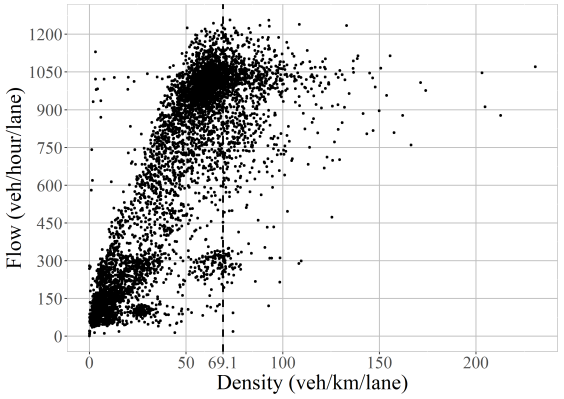


Figure 2: The quasi-density Fundamental Diagram for the Sheffield ring road, aggregated across two sensors, capturing North/South bound traffic respectively. The critical density is calculated as 69.1 veh/km.

each road were computed using the road's length and speed limit. Additional road characteristics were obtained using the following data.

2.2.2. Traffic data

The traffic data in the city was collected using inductive loop sensors and an automatic vehicle counting camera, co-located with an AQMesh air quality sensor (Section 2.2.3. Three types of data were gathered for this research:

- Traffic flow: the number of vehicles passing the sensor (veh/min).
- Sensor occupation: the percentage of time the sensor area was occupied by a vehicle, used as a proxy for traffic density (veh/km).
- Traffic speed: the average speed of all vehicles recorded in each observation period (km/hour).

Data points were collected and aggregated with a 30-minute frequency.

Figure 2 shows the relationship between traffic flow and density – known as the quasi-density Fundamental Diagram – from our sensor data measure on the Sheffield ring road. The flow-density relationship is well-established in the literature (Knoop and Hoogendoorn, 2014; Helbing, 2009). Flow behaviour can be categorised in two regimes relative to the critical density: hypo- and hyper- critical. In the hypo-critical regime, the flow-density relationship on a single road is linear, that is, with low congestion. In the hyper-critical regime, increasing density results in a reduction of flow due to congestion. This behaviour highlights that traffic flow is not the most appropriate predictor for travel time; travel time increases monotonically with density but not necessarily with flow (Kucharski and Drabicki, 2017). Therefore, density provides a more appropriate measure for modelling traffic and its impact on emissions.

Traffic density (veh/km) was calculated from sensor occupancy using the following relationship:

$$\text{density (veh/km)} = 10 \times \frac{\text{occupancy (\%)}}{\text{length of sensor (m)}}. \tag{1}$$

When a road-specific Fundamental Diagram is available, the critical density can be estimated as the density corresponding to the 95th percentile of observed traffic flows (Kucharski and Drabicki, 2017). For the Sheffield ring road, the critical density was 69.1 veh/km.

Relationships between NO_2 concentrations and traffic flow and density are shown in Figure 13 in Appendix.

2.2.3. Air quality and meteorological data

Air quality data was collected by an AQ Mesh sensor collocated with a traffic sensor (Section 2.2.2). The AQMesh (Environmental Instruments Ltd., UK, Gas algorithm v4.2.3) is a multi-sensor platform capable of measuring several gases with electrochemical (EC) gas sensors, as well as temperature and humidity (AQMesh, 2017). For this study, the AQMesh was equipped with EC sensors (B4-series, Alphasense, UK) to measure NO, NO_2, and O_3. The manufacturer-reported limits of detection are 6.25 µg/m^3 for NO and 19.2 µg/m^3 for NO_2 (AQMesh, 2017). The AQMesh also

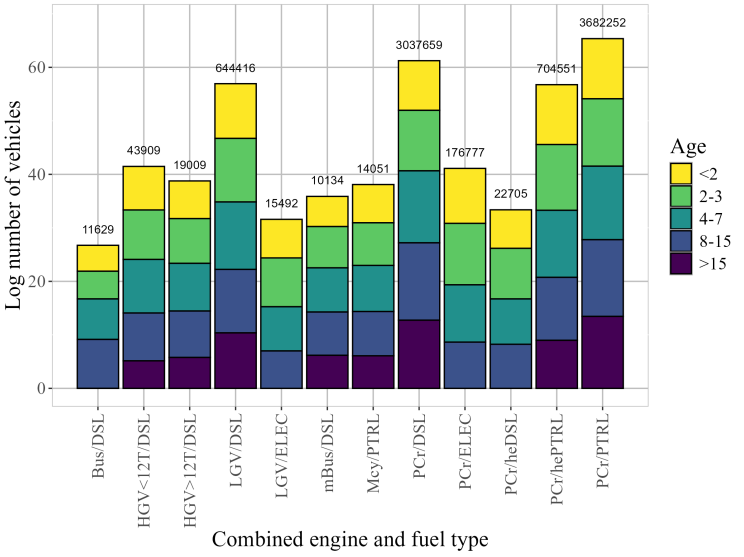


Figure 3: Breakdown of vehicle types passing the automated number plate recognition (ANPR) camera in Sheffield from Feb-Oct 2023. Abbreviations: diesel (DSL), petrol (PTRL), electric (ELEC), heavy/light goods vehicle (H/LGV), minibus (mBus), motorcycle, excluding motor-assisted bicycles (Mcy), passenger car (PCr), hybrid electric (he), tonnes (T).

included an optical particle counter for measuring particulate matter in different size fractions (PM_1, $PM_{2.5}$, PM_{10}); however, PM data was not used in the analysis.

All individual sensors were factory-calibrated and mounted into the AQMesh unit. AQMesh sensors have been extensively used in air pollution studies across Sheffield (Munir, Mayfield, Coca, Jubb and Osammor, 2019; Redondo Bermúdez, Chakraborty, Cameron, Inkson and Val Martin, 2023), as well as in other locations (Hickman, Baker, Cai, Delgado-Saborit and Thornes, 2018; Wahlborg, Björling and Mattsson, 2021), highlighting their reliability and applicability in urban air quality monitoring. In this study, data from the AQMesh sensor were uploaded via GPRS communication to a remote database. The transmitted raw data was cross-referenced and validated against air quality data from nearby UK DEFRA and SCC monitoring sites. AQMesh data was then aggregated into 30-minute averages to align with the temporal resolution of the traffic and meteorological datasets. The NO_2 sensor readings used in this study cover the period from 24th Jan to 31st Oct 2023 (Figure 12 in Appendix).

Meteorological data for this study was obtained from the automated weather station at Sheffield Hallam University City Athletics Stadium (location: 53°23'20.4"N 1°25'55.2"W). The station provided measurements of temperature, air pressure, rainfall, and wind speed/direction at 10m and 24m above the ground.

2.2.4. Vehicle fleet composition data

Data on vehicle fleet composition was obtained from camera footage using ANPR technology. ANPR data was interfaced with the Driving and Vehicle Licensing Agency (DVLA) register to retrieve vehicle categories, age, engine type and fuel used. Privacy was maintained, as vehicle data was anonymised at the source by removing license plate details. The function of our ANPR camera was solely number plate recognition. Vehicle type data came directly from cross-correlating plate numbers with DVLA vehicle information; a procedure performed by the ANPR camera's computer. This approach provided a census of the circulating vehicle fleet over nine months (Feb-Oct 2023), as summarised in Figure 3. This data was subsequently used to calibrate the air pollution cost function.

2.3. Building the traffic assignment model

The TA problem involves assigning road-wise optimal flows in order to meet the given travel demand between OD pairs. The assignment depends on the congestion function associated with each road.

In this work, we use two types of cost function that relate traffic density either to NO_2 concentrations or travel time. Both of these functional forms are specific to road segments within the transport network. The data-informed formulation of these cost functions, along with the formalisation of the TA problem, is detailed in this section.

2.3.1. Traffic-associated NO_2 concentration

No consensus exists in the literature for the most appropriate choice of function representing the relationship between traffic parameters and NO_2. Most existing studies model NO_2 primarily as a function of vehicle speed (Patil, 2016; Macedo, Tomás, Fernandes, Coelho and Bandeira, 2020b). In this work, we seek to develop a pollution-based cost function dependent solely on traffic density.

To assess the reliability of the AQMesh measurements, we first compared them to NO_2 observations from three nearby DEFRA reference monitoring stations: three classified as urban traffic (Barnsley Road, Lowfield and Wicker) and three as urban background (Devonshire Green, Tinsley and Firvale) (Figure 18 in Appendix). The AQMesh sensor showed good agreement with the urban traffic sites (r^2=0.4-0.5, mean bias=-0.3-9 $\mu g/m^3$) and a consistent, although elevated, pattern relative to the urban background sites (r^2=0.20-0.55, mean bias=14-17 $\mu g/m^3$) – confirming its sensitivity to local traffic emissions while aligning with expected spatial gradients within the city.

The AQMesh NO_2 measure in parts per billion (ppb) were then de-weatherised by accounting for meteorological effects, such as temperature, humidity, and wind speed, in order to isolate traffic-driven changes. In addition, cyclic seasonal, weekly, and daily components were resolved through de-seasonalisation to eliminate long-term temporal trends unrelated to traffic. The dataset was randomly partitioned into training and testing subsets with a 4:1 ratio. De-weatherisation and de-seasonalisation was performed using predictive random forest models for meteorological normalisation (Grange and Carslaw, 2019; Grange, Carslaw, Lewis, Boleti and Hueglin, 2018).

The random forest model achieved a correlation coefficient of $r = 0.77$ between predicted and actual NO_2 (Figure 14 in Appendix). The ranked feature importances for training the random forest model are shown in Figure 15 in the Appendix, indicating that temperature (21%) was the strongest predictor of NO_2 concentrations, followed by traffic density (13%).

While de-weatherisation removes meteorological confounding factors, it does not inform us of the direct relationship between NO_2 and traffic density that is independent of other variables. To isolate the specific influence of traffic density on NO_2, the partial dependence between NO_2 and density was computed using established methods (Friedman, 2001). The partial function $\hat{f}_S(x_S)$ was estimated from the training data using the equation

$$\hat{f}_S(x_S) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_S, \mathbf{x}_C^{(i)}) \tag{2}$$

for the random forest model $\hat{f}$ with n observations, where x_S is the density and $\mathbf{x}_C$ is the set of remaining features used for model training. Traffic flow was excluded from the random forest model, as flow and density are highly correlated, which violates the conditions for computing partial dependency.

The estimated partial dependence was then used to train a simple linear model with general formulation

$$t_{(ij)}^{NO_2}(\Omega_{(ij)}) = m \times \Omega_{ij}, \tag{3}$$

where m is a constant gradient. A linear relationship was selected to represent the approximately linear increase in NO_2 concentrations with traffic density up to the onset of ozone titration effects (Richmond-Bryant et al., 2018).

To calibrate the model (Equation (3)), a zero-intercept offsetting was applied, assuming NO_2 concentration approaches zero at zero traffic density. A parameter search was then used to fit the gradient $m = 0.0137$, such that the model intersected the observed NO_2 saturation point at traffic density= 240.125 veh/km. The resulting model achieved a coefficient of determination (r^2) of 0.91, as shown in Figure 4. This linear fitted function, derived from de-weatherised data and expressed as a partial dependency on traffic density, allows extrapolation to continuous density values across different roads. As such, it served as a road-specific cost function for use in the TA modelling framework.

NO_2 concentrations output by our TA models are faithful to the model (Equation 3), i.e., road-side pollutant concentrations directly caused by the density of vehicles. This relationship is used for TA modelling as it is these concentrations which are able to be tackled through vehicle re-routing. To re-incorporate background levels of NO_2, a scaling factor is applied to our modelled NO_2 outputs. This factor is derived from the median multiplicative difference between the zero-offset partial dependency values for NO_2 concentrations and traffic density, and the non-offset values which inherently include background pollutant levels. Unit conversion from ppb to $\mu g/m^3$ is achieved using the ideal gas law.

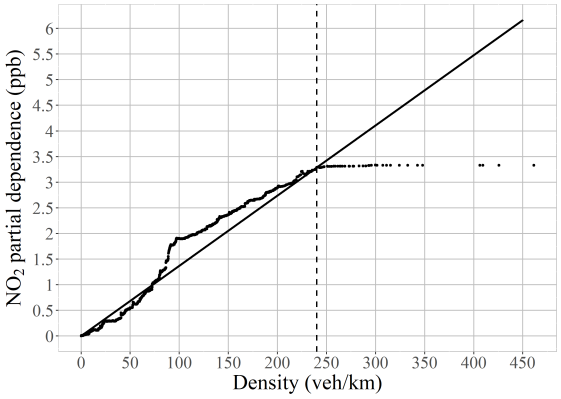


Figure 4: Scatter plot with fitted linear function for the partial dependency of NO_2 concentration on traffic density.

2.3.2. Travel time

The BPR congestion function is typically used to model travel time with respect to road-wise traffic flows (Carlier et al., 1964). Here, we convert flow to density as is more appropriate for the application of understanding traffic-related emissions (Shi et al., 2018). The BPR congestion function thus takes the form

$$t^{\mathrm{BPR}}_{(ij)}(\Omega_{(ij)}) = t^0_{(ij)}\left(1 + \alpha\left(\frac{\Omega_{(ij)}}{\Omega_{(ij)}(\kappa_{(ij)})}\right)^{\beta}\right), \tag{4}$$

where $t^0_{(ij)}$ is the free-flow travel time for edge (ij), $\Omega_{(ij)}(\kappa_{(ij)})$ is the road-wise critical density, and α, β are constants. We use parameter values $\alpha = 0.15$ and $\beta = 4$ which are widely accepted in the literature (Patriksson, 2015).

The road-wise critical density $\Omega_{(ij)}(\kappa_{(ij)})$ is determined using the linear conversion

$$\Omega_{(ij)}(\kappa_{(ij)}) = \frac{\kappa_{(ij)}}{s_{(ij)}} \tag{5}$$

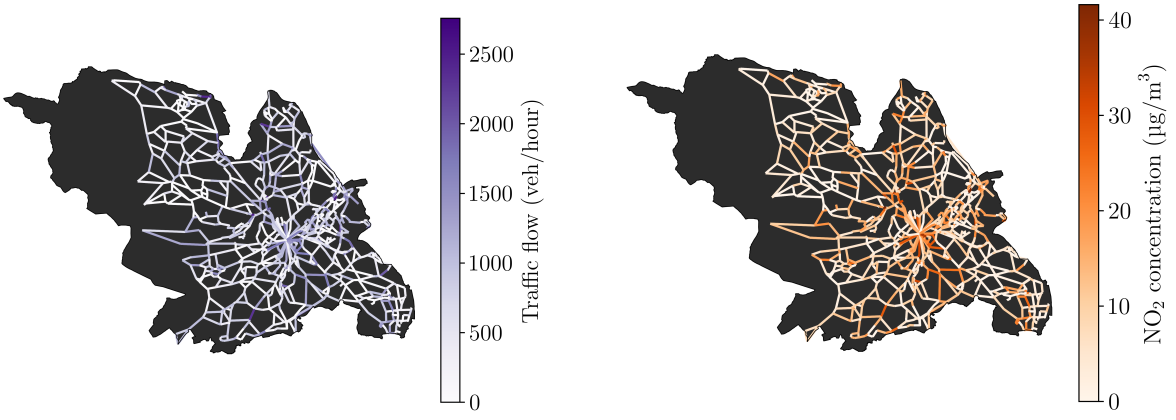
for average speed $s_{(ij)}$ (km/time) and road capacity $\kappa_{(ij)}$ (veh/time), where the latter corresponds to maximum observed flow along the edge (ij). Calculated from observed sensor data, the average vehicle speed on the ring road was estimated at approximately 30% of the posted speed limit. For motorways, where congestion is typically lower and traffic flows more freely, a higher scaling of 60% of the speed limit was applied. This factor aligns with practical estimates commonly used in transport modelling when empirical speed data is unavailable (TfL, 2020; DfT, 2023). Density-dependent travel times output from the BPR function applied to the Sheffield ring road are shown in Figure 16 in Appendix.

Whilst popular, the BPR equation is not the only function appearing in the literature relating travel time with flow/density. Alternative choices include Davidson (Davidson, 1966), Akçelik (Akçelik, 1991), and conical (Spiess, 1990), which - depending on the traffic characteristics and the cost considered - may be more or less appropriate than the BPR function. When fitted to traffic data, the BPR function out-performs alternatives in terms of computational time, which is of particular importance when dealing with large transport networks (Roocroft, Ramli and Punzo, 2023).

2.3.3. The traffic assignment problem

To formalise the TA problem, the road network is modelled as a directed graph $G(V, E)$, with a set of nodes V (with $|V| = N$) and a set of ordered pairs of nodes, called edges, $E \subseteq V \times V$. The set of OD pairs is given by $C = \{(p, q)|p, q \in V\}$, which have fixed demands $d_{p,q}$. The road-specific objective function is denoted $T_{(ij)}(\Omega_{(ij)})$ for edge $(ij) \in E$. An OD pair may be denoted $(p, q) = c \in C$. With these premises, the TA problem can be expressed as:

$$\min_{\mathbf{x},\mathbf{y}} \; T_{(ij)}(\Omega_{(ij)})$$
$$\text{s.t. } x_{(ij)} = \sum_{c\in C}\sum_{k\in K_c} a^{kc}_{ij} y^k_c \quad \forall (ij) \in E,$$



(a) Traffic flows.

(b) NO_2 concentrations.

Figure 5: Road-wise TA outputs for the current transport scenario in Sheffield, from Feb - Oct 2019 during peak hours.

$$
\begin{aligned}
& d_c = \sum_{k \in K} y_c^k \quad \forall c \in C, \\
& x_{(ij)} \geq 0 \quad \forall (ij) \in E, \\
& y_c^k \geq 0 \quad \forall c \in C \quad \forall k \in K_c,
\end{aligned}
\tag{6}
$$

where $k \in K_c$ is a path between an OD pair, and y_c^k is the flow between the OD pair c on path k. The entry $a_{(ij)}^{kc}$ in the node adjacency matrix bounds the flow $x_{(ij)}$ along edge (ij) to the flow of paths passing through the edge.

For time-based TA to calculate the user-equilibrium flows (i.e. the current scenario), the objective function is

$$
T_{(ij)}(\Omega_{(ij)}) = \sum_{(ij) \in E} \int_0^{\Omega_{(ij)}} t_{(ab)}^{\mathrm{BPR}}(\omega_{(ab)})\, d\omega_{(ab)}. \tag{7}
$$

Combining this with Equation (6) gives flows such that trip travel times are minimised for each road user, but are unlikely to be optimal for the whole network.

For a system-optimal TA, the objective function is

$$
T_{(ij)}(\Omega_{(ij)}) = \sum_{(ij) \in E} \Omega_{(ij)} \cdot t_{(ij)}(\Omega_{(ij)}), \tag{8}
$$

as formulated by Beckmann *et al.* (Beckmann, McGuire and Winsten, 1956). Here, the cost function $t_{(ij)}(\cdot)$ is taken as $t_{(ij)}^{\mathrm{BPR}}(\cdot)$, i.e. Equation (4), for time-based TA, and respectively $t_{(ij)}^{\mathrm{NO_2}}(\cdot)$, i.e. Equation (3), for pollution-based TA.

The TA problem on a network is solved using the Frank-Wolfe algorithm (Frank and Wolfe, 1956).

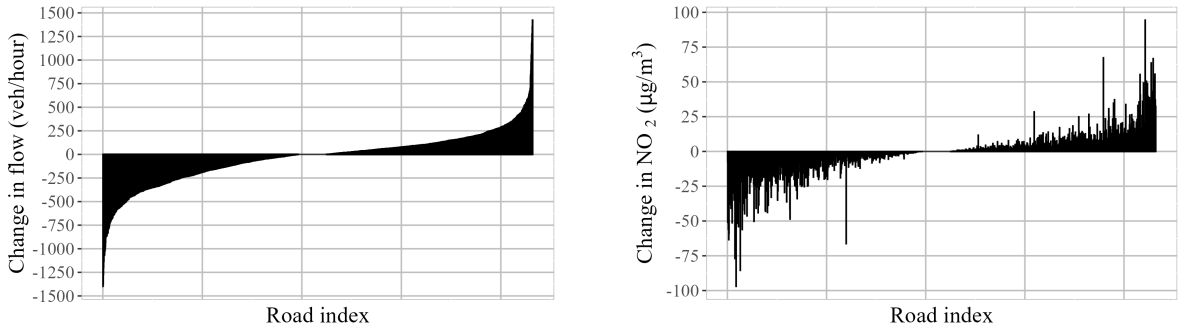
3. Results of data-driven traffic assignment

With the above formulations, TA models were created to capture the current scenario for the mobility of private cars over the considered date range, and, importantly, the pollution-optimal assignment of traffic flows while meeting the same OD demands. In this section, we present the results from these two scenarios.

3.1. Road-wise results for the current and pollution-optimal scenarios

First consider the road-wise outputs for each of the TA model scenarios, reflecting the present state[1], mapped in Figure 5.

[1] Roads are considered as one or more consecutive segments.



(a) Required change in traffic flows per road.

(b) Change in NO_2 concentrations per km of road.

Figure 6: Road-wise changes in NO_2 concentrations and traffic flows achieved through taking the pollution-optimal routings. Road indices are ordered based on ascending change in flows from (a).

Based on this, the road-wise changes in traffic flows required to move from the current to the pollution-optimal scenario is shown in Figure 6a. In total, 981 roads (45.4%) require flow reduction, 1,039 (48.1%) require an increase, and 140 (6.5%) require negligible – i.e., modelled as zero – change to flows. Under pollution-optimal routing, increased traffic flows are assigned to major roads that have high flow capacities. Meanwhile, roads within the same region as these, yet of a lower road class, require less flow.

For reducing traffic congestion, maintaining traffic flows less than or equal to the road capacity is pivotal. In the current scenario, our TA output shows that flows on 671 (31.1%) of roads exceed capacity. Many of these roads are situated at the city centre, extending through major arterial routes connecting the centre to the Northwest, Northeast, and South areas of the city in particular. In the pollution-optimal case, only 499 (32.1%) of roads exceed flow capacity: the city centre and the high-flow route towards the upper East of the city still experience congested roads, though the number of these is reduced. On average, roads in the current scenario are operating at around 71% capacity. In the pollution-optimal scenario, this figure is 69%, showing that minimising NO_2 concentrations desirably does not increase network-wide traffic congestion through flow management.

Associated changes in road-wise NO_2 concentrations per km of road in taking pollution-optimal routes are shown in Figure 6b. We find that 52% of roads see either no change or a reduction in NO_2 concentration by taking the pollution-optimal routings, demonstrating overall success in improving air quality.

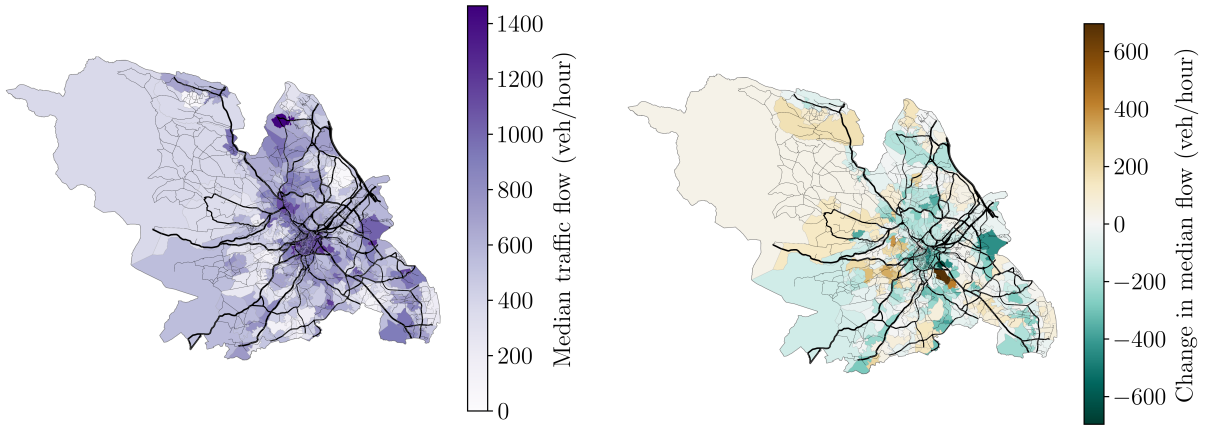
Roads with the current highest flows – and highest NO_2 concentrations – are situated within and around the city centre, diminishing towards the West of the city. Much of the East shows low flow/NO_2, with the exception of very high results on the upper East route.

Given that the preferred geographies across UK government for publishing small area statistics are lower/medium super output areas, and given the necessity of capturing small spatial variations in NO_2 for public health, we aggregate the following results to LSOA level. The median road in each LSOA is reported; it is useful to bear in mind that some LSOAs may have very few roads passing through them, particularly after performing network reduction, in which case the median road may take a high number of vehicles in comparison to LSOAs with high road counts. Click here to open an interactive LSOA map for Sheffield in 2019 (HTML). Sheffield train station is marked with a red dot, Meadowhall Shopping Centre a blue dot, and green dots mark the hospitals Northern General and the Royal Hallamshire.

3.2. Traffic flows for the current and pollution-optimal scenarios

For the current mobility scenario, Figure 7a shows the median road's traffic flows within each of Sheffield's LSOAs, output by the user-equilibrium TA model. The ten highest regional flow contributors, making up 3% of all Sheffield LSOAs in 2019, account for a combined total of 6% of modelled traffic flows along the median road per LSOA (Table 1 in Appendix). Several of these LSOAs are situated around Sheffield's train station slightly East of the centre, with high flows also in directly bordering regions. Moving away from the centre, high flows remain along major arterial roads; particularly those connecting the North and East of the city with the centre.

In terms of commuting routes, we see that a route with some of the heaviest traffic flows in the current scenario is associated with commuting to/from the East of Sheffield city towards both Rotherham and, further afield, Doncaster. Moreover, this route is situated very closely to both Meadowhall (Sheffield's major shopping complex) and the Northern



(a) Flows per LSOA in the current scenario.

(b) Current to pollution-optimal flow changes.

Figure 7: LSOA-aggregated traffic flow results from two TA models: the current scenario, and the pollution-optimal. For each LSOA, the median road is visualised.

General Hospital. The highest flow LSOA along this route is an industrial area (company count 1,676, population 1,872), while other high-flow LSOAs are residential; Figure 10 provides spatial visualisations of these metrics. These results indicate that residential LSOAs, located near major roads connecting industrial zones, are most impacted by through-traffic.

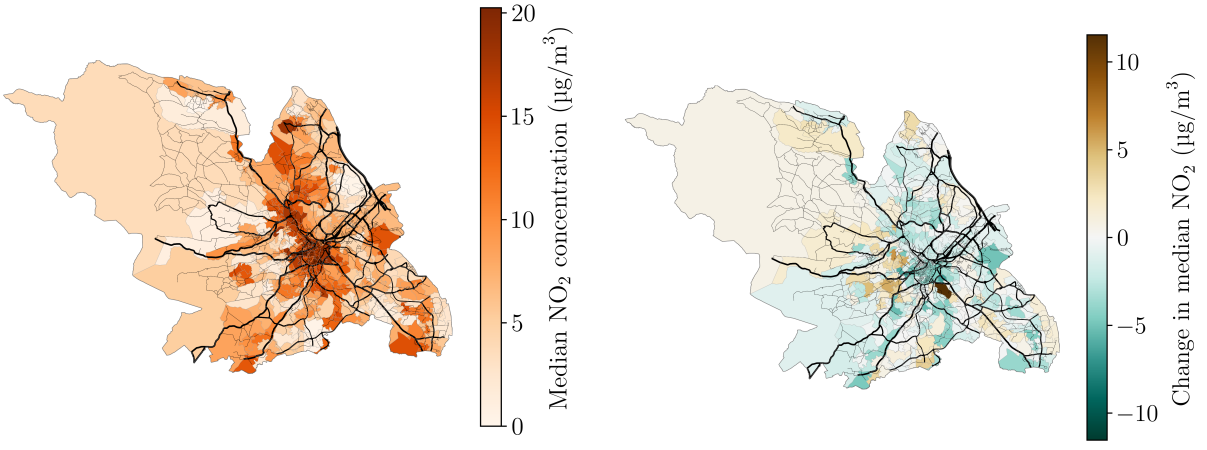
The shift in median road flows required to move from the current scenario to the pollution-optimal is shown in Figure 7b. In the pollution-optimal case, we find that the industrial district between Sheffield and Rotherham would see an increase of traffic away from the shortest route between the origin and destination, distributed instead along currently low-flow roads. In particular, increased flows are assigned within a small number of spatially heterogenous LSOAs peripheral to the city centre, with the largest flow increase directly South of the train station. This redistribution appears to facilitate flow reductions in neighbouring regions. The North/Mid-West of the city would also experience slight increases in traffic flows, as these areas are currently traversed very little. As a result, increasing the flow of vehicles in these areas provides a more balanced distribution across the city. Flow reductions are found in the East, and along spatially connected regions in the North; residential areas that are well-serviced by Sheffield's public transport networks.

The top ten LSOAs with maximum flow in the pollution-optimal scenario still contain areas surrounding the train station, which is most likely due to the high vehicle demand starting/ending in this region. However, all of the top-ten highest flow LSOAs in the pollution-optimal scenario would see lower median flows compared to the current scenario, achieving a 28% reduction. Across the whole network, an 8% total reduction in median road flows is able to be attained. This shows a redistribution of vehicles away from busy roads where traffic re-routing is feasible, reinforcing the road-wise results in Section 3.1.

3.3. NO_2 concentrations associated with the current and pollution-optimal scenarios

We now examine the associated NO_2 concentrations for each TA model scenarios. The ratio of the total NO_2 concentrations in the TA model between the current (time-optimal) and the pollution-optimal routing scenarios returns an Environmental Price of Anarchy (EPoA) (Koutsoupias and Papadimitriou, 2009). In our case, the EPoA is 1.21, meaning that selfish time-based routings are 21% more environmentally costly than the pollution-optimal routings.

Figure 8a shows the median NO_2 concentration per road within each LSOA for the current scenario. As expected from the modelled traffic flow counts, the highest NO_2 concentrations – especially the top ten – are found in and around the city centre, and generally coincide with the top ten highest traffic flows (Table 1 in Appendix). Elevated concentrations are especially prominent along spatially connected corridors running vertically through Sheffield and intersecting the centre. Some regions in the South, not experiencing excessive traffic flows, also show elevated NO_2 concentrations. This is likely due to the associated roads operating near or above their saturation rate (flow/capacity), leading to high air pollutant outputs.



(a) NO_2 concentrations per LSOA in the current scenario.

(b) Pollution-optimal changes in NO_2 concentrations.

Figure 8: LSOA-aggregated NO_2 concentration results from two TA models: the current scenario, and the pollution-optimal. For each LSOA, the median road is visualised.

Figure 8b shows the difference in modelled NO_2 concentrations per LSOA between the current and the pollution-optimal scenario. We find that 204 LSOAs (59%) experience a reduction in median NO_2 concentration under pollution-optimal routing, while 141 LSOAs (41%) see an increase. Although some regions in the city show modest increases in NO_2, the largest air quality improvements are observed within the currently most polluted LSOAs. Conversely, some of the increases occur in low-pollution LSOAs, suggesting a re-distribution of pollution that results in a more balanced spatial distribution of NO_2 across the city.

These changes in NO_2 concentrations also have implications for public health. Under the current traffic scenario, 37% of roads in the network exceed the World Health Organization (WHO) annual mean guideline of 10µg/m^3 during peak hours. In the pollution-optimal scenario, this proportion falls to 30%. Additionally, the proportion of roads exceeding 20µg/m^3 – a level associated with elevated long-term health risks (WHO, 2021) – declines from 8% to 5%. These results suggest that optimal traffic re-routing can meaningfully reduce population exposure to harmful levels of air pollution.

To investigate the role of vehicle fleet composition on NO_2 pollution in the city, we trained a linear regression model using forward feature selection to identify the largest traffic-related contributors. Data for vehicle types passing traffic sensors was split into training:validation:testing sets at a ratio 8:1:1.

We find that, passenger car flow alone predicts NO_2 levels with over 88% accuracy on the test dataset (89.2% in training). Adding further vehicle features provides only marginal improvements (<0.2%), and almost no gain beyond five features. The top five features, all relating to vehicle flow, were: passenger cars, e-hybrid cars, minibuses, standard buses, and electric cars. Notably, the presence of electric car flow among the top five features appears circumstantial, rather than causal, given its very small contribution to model performance.

These results demonstrate that the frequency at which vehicles traverse the road network over a given time period – particularly passenger cars – is the dominant factor influencing NO_2 concentrations. Fleet composition plays a relatively minor role in this context.

4. Discussion

The spatial distribution of NO_2 concentrations predicted by our TA model aligns well with existing high-resolution (100×100m) analyses of Sheffield's air quality (Munir et al., 2019; Munir, Mayfield, Coca and Mihaylova, 2020; Munir, Mayfield and Coca, 2021). In particular, the model captures the reported hotspots around the city centre and the upper East, with concentrations declining toward the West and Northwest (Munir et al., 2021). Our results also identify the city centre and areas around Sheffield's train station as exhibiting the highest modelled NO_2 levels, consistent with previous studies (Munir et al., 2021).

Notably, our model shows elevated NO_2 levels in the northern part of the city; a pattern not consistently reported in previous studies. By contrast, the industrial Northeast corridor, although showing some elevated levels, does not emerge as a major hotspot in our results. These differences reflect the presence of background non-transport sources of NO_2, which are explicitly not modelled in our study. The Northeast corridor is characterised by high intra-LSOA travel demand (Figure 9b in Appendix), and significant industrial activity (Figure 10b in Appendix), both of which increase NO_2 concentrations beyond road-traffic emissions alone. This is consistent with the national estimate that about 15% of UK NO_2 emissions arise from industrial processes (National Atmospheric Emissions Inventory (NAEI), 2021a). The North of Sheffield, meanwhile, having fewer industrial activities, has NO_2 levels more strongly governed by road traffic.

A limitation of our modelling approach is the use of a uniform background concentration scaling factor across the city, which does not capture spatial heterogeneity in non-traffic NO_2 sources. Nonetheless, the inclusion of urban background concentrations allows meaningful comparisons against reported air quality patterns, and our model performs well in representing NO_2 variations attributable to traffic.

Modelled total NO_2 concentrations were validated against DEFRA's Pollution Climate Mapping model (DEFRA, 2023), as well as SCC sensor datasets (Figure 19 in Appendix). Across these benchmarks, our predicted concentrations were found to be of the correct order of magnitude, supporting the reliability of the data-driven congestion functions developed in this work.

According to the UK's National Atmospheric Emissions Inventory (NAEI), road transport contributed approximately 1,214 tonnes of NO_2 emissions in Sheffield in 2022, accounting for 34% of the city's total of 3,556 tonnes (National Atmospheric Emissions Inventory (NAEI), 2021b). The Sheffield City Region Transport Model (SCRTM1), produced by the SCC, reported business-as-usual emissions of 65.05 g/s (equivalent to 2,053 tonnes/year) for road NO_x in 2022. As transport-related NO_2 emissions in Sheffield have been declining annually (Sheffield City Council (SCC), 2023), NO_2 concentrations for 2019 – the year modelled in our study – are likely to be higher than those given for 2022. These figures provide a reasonable benchmark for assessing modelled transport-related NO_2 concentrations in Sheffield over the Feb-Oct 2019 period during peak hours.

A key finding from our analysis is the significant environmental gain achievable through optimal routing. We find that selfish routing results in NO_2 concentrations 21% higher than pollution-optimised routing. This corresponds to an EPoA of 1.21. Importantly, this environmental gain is achieved with only a modest cost to network efficiency: the total travel time in the pollution-optimal scenario is just 9% higher than in the user-optimal scenario, corresponding to an average of just 0.3 minutes per road. These results highlight the potential for traffic management strategies that prioritise air quality without severely impacting travel times.

Our analysis also reveals that the sheer volume of vehicles, rather than fleet composition, is the dominant driver of NO_2 levels in Sheffield. Passenger car flow alone explains over 88% of the variation in actual NO_2 concentrations from sensor data, with marginal improvements when additional vehicle types are included. This result should be seen in relation to the current vehicle fleet composition and volume of circulating vehicles. In this scenario, the substitution of a limited number of zero-emission vehicles keeps the volume size, meaning maintaining the same level of congestion, which exacerbates the emissions of conventional vehicles in city traffic. This suggests that policies aimed at reducing overall traffic volumes are likely to be more effective in improving air quality than those focusing solely on changing fleet composition, e.g., changes in fuel type.

Our model shows a substantial health benefit for the overall population of Sheffield. Pollution-optimal routings reduce the number of roads modelled to exceed the WHO annual mean NO_2 guideline, 10μg/m^3, by 7% during peak hours. The proportion of roads exceeding 20μg/m^3, which is linked to an increased risk in long-term health outcomes, is decreased from 8% to 5%.

From a socioeconomic perspective, we observe that, under the current traffic scenario, lower-income areas (Figure 11 in Appendix) are more likely to experience higher NO_2 concentrations. The Kendall rank correlation, which is well-suited for comparing ordinal data or relationships between rankings, was used to compare median income and modelled NO_2 levels for MSOAs by rank. The correlation coefficient between these measures is 0.4, indicating a moderate positive association between lower income and higher pollution. In contrast, this correlation is eliminated under the pollution-optimal scenario, suggesting that optimal re-routing of traffic not only improves overall air quality, but also reduces inequalities in pollution exposure between more and less affluent communities.

We use a static TA model, meaning that traffic conditions are assumed to be constant over the modelled time period. This is a computationally efficient approach, and produces a good average estimate for traffic conditions within the time range. As the analysis focuses on a relatively short period (commuting hours only), demand and network

conditions can be assumed to remain approximately constant, justifying the use of a static TA model. However, a dynamic TA approach, which explicitly represents the temporal evolution of traffic and allows routes to adjust in response to prevailing conditions, could provide additional insight for real-time traffic management and short-term operational planning (Bliemer, Raadsen, Brederode, Bell, Wismans and Smith, 2017). To further tune the accuracy of the model, it would be beneficial to have more complete data for road characteristics across the network, such as number of lanes and speed limits, which were not always available. In these cases, the modelling approach described in Section 2.2.1 gave reasonable estimates based on widely used transport planning advice. We modelled the true driving speed seen during peak hours as one third of the road's speed limit (except for motorways), informed by traffic data from the Sheffield inner ring road during peak time. Further accuracy may be achieved by looking at the driving speed/speed limit relationship on additional types of roads, such as residential, using traffic flow sensor data.

We also assume that the likelihood of a node in the road network being an origin or destination point is inversely proportional to its distance from the corresponding LSOA's population weighted centroid. This means that a node is more likely to be selected if it is close to where people in the area live. However, it may be interesting to compare the results of this study with a slightly alternative approach: modelling the likelihood of a node being a commuting destination with respect to its distance from the economic centre, e.g., by using Companies House data to map business densities. As we present the majority of our results at LSOA level, is expected that this alternative approach would not produce substantially different results, though at road level this may differ.

Finally, although our model excludes through-traffic (e.g., motorway freight), this omission is justified. Motorway traffic accounts for only ∼4.6% of traffic volume in South Yorkshire (to which Sheffield belongs) (Ivings, Arbabi and Punzo, 2024), and such journeys typically cannot re-routed using local traffic management strategies (Levinson and Zhu, 2012). As a result, our proposed interventions are both robust and practically actionable for urban traffic management.

5. Conclusions

In this study, we presented a novel, data-driven traffic assignment framework. Our main contribution, here, directly relates NO_2 to traffic density in a time-independent and weather-corrected manner that is transferable and can be generalised within and across cities.

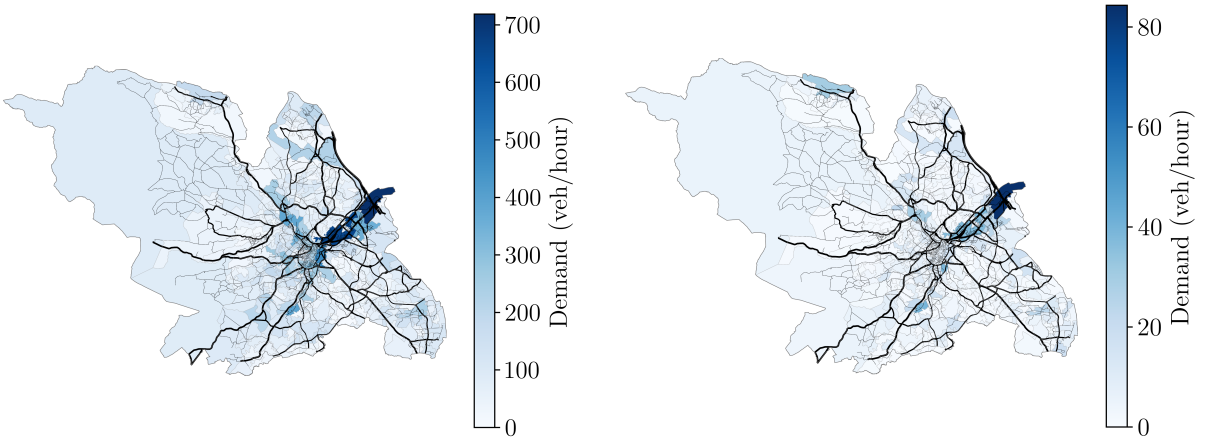
Our approach enables the evaluation of environmentally optimal traffic flows using observed pollutant concentrations, bypassing the need for tailpipe emissions estimates. By integrating de-weatherised NO_2 measurements with traffic data in a predictive modelling framework, we demonstrate the feasibility of routing strategies that reduce urban air pollution with minimal compromise to travel efficiency.

Applied to Sheffield, UK, our model achieves a 21% reduction in NO_2 concentrations relative to conventional time-minimising traffic assignment, with only a 9% increase in total travel time: equivalent to 0.3 minutes per road on average. These environmental gains result from redistributing flows away from congested residential and industrial zones toward higher-capacity arterial routes.

In the pollution-optimal scenario, the proportion of roads exceeding safety thresholds, set by the WHO, declines. Additionally, the observed correlation between NO_2 concentrations and lower-income areas is reduced, indicating potential equity co-benefits.

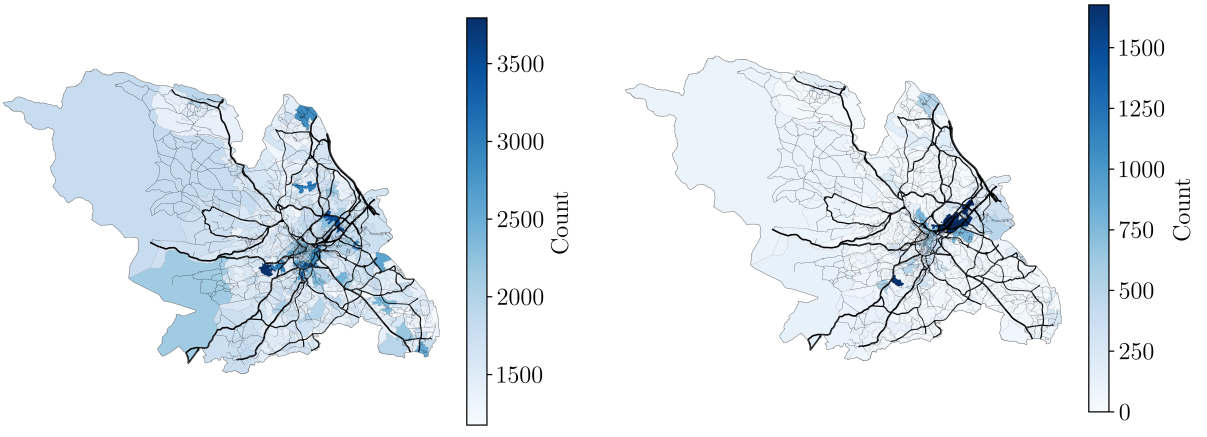
We also find that overall vehicle flow, rather than fleet composition, is the dominant driver of NO_2 concentrations under current conditions. This suggests that a transition to electric vehicles alone may not deliver immediate air quality improvements unless accompanied by broader traffic reduction strategies.

Our results underscore the potential for low-cost, data-driven traffic management tools to improve urban air quality, reduce environmental health disparities, and maintain transport efficiency. To promote optimal route choices, drivers' behavioural change may be realised through low-cost interventions such as changing speed limits along certain roads, altering traffic light timings, adding speed bumps, or active management via dynamic traffic signage. The proposed approach is widely transferable and can inform urban planning and clean air policy in cities with similar traffic and monitoring infrastructures.



(a) Inter-LSOA. (b) Intra-LSOA.

Figure 9: Map visualisations of May - July 2019 OD demand data by private motor vehicle for LSOAs in Sheffield (The Floow, 2023). Correlation coefficient $r = 0.78$. Referenced in Section 2.1.1.



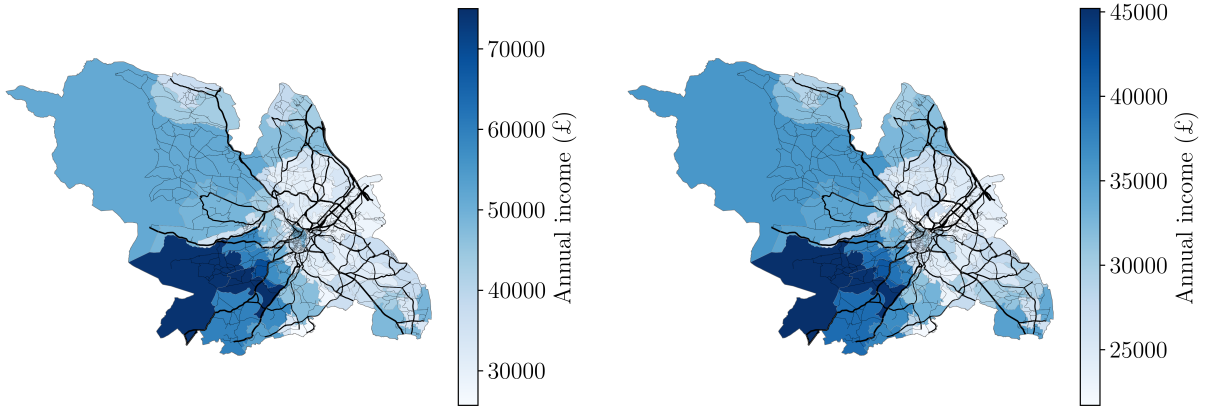
(a) Total population. (b) Number of registered companies.

Figure 10: LSOA-aggregated statistics from 2022 census data for Sheffield. Correlation coefficient $r = 0.28$. Referenced in Section 3.2.

A. Supplementary data

CRediT authorship contribution statement

Samantha Ivings: Writing, Resources, Methodology, Investigation, Conceptualization. **James A. King:** Resources, Methodology. **Alexander Roocroft:** Resources, Methodology, Writing (review & editing). **Patricio Ortiz:** Resources, Methodology, Investigation, Data curation. **Toby Willis:** Resources, Investigation. **Maria Val Martin:** Supervision, Resources, Writing (review & editing). **Hadi Arbabi:** Supervision, Writing (review & editing). **Giuliano Punzo:** Supervision, Conceptualization, Writing (review & editing).



(a) Total. (b) Net.

Figure 11: MSOA-aggregated household income statistics, from 2020 census data for Sheffield. Referenced in Section 4.

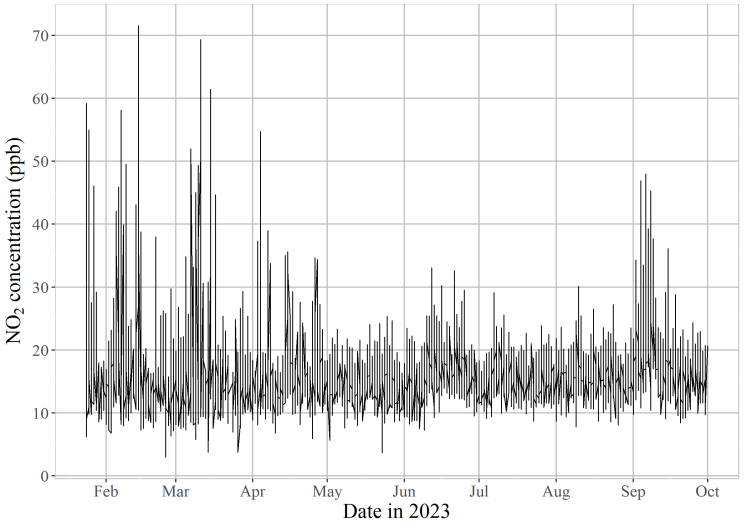
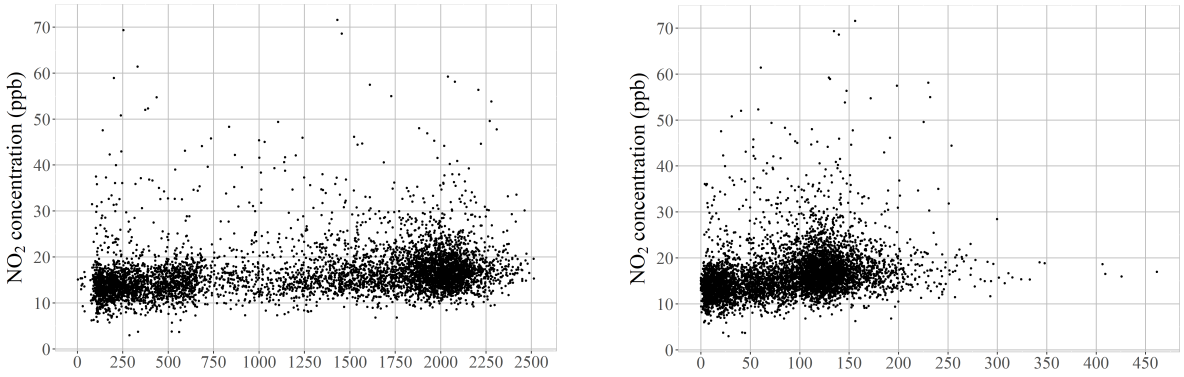


Figure 12: Time-series data for NO_2 ranging from 24th Jan - 31st Oct 2023, measured from the AQMesh sensor. Referenced in Section 2.2.3.

Acknowledgments

Funding was from TRANSITION Clean Air Network (NE/V002449/1), and the South Yorkshire Sustainability Centre (Research England Development Fund). MVM acknowledges funding from the UKRI Future Leaders Fellowship Programme (MR/T019867/1). We thank James Weber for feedback during the earlier stages of the analysis.



(a) Flow (veh/hour). (b) Density (veh/km).

Figure 13: Scatter plots showing the relationships between NO_2 concentration and measures of traffic on the Sheffield ring road. Traffic sensor readings were aggregated over hourly intervals, ranging from 24th Jan - 31st Oct 2023. Referenced in Section 2.2.2.

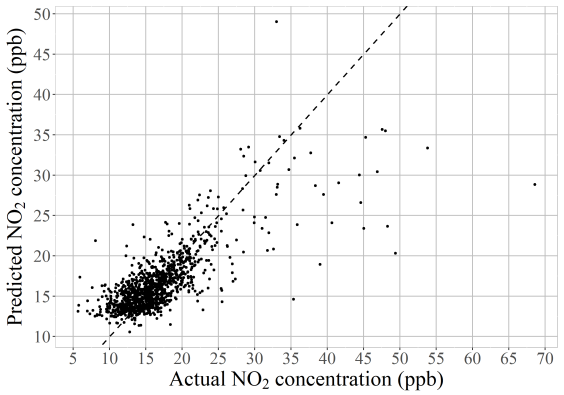


Figure 14: Scatter plot showing predicted NO_2 concentrations output by the trained random forest model, vs. actual NO_2 concentrations from sensor readings. Correlation coefficient $r = 0.77$. Guide line in $y = x$ indicates a perfect fit ($r = 1$). Referenced in Section 2.3.1.

References

Akçelik, R., 1991. Travel time functions for transport planning purposes: Davidson's function, its time dependent form and alternative travel time function. Australian Road Research 21.

AQMesh, 2017. AQMesh Website, UK, http://www.aqmesh.com/.

Barrington-Leigh, C., Millard-Ball, A., 2017. The world's user-generated road map is more than 80% complete. PLOS ONE 12, e0180698. doi:10.1371/journal.pone.0180698.

Beckmann, M., McGuire, C.B., Winsten, C.B., 1956. Studies in the economics of transportation. RAND Corporation, Santa Monica, CA .

Bliemer, M.C., Raadsen, M.P., Brederode, L.J., Bell, M.G., Wismans, L.J., Smith, M.J., 2017. Genetics of traffic assignment models for strategic transport planning. Transport reviews 37, 56–78.

Boeing, G., 2017. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems 65, 126–139. doi:10.1016/j.compenvurbsys.2017.05.004.

Boeing, G., 2024. Modeling and analyzing urban networks and amenities with osmnx. working paper. https://geoffboeing.com/publications/osmnx-paper/.

Carlier, G., Jimenez, C., Santambrogio, F., 1964. Traffic assignment manual. US Bureau of Public Roads .

Ciscal-Terry, W., Dell'Amico, M., Hadjidimitriou, N.S., Iori, M., 2016. An analysis of drivers route choice behaviour using gps data and optimal alternatives. Journal of Transport Geography 51, 119–129. URL: https://www.sciencedirect.com/science/article/pii/S0966692315002239, doi:https://doi.org/10.1016/j.jtrangeo.2015.12.003.

Davidson, K., 1966. A flow travel time relationship for use in transportation planning, in: Australian Road Research Board (ARRB) Conference, 3rd, 1966, Sydney.

DEFRA, 2023. Modelled background pollution data. https://uk-air.defra.gov.uk/data/pcm-data.

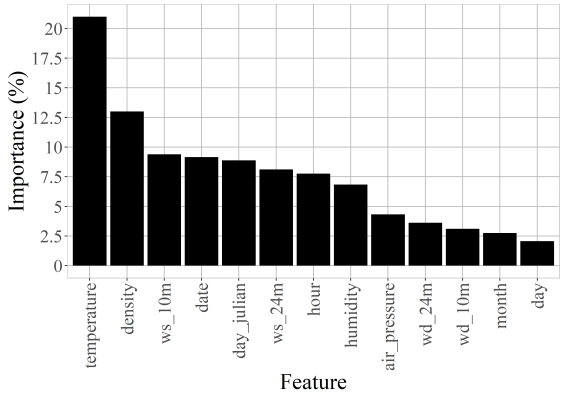


Figure 15: Bar chart showing the importance of features input to the random forest model for predicting the NO_2 output. Abbreviations: wind speed (ws), wind direction (wd), meters above ground (m). Referenced in Section 2.3.1.

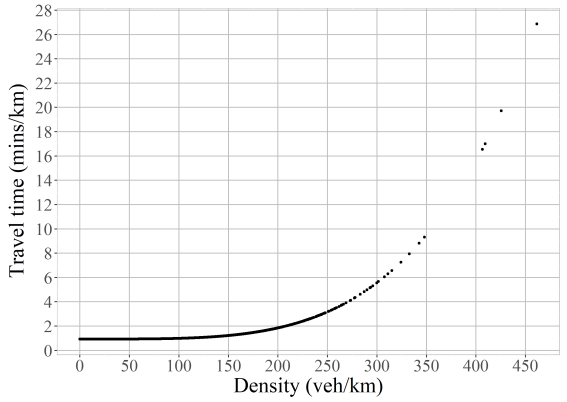


Figure 16: Travel time per km, output from the density-wise Bureau of Public Roads function applied to the Sheffield ring road (Carlier et al., 1964). Parameters: speed limit=64.4 km/hour, critical density=125 veh/km spanning both directions, α=0.15, β=4. Referenced in Section 2.3.2.

DEFRA, 2025. Emissions of air pollutants in the uk – nitrogen oxides (nox). https://www.gov.uk/government/statistics/emissions-of-air-pollutants/emissions-of-air-pollutants-in-the-uk-nitrogen-oxides-nox.

DfT, 2018. Road classification and the primary route network. https://www.gov.uk/government/publications/road-classification-and-the-primary-route-network. Accessed: 2025-04-22.

DfT, 2023. Transport analysis guidance (tag). https://www.gov.uk/guidance/transport-analysis-guidance-tag. Accessed: 2025-04-22.

DfT, 2024. National travel survey: 2023. https://www.gov.uk/government/statistics/national-travel-survey-2023.

European Environment Agency, 2011. COPERT 4 (COmputer Programme to calculate Emissions from Road Transport). Technical Report. European Union.

Frank, M., Wolfe, P., 1956. An algorithm for quadratic programming. Naval Research Logistics Quarterly 3, 95–110. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800030109, doi:https://doi.org/10.1002/nav.3800030109, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.3800030109.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics , 1189–1232.

Grange, S.K., Carslaw, D.C., 2019. Using meteorological normalisation to detect interventions in air quality time series. Science of the Total Environment 653, 578–588.

Grange, S.K., Carslaw, D.C., Lewis, A.C., Boleti, E., Hueglin, C., 2018. Random forest meteorological normalisation models for swiss pm 10 trend analysis. Atmospheric Chemistry and Physics 18, 6223–6239.

Grundström, M., Hak, C., Chen, D., Hallquist, M., Pleijel, H., 2015. Variation and co-variation of pm10, particle number concentration, nox and no2 in the urban air–relationships with wind speed, vertical temperature gradient and weather type. Atmospheric Environment 120, 317–327.

Helbing, D., 2009. Derivation of a fundamental diagram for urban traffic flow. The European Physical Journal B 70, 229–241.

Hickman, A., Baker, C., Cai, X., Delgado-Saborit, J., Thornes, J., 2018. Evaluation of air quality at the birmingham new street railway station. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 232, 1864–1878.

Table 1
Results from the current and pollution-optimal traffic assignment model scenarios. LSOA codes are reported for the LSOAs with the top ten highest median traffic flows, and top ten highest median NO_2 concentrations. Click here to open an interactive LSOA map for Sheffield in 2019 (HTML). Sheffield train station is marked with a red dot.
Referenced in Section 3.2 and Section 3.3.

Ranking in the current scenario		
1	1,474 (E01007908)	19.5 (E01008115)
2	1,438 (E01007879)	19.1 (E01033266)
3	1,297 (E01007893)	18.6 (E01007879)
4	1,285 (E01007878)	18.3 (E01008156)
5	1,249 (E01008102)	18.1 (E01007893)
6	1,209 (E01008115)	17.9 (E01007866)
7	1,150 (E01033266)	17.8 (E01033268)
8	1,149 (E01033268)	17.8 (E01033274)
9	1,133 (E01007965)	17.5 (E01008093)
10	1,108 (E01007902)	17.3 (E01033263)
Ranking in the pollution-optimal scenario		
1	1,222 (E01008095)	20.2 (E01008095)
2	1,156 (E01008097)	18.2 (E01007989)
3	1,118 (E01007893)	17.3 (E01033273)
4	1,114 (E01007989)	16.7 (E01033263)
5	1,034 (E01007908)	16.3 (E01007893)
6	1,007 (E01033263)	15.6 (E01008090)
7	980 (E01033273)	15.6 (E01008157)
8	956 (E01007965)	15.5 (E01008115)
9	942 (E01008090)	15.1 (E01008161)
10	939 (E01008157)	15.0 (E01033266)

Highways England, 2020. Design manual for roads and bridges (dmrb). https://www.standardsforhighways.co.uk/dmrb/. Accessed: 2025-04-22.

Ingole, D., Mariotte, G., Leclercq, L., 2020. Minimizing network-wide emissions by optimal routing through inner-city gating. Transportation Research Part D: Transport and Environment 86, 102411.

Ivings, S., Arbabi, H., Punzo, G., 2024. Report on Intra- and Inter- Generated Traffic in South Yorkshire. Technical Report. The University of Sheffield. https://doi.org/10.15131/shef.data.27868524.v1.

Kendrick, C.M., Koonce, P., George, L.A., 2015. Diurnal and seasonal variations of NO , NO 2 and PM 2 . 5 mass as a function of traf fi c volumes alongside an urban arterial. Atmospheric Environment 122, 133–141. URL: http://dx.doi.org/10.1016/j.atmosenv.2015.09.019, doi:10.1016/j.atmosenv.2015.09.019.

Knoop, V., Hoogendoorn, S., 2014. Network transmission model: A dynamic traffic model at network level (poster), in: 93rd Annual Meeting Transportation Research Board, Washington, USA, 12-16 January 2014.

Koutsoupias, E., Papadimitriou, C., 2009. Worst-case equilibria. Computer science review 3, 65–69.

Kovács, A., Leelőssy, Á., Tettamanti, T., Esztergár-Kiss, D., Mészáros, R., Lagzi, I., 2021. Coupling traffic originated urban air pollution estimation with an atmospheric chemistry model. Urban Climate 37, 100868.

Kucharski, R., Drabicki, A., 2017. Estimating macroscopic volume delay functions with the traffic density derived from measured speeds and flows. Journal of Advanced Transportation 2017, 4629792.

Kuik, F., Kerschbaumer, A., Lauer, A., Lupascu, A., Schneidemesser, E.V., 2018. Top – down quantification of no_x emissions from traffic in an urban area using a high-resolution regional atmospheric chemistry model. Atmospheric Chemistry and Physics , 8203–8225.

Levinson, D., Zhu, S., 2012. The hierarchy of roads, the locality of traffic, and governance. Transport policy 19, 147–154.

Macedo, E., Tomás, R., Fernandes, P., Coelho, M.C., Bandeira, J.M., 2020a. Quantifying road traffic emissions embedded in a multi-objective traffic assignment model traffic assignment model. Transportation Research Procedia 47, 648–655. URL: https://doi.org/10.1016/j.trpro.2020.03.143, doi:10.1016/j.trpro.2020.03.143.

Macedo, E., Tomás, R., Fernandes, P., Coelho, M., Bandeira, J., 2020b. Quantifying road traffic emissions embedded in a multi-objective traffic assignment model. Transportation Research Procedia 47, 648–655. doi:10.1016/j.trpro.2020.03.143.

Munir, S., Mayfield, M., Coca, D., 2021. Understanding spatial variability of no2 in urban areas using spatial modelling and data fusion approaches. Atmosphere 12, 179.

Munir, S., Mayfield, M., Coca, D., Jubb, S.A., Osammor, O., 2019. Analysing the performance of low-cost air quality sensors, their drivers, relative benefits and calibration in cities—a case study in sheffield. Environmental monitoring and assessment 191, 1–22.

Munir, S., Mayfield, M., Coca, D., Mihaylova, L.S., 2020. A nonlinear land use regression approach for modelling no2 concentrations in urban areas—using data from low-cost sensors and diffusion tubes. Atmosphere 11, 736.

National Atmospheric Emissions Inventory (NAEI), 2021a. Air quality pollutant inventories for england, scotland, wales and northern ireland: 1990–2019. https://naei.beis.gov.uk/reports/reports?report_id=1017.

National Atmospheric Emissions Inventory (NAEI), 2021b. Uk emissions interactive map. https://naei.energysecurity.gov.uk/emissionsapp/.

Ntziachristos, L., Gkatzoflias, D., Kouridis, C., Samaras, Z., 2009. Copert: a european road transport emission inventory model, in: Information Technologies in Environmental Engineering: Proceedings of the 4th International ICSC Symposium Thessaloniki, Greece, May 28-29, 2009, Springer. pp. 491–504.

OpenStreetMap contributors, 2017. Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org.

Patil, G.R., 2016. Emission-based static traffic assignment models. Environmental Modeling and Assessment , 629–642URL: http://dx.doi.org/10.1007/s10666-015-9498-7, doi:10.1007/s10666-015-9498-7.

Patriksson, M., 2015. The traffic assignment problem: models and methods. Courier Dover Publications.

Pope, R., Savage, N., Chipperfield, M., Arnold, S., Osborn, T., 2014. The influence of synoptic weather regimes on uk air quality: analysis of satellite column no_2. Atmospheric Science Letters 15, 211–217.

Pope, R.J., Arnold, S.R., Chipperfield, M.P., Latter, B.G., Siddans, R., Kerridge, B.J., 2018. Widespread changes in uk air quality observed from space. Atmospheric Science Letters 19, e817.

Redondo Bermúdez, M.d.C., Chakraborty, R., Cameron, R.W., Inkson, B.J., Val Martin, M., 2023. A practical green infrastructure intervention to mitigate air pollution in a uk school playground. Sustainability 15, 1075.

Richmond-Bryant, J., Snyder, M., Owen, R., Kimbrough, S., 2018. Factors associated with no2 and nox concentration gradients near a highway. Atmospheric environment 174, 214–226.

Rodríguez, S., Van Dingenen, R., Putaud, J.P., Dell'Acqua, A., Pey, J., Querol, X., Alastuey, A., Chenery, S., Ho, K.F., Harrison, R., et al., 2007. A study on the relationship between mass concentrations, chemistry and number size distribution of urban fine aerosols in milan, barcelona and london. Atmospheric Chemistry and Physics 7, 2217–2232.

Roocroft, A., Ramli, M.A., Punzo, G., 2023. Data-driven traffic assignment through density-based road-specific congestion function estimation. IEEE Access .

Sheffi, Y., 1985. Urban transportation networks. volume 6. Prentice-Hall, Englewood Cliffs, NJ.

Sheffield City Council (SCC), 2023. Air quality annual status report 2023. https://www.sheffield.gov.uk/sites/default/files/2024-02/2023-air-quality-annual-status-report_0.pdf.

Shen, Y., Jiang, F., Feng, S., Zheng, Y., Cai, Z., Lyu, X., 2021. Impact of weather and emission changes on no2 concentrations in china during 2014–2019. Environmental Pollution 269, 116163.

Shi, K., Di, B., Zhang, K., Feng, C., Svirchev, L., 2018. Detrended cross-correlation analysis of urban traffic congestion and no2 concentrations in chengdu. Transportation Research Part D: Transport and Environment 61, 165–173.

Spiess, H., 1990. Conical volume-delay functions. Transportation Science 24, 153–158.

TfL, 2013. Raods task force - technical note 10. what is the capacity of the road network for private motorised traffic and how has this changed over time? "https://content.tfl.gov.uk/technical-note-10-what-is-the-capacity-of-the-road-network-for-private-motorised-traffic.pdf".

TfL, 2020. Strategic transport and land use models. https://tfl.gov.uk/corporate/publications-and-reports/strategic-transport-and-land-use-models. Accessed: 2025-04-22.

The Floow, 2023. Smart mobility observatory. https://www.thefloow.com/.

Tsanakas, N., 2021. Data-driven approaches for traffic state and emission estimation. Ph.D. thesis. Linköping University Electronic Press.

Tsanakas, N., Ekström, J., Olstam, J., 2017. Reduction of errors when estimating emissions based on static traffic model outputs. Transportation research procedia 22, 440–449.

Tsanakas, N., Ekström, J., Olstam, J., 2020. Estimating emissions from static traffic models: problems and solutions. Journal of Advanced Transportation 2020, 5401792.

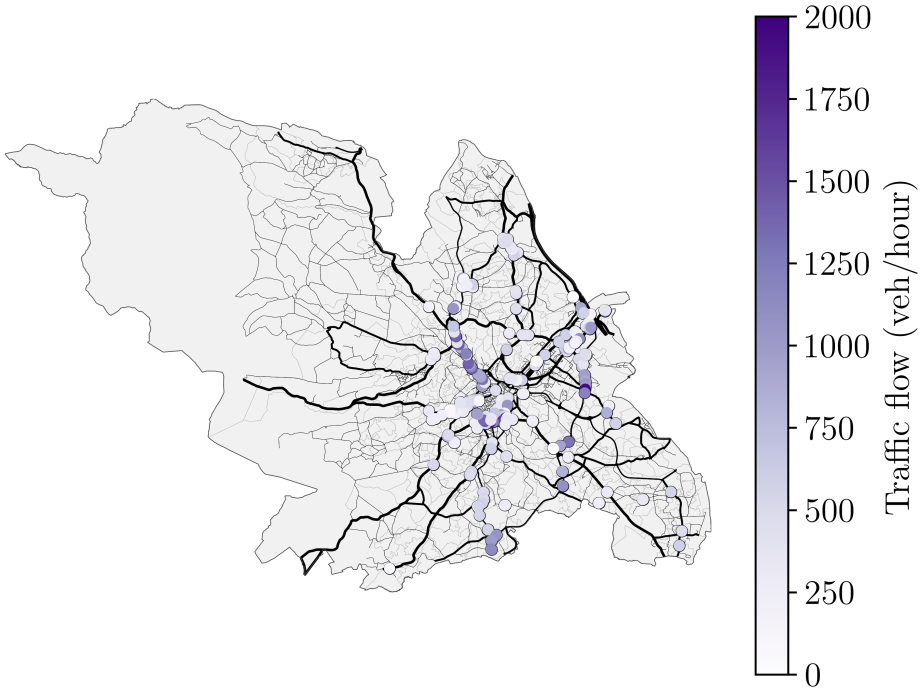
Tu, Y., Xu, C., Wang, W., Wang, Y., Jin, K., 2021. Investigating the impacts of driving restriction on no2 concentration by integrating citywide scale cellular data and traffic simulation. Atmospheric Environment 265, 118721.

UK Government, 2010. The air quality standards regulations 2010. https://www.legislation.gov.uk/uksi/2010/1001/contents.

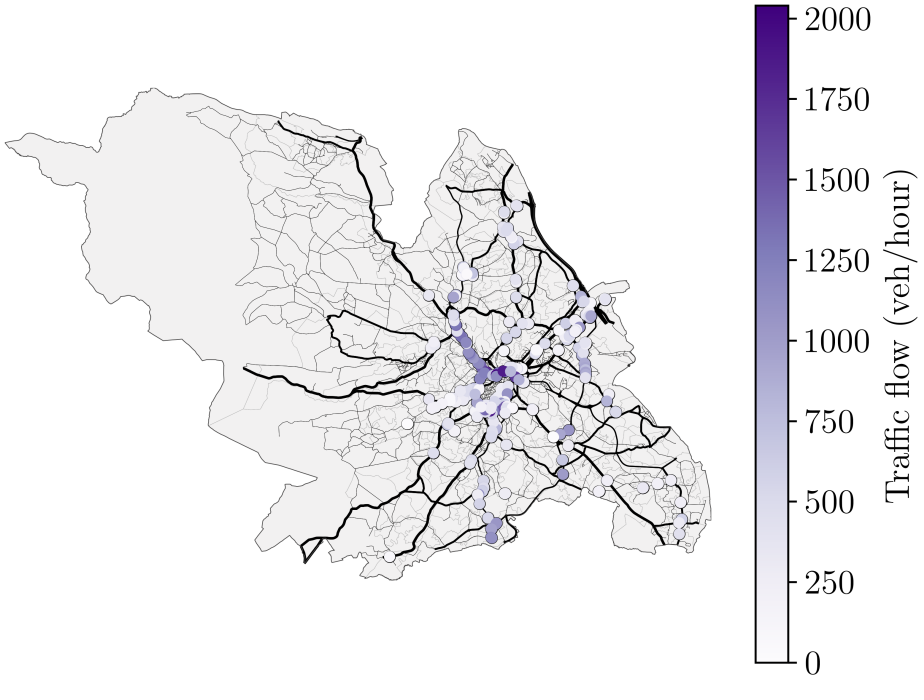
Wahlborg, D., Björling, M., Mattsson, M., 2021. Evaluation of field calibration methods and performance of aqmesh, a low-cost air quality monitor. Environmental monitoring and assessment 193, 251.

WHO, 2021. WHO Global Air Quality Guidelines: Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization, Geneva. URL: https://www.who.int/publications/i/item/9789240034228.

Zhang, H., Wang, Y., Hu, J., Ying, Q., Hu, X.M., 2015. Relationships between meteorological parameters and criteria air pollutants in three megacities in china. Environmental research 140, 242–254.



(a) 2019.



(b) 2023.

Figure 17: Traffic flows during 2019 and 2023 in Sheffield, UK. Measurements are from Sheffield City Council traffic sensors. Plots show hourly averages across peak times (7-9am, 4-7pm) for Feb - Oct of each year. Referenced in Section 2.1.1.

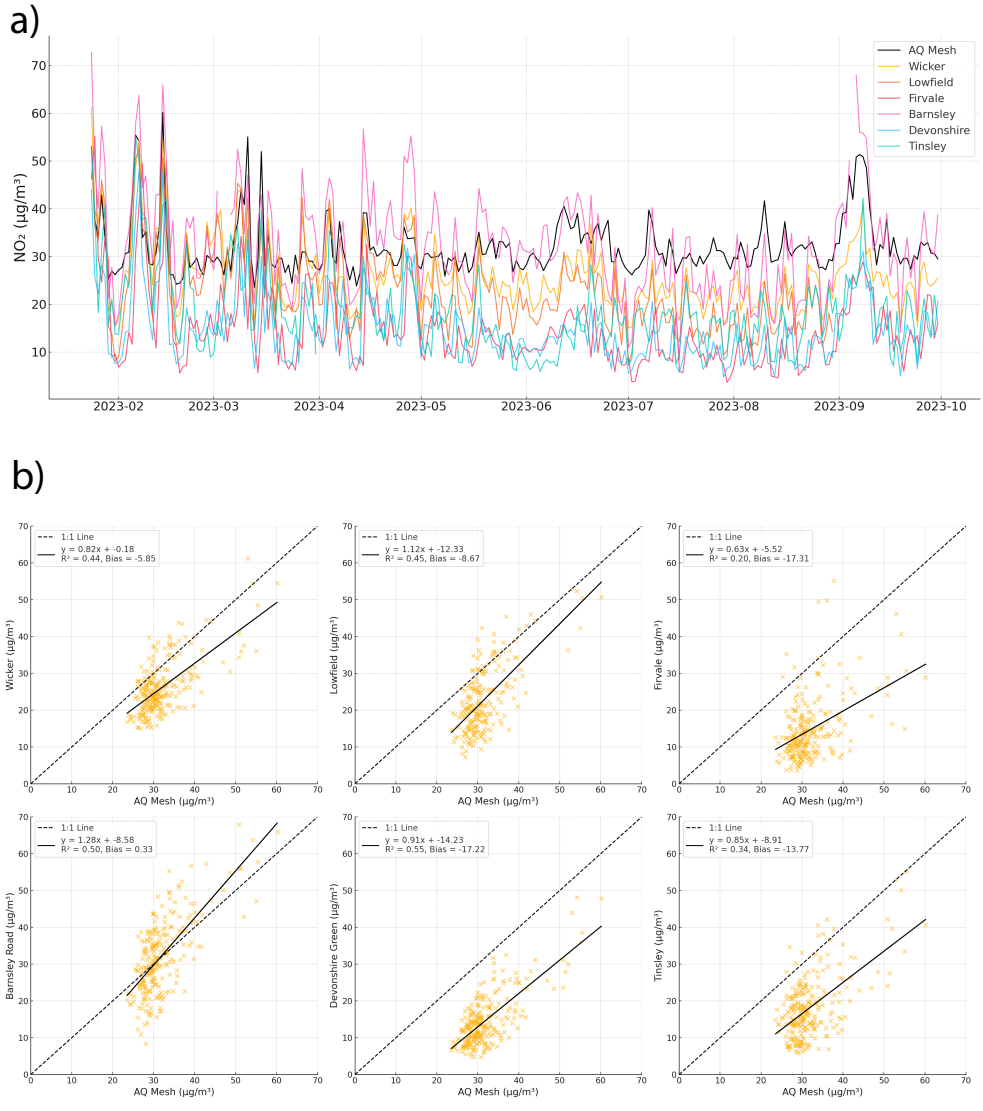


Figure 18: Comparison of the AQMesh data against DEFRA and Sheffield City Council NO_2 datasets. Referenced in Section 2.3.1.

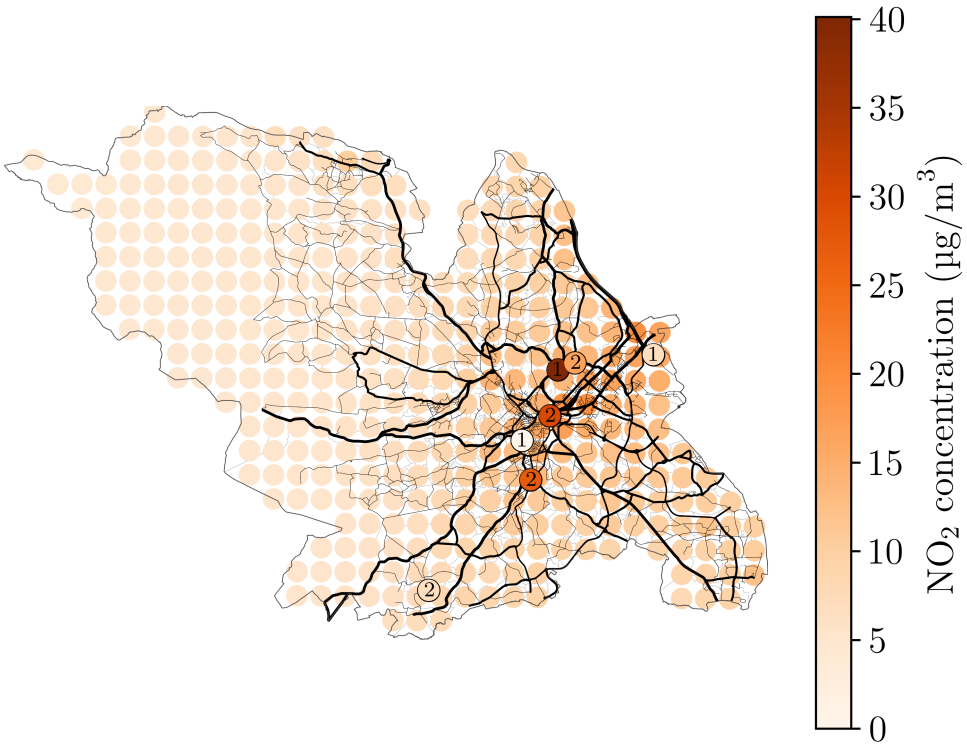


Figure 19: NO_2 concentration levels during 2023 in Sheffield, UK. City-wide background NO_2 concentrations are modelled by DEFRA (DEFRA, 2023), provided as hourly averages across 2023. Points labelled with (1) show measurements from Sheffield City Council (SCC) sensor datasets, averaged hourly across peak times (7-9am, 4-7pm) for Feb - Oct 2023, with NO_2 measurements provided in the unit $\mu g/m^3$. Points labelled with (2) show additional SCC sensor readings, averaged hourly across the same time frame, with NO_2 measurements provided in the unit parts per billion (ppb). These measurements in ppb were then converted to $\mu g/m^3$ using the ideal gas law. Referenced in Section 4.

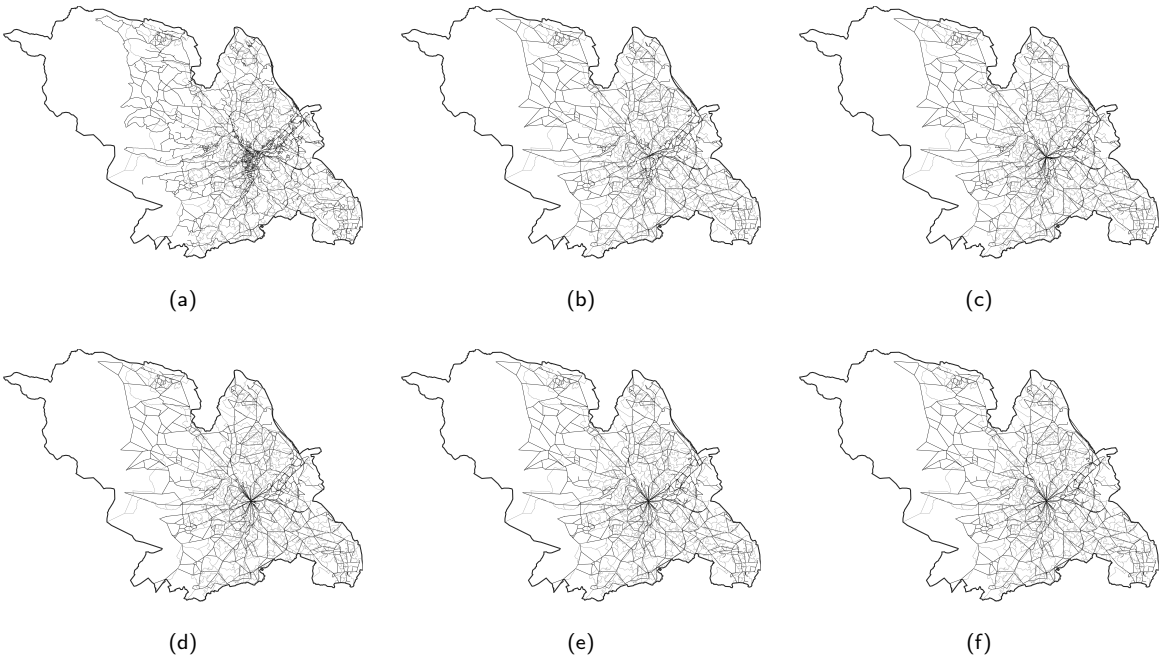


Figure 20: Road networks for the city of Sheffield, with LSOA boundaries outlined in grey. Networks (b-f) have been computationally reduced by merging nodes within a given distance threshold. The threshold used for analysis is 58.6m: the median diameter of Sheffield's LSOAs. Additional thresholds were tested within a ±20% range. **(a)** The full road network; **(b)** 46.88m threshold (20% decrease); **(c)** 52.74m threshold (10% decrease); **(d)** 58.6m threshold (used value); **(e)** 64.46m threshold (10% increase); **(f)** 70.32m threshold (20% increase).